



**FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE COMPUTACIÓN Y SISTEMAS**

**MODELO DE IMPLEMENTACIÓN DE UN DATAMART PARA
EL ANÁLISIS DE INFORMACIÓN DE POSTULANTES A LA
FACULTAD DE INGENIERÍA Y ARQUITECTURA DE LA
UNIVERSIDAD DE SAN MARTÍN DE PORRES**

**PRESENTADA POR
LUIS MIGUEL ARBULÚ HERRERA
DIEGO MARTICORENA FOSSA**

**ASESORA
NORMA BIRGINIA LEON LESCANO**

**TESIS
PARA OPTAR EL TÍTULO PROFESIONAL DE INGENIERO DE
COMPUTACIÓN Y SISTEMAS**

**LIMA – PERÚ
2021**



CC BY-NC-SA

Reconocimiento – No comercial – Compartir igual

El autor permite transformar (traducir, adaptar o compilar) a partir de esta obra con fines no comerciales, siempre y cuando se reconozca la autoría y las nuevas creaciones estén bajo una licencia con los mismos términos.

<http://creativecommons.org/licenses/by-nc-sa/4.0/>



USMP
UNIVERSIDAD DE
SAN MARTÍN DE PORRES

**FACULTAD DE
INGENIERÍA Y ARQUITECTURA**

**ESCUELA PROFESIONAL DE INGENIERÍA DE COMPUTACIÓN Y
SISTEMAS**

**MODELO DE IMPLEMENTACIÓN DE UN DATAMART PARA
EL ANÁLISIS DE INFORMACIÓN DE POSTULANTES A LA
FACULTAD DE INGENIERÍA Y ARQUITECTURA DE LA
UNIVERSIDAD DE SAN MARTÍN DE PORRES**

TESIS

**PARA OPTAR EL TÍTULO PROFESIONAL DE
INGENIERO DE COMPUTACIÓN Y SISTEMAS**

PRESENTADA POR

ARBULÚ HERRERA, LUIS MIGUEL

MARTICORENA FOSSA, DIEGO

ASESORA

MG. LEON LESCANO, NORMA BIRGINIA

LIMA – PERÚ

2014

ÍNDICE

	Página
RESUMEN	xi
ABSTRACT	xii
INTRODUCCIÓN	xiii
CAPÍTULO I PLANTEAMIENTO DEL PROBLEMA	1
1.1 Situación problemática	1
1.2 Definición del problema	1
1.3 Objetivos	2
1.4 Justificación	2
1.5 Alcance	3
1.6 Premisas	3
CAPÍTULO II MARCO TEÓRICO	4
2.1 Antecedentes	4
2.2 Bases teóricas	6
2.3 Definición de términos básicos	38
CAPÍTULO III METODOLOGÍA	44
3.1 Materiales	44
3.2 Métodos	45
3.3 Cronograma de la tesis	56
CAPÍTULO IV DESARROLLO DEL MODELO	57

4.1	Planificación y gestión del desarrollo de la tesis	57
4.2	Definición de los requerimientos del negocio	61
4.3	Modelamiento dimensional	67
4.4	Diseño físico	84
4.5	Diseño y desarrollo de los procesos de ETL	106
4.6	Diseño de la arquitectura técnica	125
4.7	Selección de los productos e instalación	126
4.8	Diseño y desarrollo de aplicaciones BI	127
4.9	Implementación (simulación)	137
CAPÍTULO V PRUEBAS Y RESULTADOS		162
5.1	Caso de prueba #1 – ETL_ FACT_ALUMNO	162
5.2	Caso de prueba #2 - Consistencia de datos	185
5.3	Caso de prueba #3 – Consistencia de Datos_Reportes	187
5.4	Caso de prueba #4 – Indicador No_Matriculados	189
CAPÍTULO VI DISCUSIÓN Y APLICACIONES		191
6.1	Discusión y aplicación	191
CONCLUSIONES		193
RECOMENDACIONES		195
FUENTES DE INFORMACIÓN		196
ANEXOS		198

ÍNDICE DE ILUSTRACIONES

	Página
Ilustración I - Componentes de una solución de inteligencia de negocios	12
Ilustración II – Proceso ETL	14
Ilustración III – Data Waterhouse	16
Ilustración IV: Datamart	17
Ilustración V – Estructura cubo OLAP	18
Ilustración VI – DashBoard	19
Ilustración VII - Interacción entre sistemas OLTP y OLAP	20
Ilustración VIII - Arquitectura MOLAP	22
Ilustración IX - Arquitectura ROLAP	24
Ilustración X - Cubo de información	26
Ilustración XI - Modelo estrella	27
Ilustración XII - Modelo copo de nieve	28
Ilustración XIII – Modelo propuesto por Ralph Kimball	31
Ilustración XIV - Ciclo de vida metodología de Kimball	47
Ilustración XV - Requerimientos funcionales	49
Ilustración XVI - Requerimientos no funcionales	50
Ilustración XVII - Análisis de fuentes de datos	50
Ilustración XVIII - Roles del equipo de trabajo	57
Ilustración XIX - Modelo lógico dimensional	83
Ilustración XX - Diseño físico	105
Ilustración XXI - Proceso ELT_DIMENSIONES_NCD	110
Ilustración XXII - Paquete ETL_DIMENSIONES_SCD	112
Ilustración XXIII - ELT_DIM_ESPECIALIDAD	112
Ilustración XXIV - ETL_DIM_MODALIDAD	113
Ilustración XXV - ELT_DIM_UBIGEO	114

Ilustración XXVI - ELT_DIM_UBIGEO_COLEGIO	115
Ilustración XXVII - ETL_DIM_COLEGIO	116
Ilustración XXVIII - Carga de severidades	116
Ilustración XXIX - Limpieza de temporales	117
Ilustración XXX – Paquete ETL_FACT_ALUMNO	118
Ilustración XXXI - EX_FUENTES_FACT_ALUMNO	119
Ilustración XXXII - RN_CW_DIM_PROCESO	120
Ilustración XXXIII - RN_FACT_ALUMNO_GRP1	120
Ilustración XXXIV - RN_FACT_ALUMNO_GRP2	121
Ilustración XXXV - CW_FACT_ALUMNO	122
Ilustración XXXVI - Carga de severidades	122
Ilustración XXXVII - Limpieza de temporales	123
Ilustración XXXVIII - Paquete mantenimiento índices	123
Ilustración XXXIX - Secuencia de ejecución paquetes ETL	124
Ilustración XL – Arquitectura técnica	125
Ilustración XLI - Cuerpo del reporte base	128
Ilustración XLII - Desarrollo del cubo OLAP	136
Ilustración XLIII - Creación de base de datos	137
Ilustración XLIV - Ejecución del script BD_DATAMART_FIA_1RA_EJEC.SQL	138
Ilustración XLV - Ejecución del script SEVERIDADES.SQL	138
Ilustración XLVI - Explorador de objetos	139
Ilustración XLVII - Mantenimiento índices	140
Ilustración XLVIII - Ejecución de dimensiones NCD	140
Ilustración XLIX - ETL dimensiones SCD	141
Ilustración L - ETL FACT_ALUMNO	142
Ilustración LI - Nuevo proyecto SSAS	143
Ilustración LII – Nuevo origen de datos	143
Ilustración LIII - Definir conexión	144
Ilustración LIV - Seleccionar tablas y vistas	145
Ilustración LV - Nuevo cubo	145
Ilustración LVI – Seleccionar método de creación de cubo	146
Ilustración LVII – Seleccionar tablas para el cubo	146
Ilustración LVIII – Crear dimensión	147
Ilustración LIX – Seleccionar método de creación de dimensión	147

Ilustración LX – Especificar información de origen	148
Ilustración LXI – Tablas relacionadas	148
Ilustración LXII – Seleccionar atributos de la dimensión	149
Ilustración LXIII – Estructura de la dimensión	149
Ilustración LXIV – Vista de dimensión	150
Ilustración LXV – Vista general de dimensiones	150
Ilustración LXVI – Nueva medida	151
Ilustración LXVII – Editar medida	151
Ilustración LXVIII – Estructura del cubo	152
Ilustración LXIX - Relacionar dimensiones con tablas	153
Ilustración LXX – Visión general del cubo	154
Ilustración LXXI – Creación de indicadores (KPI)	155
Ilustración LXXII – KPI implementados	155
Ilustración LXXIII – Pagina inicial de reportes	156
Ilustración LXXIV – Crear conexión con cubo OLAP	156
Ilustración LXXV – Actualizar conexión con cubo OLAP	157
Ilustración LXXVI - Filtros de reportes	157
Ilustración LXXVII – Reporte ejemplo	158
Ilustración LXXVIII – Filtro de KPI's	159
Ilustración LXXIX – Actualización de KPI's	160
Ilustración LXXX – Creación de tabla dinámica	160
Ilustración LXXXI – Área de trabajo tabla o gráfico dinámico	161
Ilustración LXXXII - Extracción fuente postulantes	166
Ilustración LXXXIII - Extracción fuente matrícula	166
Ilustración LXXXIV - Carga de la tabla FACT_ALUMNO	166
Ilustración LXXXV - Log de ejecución de carga	167
Ilustración LXXXVI - Tabla FACT_ALUMNO	167
Ilustración LXXXVII - Extracción fuente postulantes	170
Ilustración LXXXVIII - Extracción fuente matrícula	170
Ilustración LXXXIX - Carga de la tabla FACT_ALUMNO	170
Ilustración XC - Log EX_SRC_POSTULANTES	171
Ilustración XCI - Log EX_SRC_Matrícula	171
Ilustración XCII - Log FACT_ALUMNO	171
Ilustración XCIII - Tabla FACT_ALUMNO	172

Ilustración XCIV - Extracción fuente postulantes	175
Ilustración XCV - Carga de severidades	176
Ilustración XCVI - Carga tabla FACT_ALUMNO	176
Ilustración XCVII - Log de EX_SRC_POSTULANTES	177
Ilustración XCVIII - Log de CW_SEVERIDADES	177
Ilustración XCIX - Log de carga FACT_ALUMNO	177
Ilustración C - Tabla severidades	178
Ilustración CI - Tabla FACT_ALUMNO	178
Ilustración CII – RN_FACT_ALUMNO_GRP2_2	180
Ilustración CIII - Carga de severidades	181
Ilustración CIV - Carga tabla FACT_ALUMNO	181
Ilustración CV - Log de RN_FACT_ALUMNO_GRP2_2	182
Ilustración CVI - Log de CW_SEVERIDADES	182
Ilustración CVII - Log de carga FACT_ALUMNO	182
Ilustración CVIII - Tabla severidades	183
Ilustración CIX - Tabla FACT_ALUMNO	183
Ilustración CX - Limpieza archivos temporales	184
Ilustración CXI - Eliminación archivos temporales	184
Ilustración CXII - Carpeta de archivos temporales	184
Ilustración CXIII - Resultado datamart consulta	186
Ilustración CXIV - Resultado del cubo	186
Ilustración CXV – Resultado cubo	188
Ilustración CXVI – Resultado Excel	188
Ilustración CXVII – Resultado Cubo_Indicador	190
Ilustración CXVIII – Resultado Excel_Indicador	190

ÍNDICE DE TABLAS

	Página
Tabla I - Comparación de soluciones	37
Tabla II - Herramientas de software	44
Tabla III – Tiempo total del desarrollo	60
Tabla IV - Definición de requerimientos funcionales	62
Tabla V - Definición de requerimientos no funcionales	63
Tabla VI – Análisis de fuentes de datos 1 (DATOS_POSTULANTE.XLS)	64
Tabla VII – Análisis de fuentes de datos 2 (DATOS_MATRICULA.XLS)	65
Tabla VIII - Análisis de fuentes de datos 3	66
Tabla IX - Comparación modelo estructura de datos	67
Tabla X - Establecimiento del nivel de granularidad	68
Tabla XI - Variables de análisis	69
Tabla XII - Dimensión Carrera	70
Tabla XIII - Dimensión Colegio	71
Tabla XIV - Dimensión Estado Civil	71
Tabla XV - Dimensión Facultad	72
Tabla XVI - Dimensión Género	72
Tabla XVII - Dimensión Ingreso	73
Tabla XVIII - Dimensión Letra Inicial	73
Tabla XIX - Dimensión Matricula	74
Tabla XX - Dimensión Modalidad	74
Tabla XXI - Dimensión Proceso	75
Tabla XXII - Dimensión Rango Edad	75
Tabla XXIII - Dimensión Rango Egreso Colegio	76
Tabla XXIV - Dimensión Rango Mérito	76

Tabla XXV - Dimensión Rango Puntaje	77
Tabla XXVI - Dimensión Sede	77
Tabla XXVII - Dimensión Tipo Prueba	78
Tabla XXVIII - Dimensión Tipo Colegio	78
Tabla XXIX - Dimensión Trabaja	79
Tabla XXX - Dimensión Ubigeo	79
Tabla XXXI - Dimensión Ubigeo Colegio	80
Tabla XXXII - Tabla de hechos	81
Tabla XXXIII - Def. columnas dimensión Carrera	85
Tabla XXXIV - Def. columnas dimensión Colegio	85
Tabla XXXV - Def. columnas dimensión Estado Civil	86
Tabla XXXVI - Def. columnas dimensión Facultad	86
Tabla XXXVII - Def. columnas dimensión Género	86
Tabla XXXVIII - Def. columnas dimensión Ingreso	87
Tabla XXXIX - Def. columnas dimensión Letra Inicial	87
Tabla XL - Def. columnas dimensión Matricula	87
Tabla XLI - Def. columnas dimensión Modalidad	88
Tabla XLII - Def. columnas dimensión Proceso	88
Tabla XLIII - Def. columnas dimensión Rango Edad	89
Tabla XLIV - Def. columnas dimensión Rango Egreso Colegio	89
Tabla XLV - Def. columnas dimensión Rango Mérito	90
Tabla XLVI - Def. columnas dimensión Rango Puntaje	90
Tabla XLVII - Def. columnas dimensión Sede	91
Tabla XLVIII - Def. columnas dimensión Tipo Prueba	91
Tabla XLIX - Def. columnas dimensión Tipo Colegio	92
Tabla L - Def. columnas dimensión Trabaja	92
Tabla LI - Def. columnas dimensión Ubigeo	93
Tabla LII - Def. columnas dimensión Ubigeo Colegio	93
Tabla LIII – Def. columnas tabla de hechos	94
Tabla LIV - Índices dimensión Carrera	96
Tabla LV - Índices dimensión Colegio	96
Tabla LVI - Índices dimensión Estado Civil	97
Tabla LVII - Índices dimensión Facultad	97
Tabla LVIII - Índices dimensión Género	97

Tabla LIX - Índices dimensión Ingreso	98
Tabla LX - Índices dimensión Letra Inicial	98
Tabla LXI - Índices dimensión Matricula	98
Tabla LXII - Índices dimensión Modalidad	99
Tabla LXIII - Índices dimensión Proceso	99
Tabla LXIV - Índices dimensión Rango Edad	99
Tabla LXV - Índices dimensión Rango Egreso Colegio	100
Tabla LXVI - Índices dimensión Rango Mérito	100
Tabla LXVII - Índices dimensión Rango Puntaje	101
Tabla LXVIII - Índices dimensión Sede	101
Tabla LXIX - Índices dimensión Tipo Prueba	101
Tabla LXX - Índices dimensión Tipo Colegio	102
Tabla LXXI - Índices dimensión Trabaja	102
Tabla LXXII - Índices dimensión Ubigeo	102
Tabla LXXIII - Índices dimensión Ubigeo Colegio	103
Tabla LXXIV – Índices tabla de hechos Alumno	103
Tabla LXXV - Cabecera para la impresión	127
Tabla LXXVI - Pie de página para la impresión	128
Tabla LXXVII - Reportes pre-establecidos	129
Tabla LXXVIII - Diseño de indicadores (KPI)	131
Tabla LXXIX - Especificación cubo postulantes	134
Tabla LXXX - Caso prueba #1	162
Tabla LXXXI – Datos del postulante	164
Tabla LXXXII – Datos de matricula	165
Tabla LXXXIII – Datos postulantes duplicados	168
Tabla LXXXIV – Datos matriculas duplicadas	169
Tabla LXXXV – Datos nullos de postulantes	173
Tabla LXXXVI – Datos de ingresantes	174
Tabla LXXXVII – Datos de postulantes	179
Tabla LXXXVIII – Datos de ingresantes	180
Tabla LXXXIX - Caso prueba #2	185
Tabla XC - Caso prueba #3	187
Tabla XCI - Caso prueba #4	189
Tabla LXXXIV - Discusión y aplicación	191

RESUMEN

La presente tesis consiste en el desarrollo de un modelo que permita implementar un Datamart para analizar la información de los alumnos postulantes e ingresantes a la Facultad de Ingeniería y Arquitectura. Dicha información la brinda la Oficina de Admisión de la Universidad San Martín de Porres a la Oficina de Registros Académicos de la Facultad. Actualmente, la consolidación de la información la realiza la Oficina de Admisión de forma manual, lo que comprende un alto costo en recursos, tiempo y un alto riesgo operativo. La metodología utilizada para el desarrollo del trabajo fue la Metodología Kimball, que se alineó al alcance y los objetivos de la investigación, con el fin de cubrir los requerimientos del usuario. Como resultado, el trabajo marcó la pauta para una futura implementación de un Datamart que permita un mejor análisis de la información, como apoyo a la Oficina de Registros Académicos o al nivel estratégico de la Facultad, en la toma de decisiones proactivas, el desarrollo de campañas dirigidas a diversos sectores, visitas guiadas a instituciones educativas, ferias estudiantiles, entre otros. La investigación aplicada, permite concluir que una futura implementación de un Datamart en la Oficina de Registros Académicos reduce los tiempos de elaboración y entrega de reportes a usuarios finales. También facilita el análisis de la información en el modelo multidimensional. Asimismo, comprueba la capacidad y el potencial beneficio del Datamart y, así lograr mejorar el análisis de información para la toma de decisiones.

Palabras clave: datamart, cubo, olap, modelos, inteligencia de negocios (BI), postulantes, ETL, información, análisis, decisiones, Oficina de Admisión, reportes.

ABSTRACT

This thesis involves the development of a model for a Datamart implementation, to analyze information from applicants to the Faculty of Engineering and Architecture. This information is provided by the University San Martin de Porres Admissions Office to the Academic Records Office of the faculty. Currently, the information consolidation is manually made by the Admissions Office, which comprises a high cost in resources and time and turns a higher operational risk. The methodology used to develop the product was the Kimball methodology, which is aligned to the scope and objectives of the investigation, to meet user requirements. As a result, this research work set the tone for a future Datamart implementation, to allow better information analysis, supporting the Academic Records Office, in making proactive decisions such as developing campaigns for various sectors, guided visits to educational institutions, student fairs, among others. This research leads to the conclusion that a future Datamart implementation, for the Academic Records Office, reduces development and deliver reports time to end users, as well as facilitates the analysis information because of the multidimensional model design. Likewise, it verifies the potential benefit that would have the Datamart implementation for achieving better information analysis in decision making.

Key words: datamart, cube, olap, models business intelligence applicants, ETL, information, analysis, decisions, admissions office, reports.

INTRODUCCIÓN

La presente tesis se ha realizado con el objetivo de brindar a los altos directivos de la Facultad de Ingeniería y Arquitectura de la USMP las herramientas de gestión de información necesarias para mejorar los procesos de admisión y selección de la facultad, así como para soportar la toma de decisiones estratégicas de manera oportuna y acertada, contribuyendo al crecimiento y desarrollo de la universidad, teniendo en cuenta que es una de las más grandes del país.

El diseño de la solución se encuentra enfocado en el desarrollo de un modelo de Datamart para poder realizar un análisis con cubos OLAP con la información que la facultad de ingeniería y arquitectura de la USMP genera en los procesos de admisión, este análisis esta estrictamente acotado al universo de postulantes e ingresantes de dicha facultad.

Actualmente, la Facultad cuenta con todos los registros de los postulantes e ingresantes, los cuales van desde su nombre y apellido hasta nombre de carrera, puntaje de ingreso, ubicación geográfica del colegio, modalidad de admisión, entre otros. Todo este vasto universo de datos permite realizar diversos tipos de análisis y para ello los cubos OLAP son una pieza fundamental para la explotación de la información.

El problema principal es que la elaboración de los diferentes reportes ad hoc y las respectivas estadísticas por cada proceso de admisión en la universidad se encuentra estrechamente relacionada con una alta manualidad que trae un efecto negativo generando una entrega dilatada de estos.

Por ello, el objetivo general consiste en elaborar un modelo que permita la implementación de un datamart en la FIA - USMP para mejorar el análisis de la información de los postulantes a las diferentes escuelas profesionales, así como reducir el tiempo que lleva la elaboración de reportes para la toma de decisiones.

La investigación está compuesta por seis (6) capítulos. En el primero se presenta la situación problemática, definición del problema, alcance de la investigación y los respectivos objetivos. En el segundo, se expone el marco teórico. En el tercero, se presenta la metodología que señala la línea base de la investigación con la respectiva infraestructura de tecnología de información con la que se desplegará el desarrollo. En el cuarto, se plantea la solución, desde la planificación hasta el desarrollo. En el quinto, se explican los resultados obtenidos. Y, en el sexto capítulo, son ponderados y analizados permitiendo un desglose de conclusiones y la presentación de contrastes versus los objetivos planteados.

CAPÍTULO I

PLANTEAMIENTO DEL PROBLEMA

1.1 Situación problemática

Actualmente, la Oficina de Registros Académicos recibe reportes con las estadísticas de los ingresantes a las diversas carreras de la facultad, a través de la Oficina de Admisión de la USMP. Esta unidad utiliza los sistemas transaccionales como fuente para realizar reportes frecuentes, para una potencial toma de decisiones, teniendo en cuenta que dichos sistemas transaccionales no fueron elaborados como una herramienta para la síntesis, análisis y consolidación de datos.

En cada proceso de admisión, se elaboran de forma manual estadísticas y reportes a pedido (ad-hoc), cuya elaboración y entrega tarda aproximadamente un mes calendario. Esto a su vez, genera un gran nivel de riesgo operativo debido a que existe la posibilidad de cometer algún error en la consolidación de la data. Por otro lado, el obtener la información directamente de los sistemas transaccionales genera una degradación de estos afectando su desempeño.

1.2 Definición del problema

La consolidación, procesamiento y generación de los reportes frecuentes y personalizados, para la FIA - USMP por parte de la Oficina de Admisión, se realiza de manera manual y comprende un elevado tiempo de elaboración.

1.3 Objetivos

1.3.1 Objetivo general:

Desarrollar un modelo que permita la implementación de un Datamart en la FIA - USMP para mejorar el análisis de la información de los postulantes a las diferentes escuelas profesionales, así como reducir el tiempo que lleva la elaboración de reportes para la toma de decisiones.

1.3.2 Objetivos específicos:

- Realizar un análisis de negocio para la oficina de admisión y para la FIA - USMP, obteniendo los requerimientos e indicadores claves.
- Identificar y analizar las diversas fuentes de datos que servirán como input para el Datamart de la Oficina de Registros Académicos.
- Elaborar un modelo de Base de datos multidimensional que permita la explotación de la información de ingresantes.
- Desarrollar los procesos de extracción, transformación y carga de los datos al modelo de base de datos multidimensional
- Explotar la información a través de cubos OLAP y Reportes.

1.4 Justificación

La elaboración de la tesis beneficia, a nivel estratégico, a la FIA - USMP, al realizar un análisis y explotación de la información proporcionada por Oficina de Admisión, al brindar patrones y tendencias para generar estrategias de negocio.

Por otro lado, permite a la Oficina de Registros Académicos acceder a la información con alta disponibilidad y contar con herramientas de explotación de información.

Además, el análisis BI ofrece, a nivel estratégico, la posibilidad de tomar decisiones proactivas tales como el desarrollo de campañas dirigidas a diversos sectores, visitas guiadas a instituciones educativas, ferias estudiantiles, entre otros.

Por último, el uso apropiado de las herramientas de BI genera ventajas competitivas, ya que el análisis de la información reduce notablemente la incertidumbre para la toma de decisiones y da la posibilidad de trabajar con data histórica para poder elaborar proyecciones.

1.5 Alcance

El presente trabajo fue desarrollado exclusivamente para la oficina de registros académicos de la FIA - USMP. Además, hay que tener en cuenta que se realizó el análisis, diseño y construcción del Datamart, mas no la implementación de este, porque existe una limitación de tiempo de solo cuatro meses para culminar con la investigación.

Debido a la migración de la base de datos que actualmente se está realizando a nivel de la Universidad, que contempla la estandarización del actual motor de base de datos a DB2, es posible extender, en un futuro, el alcance de esta propuesta a las facultades restantes de la universidad.

1.6 Premisas

- La FIA - USMP no cuenta con una solución de Inteligencia de Negocios.
- Actualmente existe un proyecto en ejecución que integrará las bases de datos de las distintas facultades en una sola.
- La implementación del modelo es avalada por el decanato de la FIA - USMP.
- El proceso de análisis de datos utilizando BI, es más eficiente que el proceso manual de análisis de data.

CAPÍTULO II

MARCO TEÓRICO

2.1 Antecedentes

2.1.1 Inicios de la inteligencia de negocios (BI)

Desde un punto de vista tecnológico, la etapa que se puede definir como "pre-BI" no se encuentra muy lejos, hace más de cuarenta años, cuando los ordenadores dejaron de ocupar habitaciones enteras, porque se han reducido en tamaño y, a la vez, las empresas han comenzado a considerar los datos como una potencial fuente de información que se puede procesar y explotar. explotar

Sin embargo, en ese momento, aún no se desarrollaban los recursos efectivos para permitir un análisis estructurado y consistente de los datos que influya al momento de tomar decisiones.

De igual manera se logró recolectar información de forma integrada, resultado de sistemas transaccionales establecidos con predominio de datos jerárquicos, pero que, al momento de ser recolectados en forma de bloques delimitados de información, nos dan una vista global de la organización, pero no aportaban al resultado final ni beneficios negociables, estamos maso menos a finales de la década de 1960, una época en la que las tarjetas perforadas, los transistores y el lenguaje COBOL eran la tecnología vigente para la informática. Una época en la que la informática se consideraba desconocida era una realidad y modernidad lejana.

En la época de 1970 las cosas empezaron a cambiar con la aparición de tecnologías que brindaban almacenamiento y un mejor acceso a los datos. A partir de esa época, la computadora se considera una herramienta que permitía coordinar

actividades comerciales y a su vez las bases de datos se consideran piezas clave para asegurar que la empresa tenga un valor diferencial en el mercado versus su competencia.

A principios de la década de 1990, la gran mayoría de las corporaciones solo contaban con Centros de Información que, teniendo un stock de datos, ofrecían muy poca información, y estos centros respondían, de alguna manera, a los requerimientos de los líderes y altos ejecutivos. El mercado comenzó a sofisticarse y se movió hacia la mejora de las herramientas de software para ofrecer información en el momento oportuno y definir acciones, las cuales apuntaban a mejorar el desempeño en el mundo de los negocios.

Entre los años 1992 y 1993 se estaba formado el Datawarehouse, una base de datos de información, datos consolidados, limpios y estandarizados, considerado por los expertos de esta rama como uno de los engranajes principales para que se logre ejecutar un desarrollo de BI. Sin embargo, las opiniones no siempre son unánimes cuando se habla del BI. Según varios expertos, es importante que la organización que quiera implantar análisis de data cuente con un repositorio específico para recoger los datos ya transformados en información. El repositorio no tiene que ser un Datawarehouse, sino algo con menos complejidad, como un Datamart (base de datos personalizada para temas o aspectos específicos), o una base de datos relacional, pero dividida del entorno operacional y enfocada a la preservación de información " ejecución de diversos análisis y proyecciones. (Intel Corporation, 2009)

El Business Intelligence (Inteligencia de Negocios) como definición, formalmente fue mencionada en 1996 cuando Gartner Group lo incluyó en uno de sus reportes donde, a grandes rasgos y entre líneas, indico que se requieren herramientas que apoyen a los usuarios de negocios en la toma decisiones porque se requiere más que intuición para las decisiones correctas. (Gartner Group, 1996)

2.2 Bases teóricas

2.2.1 Las empresas en la era de la información

El Instituto Tecnológico de Buenos Aires nos comenta que cuando las empresas iniciaron a registrar la data de sus operaciones en soportes físicos de almacenamiento, con el objetivo de administrar y controlar la información, surgió la necesidad de usar esa información para potenciar y mejorar el negocio.

La importancia estratégica de la información empezó a tomar relevancia cuando el mercado se hizo muy competitivo y cada vez se ofrecían a la venta más productos similares, de diferentes empresas, momento en el cual el consumidor podía seleccionar el que por algún motivo más le agradaba. Surgió entonces la necesidad de fidelizar a los clientes a través de nuevos servicios o productos, los clientes empezaron gradualmente a ver más allá del producto que ellos adquirirían y de forma tangible podían verificar las ventajas y la diferenciación entre los distintos productos que adquirirían.

Las empresas actualmente definen a la información como un activo de la empresa, por lo que a los datos se le empiezan a tratar de una forma más metodológica. Además, se da un peso específico de alto valor a la información como principal argumento que sustenta a la empresa. Existen empresas que principalmente ofrecen servicios y su actividad principal gira en torno al manejo de la información (bancos, seguros, casas de apuestas, e-commerce, etc.), en ellas es más sencillo identificar la importancia de la información. Sin embargo, existen otros donde su línea principal gira en torno a la producción, en ellos se debe identificar información para generar análisis, evaluar y mejorar su producción (mermas, líneas de producción, procesos internos, distribución de materias, abastecimiento, stocks y almacenes, procesos internos, preferencias de los clientes, etc.). Cada empresa, de hecho, trata de sacar provecho por todos los medios posibles, para transformar la data en conocimiento que mejore a la organización a todo nivel, les genere valor y se transforme en una ventaja competitiva en el mercado

La idea de las organizaciones necesitadas de información no nace repentinamente, es más, al estar los datos archivados hay que entender que en un momento determinado tendrían una finalidad utilitaria, de lo contrario todos los datos de control serían descartados instantáneamente. Surge de repente Es la necesidad imprescindible de responder con rapidez a las necesidades de información para la toma de

decisiones con el fin de ayudar de una u otra forma a mejorar los procesos internos de la organización. (Instituto Tecnológico de Buenos Aires, 2009).

2.2.2 Información que las empresas necesitan

La tendencia hoy en día es solicitar reportes a niveles donde la gerencia se basó en la intuición, experiencia, sentido común y criterio para tomar decisiones. Si bien siempre se ha requerido información a nivel operativo, históricamente no ha habido restricciones de ningún tipo. Por el contrario, la dinámica de los mercados ha obligado a las organizaciones a poner información estratégica para los directivos y altos ejecutivos. Este comportamiento se ha extendido motivado principalmente por la agilidad y utilidad de los flujos de información compartidos en todos los niveles de la organización para diferentes propósitos independientemente de los cargos. A través de los años las empresas se han dado cuenta que los niveles gerenciales tienen una enorme responsabilidad en la toma de decisiones, porque el impacto que generan afecta a toda la empresa, pero también hay más colaboradores y mandos medios que toman decisiones y aunque estas no tienen un impacto generalizado en la empresa, también deben ser precisas y más oportunas, ya que algunos grupos dependen de ellos. Los managers, gerentes, jefes, coordinadores, directivos y todos aquellos que toman decisiones deben tener suficiente información para respaldar su trabajo diario, donde ocupan en la pirámide organizacional se vuelven secundarios cuando el enfoque es hacia la gestión del proceso. Y todas las posiciones tienen una relación y dependencia.

Para el (Instituto Tecnológico de Buenos Aires, 2009) de forma general en una pirámide organizacional, las necesidades de información se segmentan en 3 tipos:

- Información estratégica
- Información táctica
- Información técnico operacional.

a) Información estratégica

Apoya la toma de decisiones de las áreas claves de la organización para conseguir los objetivos planeados en su misión, Su característica principal es que no realizan cargas de información periódica y no tienen grandes volúmenes de datos, sin embargo, la información que almacenan está relacionada a un aspecto más

cualitativo que el aspecto cuantitativo, que puede indicar cómo funcionará la empresa hoy y en el futuro.

b) Información táctica

Esta información respalda las estrategias que se aplicarán en las actividades de la empresa, siempre están buscando que la estrategia de la alta dirección de la organización sea concretada.

c) Información técnico operacional

Este tipo de información hace referencia a las operaciones tradicionales, las cuales se ejecutan de manera rutinaria en las organizaciones a través de sistemas masivos de adquisición de datos y procesamiento de transacciones. Comúnmente, están asociados al personal operativo o de coordinación.

2.2.3 Inteligencia de negocios (BI)

La capacidad de observar a toda una organización pasa por la información, quien maneja correctamente la información tiene en sus manos el poder del éxito Actualmente, *Business intelligence* ofrece a las organizaciones, cualquiera que sea su ámbito de pertenencia, la capacidad de procesar y analizar los datos generados por sus sistemas transaccionales.

La información obtenida en tiempo real es utilizada por el nivel estratégico para la toma de decisiones, generar ventajas competitivas y orientar el crecimiento de la compañía para medir el desempeño de las diferentes áreas, así como tener una visión global alcance de toda la organización; Es sobre todo una alternativa tecnológica que permite generar conocimiento para tomar decisiones oportunas a nivel estratégico.

Business intelligence (BI) nos permite diseñar e implementar estrategias basadas en información para la organización, ofrecer soluciones a necesidades o incluso ayudar a una gestión eficaz basada en intereses y objetivos.

Las decisiones en la dirección de una organización siempre están ligadas a la incertidumbre, y esta existe cuando quien la toma no tiene suficiente información para conocer las consecuencias de haberlas tomado capaz de minimizar el nivel de riesgo.

“Se ha demostrado que todas las decisiones que se toman en un negocio deben surgir de las mismas operaciones y fenómenos comerciales que ocurren en el ambiente organizacional. Estas operaciones y fenómenos se expresan en términos de información”. (Araníbar, 2003)

Los datos son el reflejo de actividades y hechos que son recogidos en la compañía y que aún no están trabajados. Estos datos se convierten en información cuando se procesan, esto quiere decir que se organizan de forma que tengan sentido, sean útiles y de los que se puedan extraer conclusiones, todo ello ayuda a reducir la incertidumbre del receptor.

El conocimiento es información que ha sido organizada y procesada para aportar comprensión, experiencia y aprendizaje acumulados, cuando se aplica a un problema o actividad. La información que se procesa proporciona al destinatario conocimiento organizacional del más alto valor potencial.

En resumen, el concepto de inteligencia de negocios es el proceso de identificar, recopilar, entregar y analizar datos para convertirlos en conocimientos y poderosas acciones comerciales. Los colaboradores que participan en los diferentes procesos de negocio deben hacer uso de herramientas, hardware, software y diferentes tecnologías que les permitan captar, almacenar, transformar, analizar y dar acceso, a los datos de una forma simple y manejable.

2.2.4 Beneficios que aporta la inteligencia de negocios

Una compañía se vuelve competitiva si aplica BI correctamente, sin embargo, una situación muy común es que los líderes de tecnología de información pueden asegurar que los beneficios son tangibles, pero el nivel estratégico de las organizaciones no entiende los beneficios de esta inversión. y no ven con buenos ojos la salida de dinero y para tales efectos incluso lo consideran como un gasto extra, especialmente si se trata de una empresa del sector small business. Debemos tener en cuenta que las soluciones de BI tienen un alto retorno de inversión en el negocio, y esto se debe a la cantidad de decisiones críticas que se toman en base a la explotación de la información que realizan las soluciones de BI.

Uno de los objetivos fundamentales de los sistemas transaccionales es proporcionar a todos los niveles de la compañía la información necesaria

para dirigir sus actividades, De la misma manera, los gestores pueden requerir recopilar más o menos información en función al mayor o menor riesgo que deseen asumir.

Por lo tanto, el BI nos servirá como soporte para la toma de decisiones y a posteriori encontrar o descubrir patrones ocultos que hasta el momento desconocíamos, Según (Cano, 2007) podemos tener diferentes tipos de beneficios:

Beneficios tangibles (más ingresos, menos costos):

- Mejorar la captación de clientes.
- Reducir la tasa de fuga de clientes
- Aumentar la lealtad de los clientes.
- Aumento de ingresos por crecimiento de ventas.
- Evitar las pérdidas ocasionadas por las ventas de nuestros competidores.

Entre los beneficios intangibles tenemos:

- Disponer de información para la toma de decisiones supondrá que más usuarios utilizarán esta información para tomar decisiones y mejorar nuestra posición competitiva
- Obtener ventajas competitivas
- Menos dependencia de los sistemas desarrollados
- Integración más alta de la información
- Aumentar la satisfacción del cliente

Por último, los beneficios estratégicos son los que ayudan a que la estrategia de la organización sea implementada y desarrollada más fácilmente:

- Mayor visibilidad de la gestión
- Estrategias de coaching
- Mejor desempeño del proceso de toma de decisiones, rápido, más informado y basado en hechos
- Identificación y retención de clientes con mayor potencial
- Mejor capacidad para implementar estrategias de fijación de precios.

2.2.5 Características de la Inteligencia de Negocios

a) Información:

La información para el BI siempre tiene que ser proporcionada de forma oportuna al usuario que tomará las decisiones, ya sean estratégicas u operativas para que estas al momento de tomarse, tengan un menor porcentaje de incertidumbre.

b) Apoyo a la decisión:

Consiste esencialmente en diagramar y mostrar la data relevante para la toma de decisiones usando herramientas, tecnologías y metodologías que permitan generar información clara, concisa, útil y relevante.

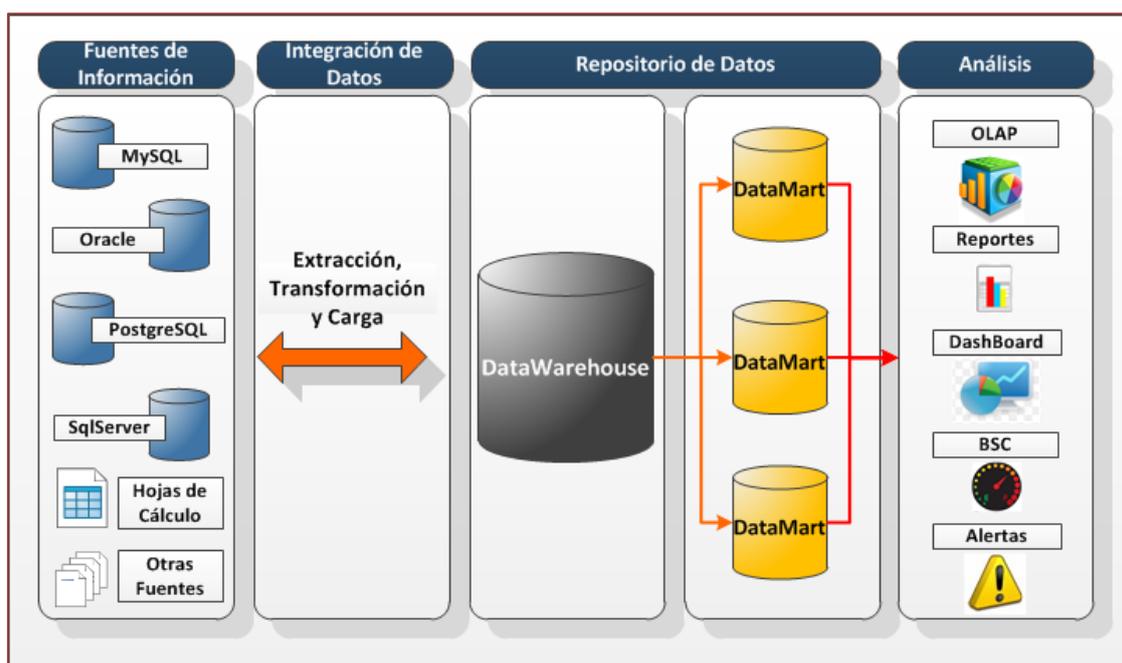
c) Orientación al usuario final:

Es común encontrar usuarios finales que carezcan de conocimientos para generar la información relevante o que tengan la capacidad de acceder a ella, por lo cual siempre se necesitaba al equipo de TI de intermediario para obtener la información, con BI se elimina este intermediario y esta dependencia de tal manera que el usuario final puede interactuar directamente con la información.

2.2.6 Componentes de una solución de inteligencia de negocios

Teniendo en cuenta el diagrama de (Cano, 2007), el Business Intelligence se compone de fuentes de información, el proceso ETL, el Data Warehouse, los Datamarts, el motor OLAP y diferentes formas de explotación de datos (gráficos, vistas, exploración de datos, etc.)

Ilustración I - Componentes de una solución de inteligencia de negocios



Elaboración: Los autores

a) Fuentes de información

Suelen ser datos generados por los sistemas operativos, que sirven de input para el Data Warehouse.

- Sistemas operativos: con aplicaciones ad-hoc.
- Sistemas de información por sector: financiera, presupuesto, planillas, etc. o fuentes de información externas, como estudios de mercado, base de datos de clientes, etc.

b) Proceso ETL

Es el proceso de extracción, transformación y carga de datos en el *Data Warehouse*. Antes de ser archivados en un *Data Warehouse*, los datos pasan por procesos de filtrado, limpieza, transformación y redefinición en un entorno de *staging* de Zona específico.

El proceso Extraer, Transformar y Cargar, también llamado ETL, es el proceso de carga y actualización de los datos obtenidos de las distintas fuentes de información presentes en el *Data Warehouse*, (Kimball Ralph, 2004) y es un proceso fundamental en la vida de cualquier proyecto de esta naturaleza.

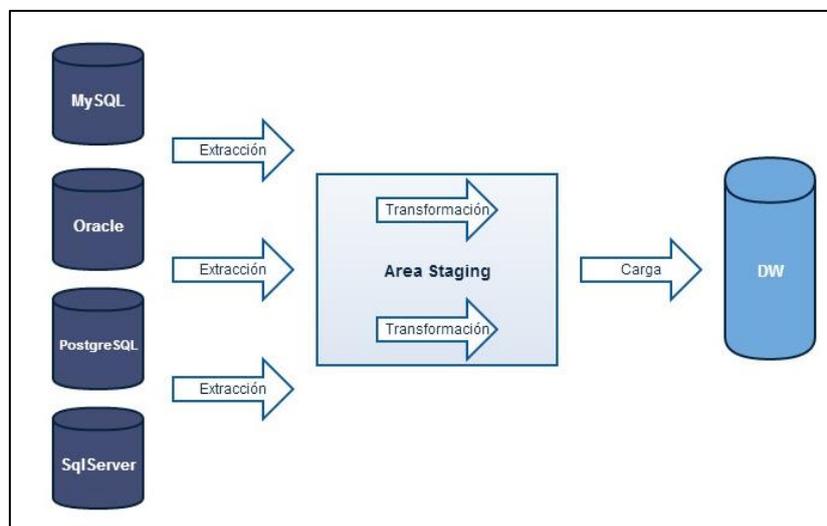
Esta fase del proceso de construcción del almacén de datos es costosa e implica una inversión considerable de tiempo y esfuerzo para completarla, por lo que requiere recursos, estrategia, habilidades y tecnologías especializadas.

El proceso ETL se divide en cinco subprocesos:

- **Extracción:** en esta primera parte se extraen físicamente los datos de diferentes bases que contengan la información transaccional.
- **Depuración:** en este subproceso se toman los datos en bruto y se comprueba su calidad, aquí se limpian los datos duplicados y, si fuera posible, se corrige cualquier valor que se encuentre incorrecto o que requiera ser completado, es decir, que los datos se transforman, todo con el fin de reducir los errores de carga. Cuando se termina el proceso se tendrán datos claros y de alta calidad.

- **Transformación:** Este paso hace uso de los datos limpios del proceso anterior y los empieza a estructurar en los diferentes modelos de análisis, obteniendo así datos coherentes y útiles.
- **Integración:** aquí se valida la información cargada en el Data Warehouse, se realiza un análisis para verificar si cumple con las definiciones y formatos, y es aquí donde se integra a los diferentes modelos de las distintas áreas de negocio.
- **Actualización:** Es el subproceso donde se agrega nueva data al Data Warehouse, para mantenerlo siempre actualizado.

Ilustración II – Proceso ETL



Elaboración: Los autores

c) Data Warehouse

Este es el repositorio donde se encuentran los datos para tener facilidad de acceso y una correcta administración y así generar reportes y explotar la información con el fin de soportar al negocio.

(Kimball Ralph, 2004) Define el almacén de datos como “una copia de transacciones de datos estructurada específicamente para consulta y análisis”

El *Data Warehouse* es la herramienta que surge como respuesta a las necesidades de los niveles de gestión de los usuarios, que requieren niveles consistentes,

integrados información histórica lista para ser analizada y utilizada para mejorar el proceso de toma de decisiones

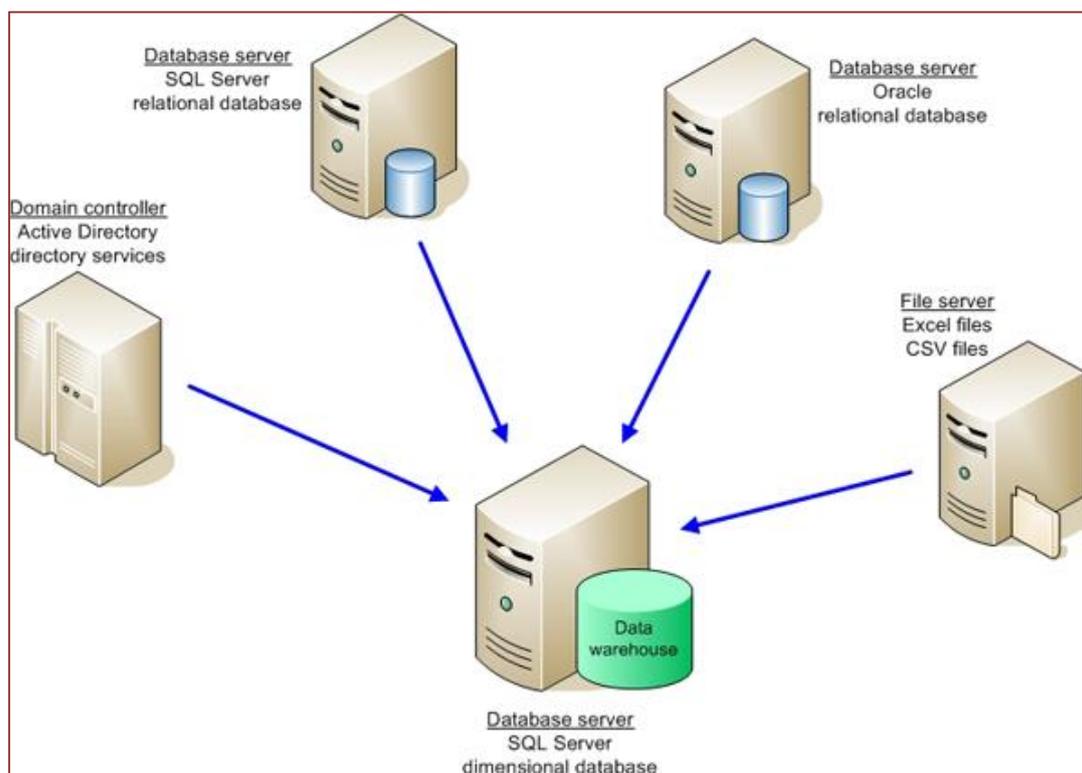
El hecho de recuperar información de los diversos sistemas con los que cuenta la empresa, ya sean transaccionales o externos, para luego archivarlos en un ambiente de información coherente, como un *Data Warehouse*, permitirá el análisis contextual y la correlación de la información dentro de la compañía.

Las características que debe cumplir un almacén de datos son:

- Temático
- Embebido
- Histórico
- No volátil

El proceso de diseño y construcción de un almacén de datos empresarial es generalmente costoso, además de requerir tiempos que las organizaciones no quieren aceptar, estas situaciones han dado lugar al nacimiento de *Datamart*, centrado en un área particular de usuarios dentro de la organización, que muy bien puede estar compuesta por miembros de un servicio o usuarios de un determinado nivel organizativo o por un grupo de trabajo multidisciplinar con objetivos comunes.

Ilustración III – Data Waterhouse



Elaboración: Los autores

d) Los Datamarts

Contienen la información relacionada con un determinado dominio, tema o función, con los datos relevantes provenientes de las diferentes aplicaciones.

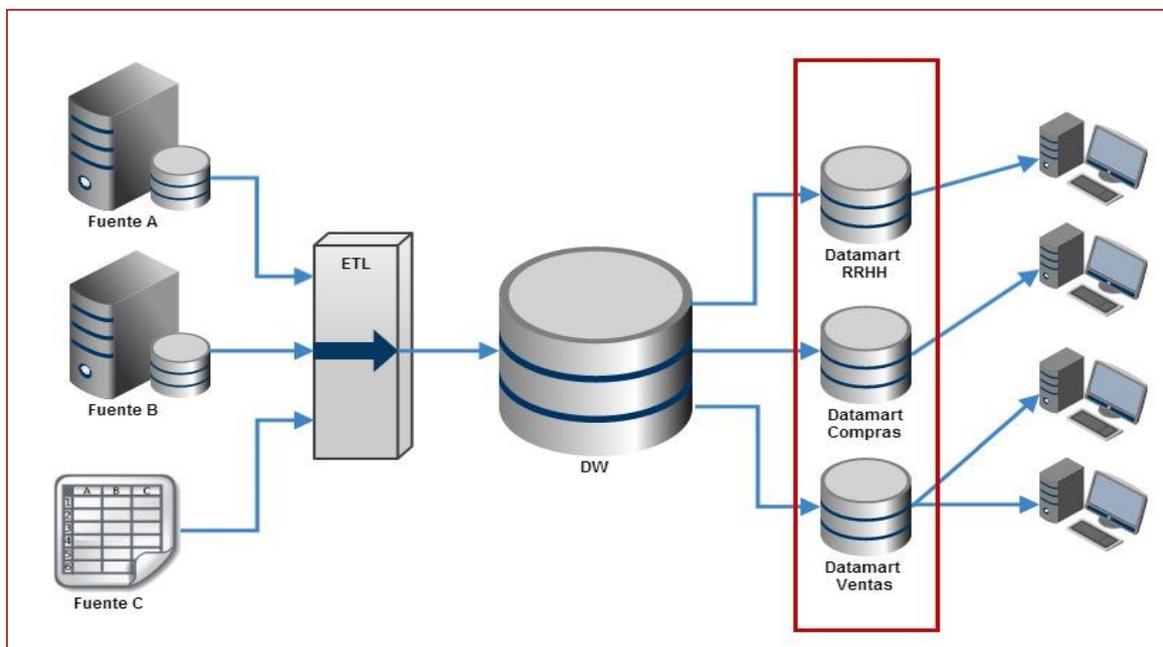
Los *datamarts* se utilizan para almacenar información para un grupo particular de áreas cuyos flujos de información coinciden: por ejemplo, marketing y ventas o producción, generalmente definidas para satisfacer necesidades específicas, usos muy específicos y concretos.

En general, los *datamarts* son más pequeños que los almacenes de datos y también almacenan menos información, menos modelos comerciales, ya que son utilizados por menos usuarios.

Según la definición de (Inmon, 2005) existen dos tipos de *Datamart*, estos pueden ser independientes o dependientes: los independientes se alimentan directamente de las fuentes de información, mientras que los dependientes obtienen la

información del *Data Warehouse* de la empresa. podrían tener problemas en su evolución, ya que podrían generar inconsistencias con otros *Datamarts*.

Ilustración IV: Datamart



Elaboración: Los autores

e) El motor OLAP

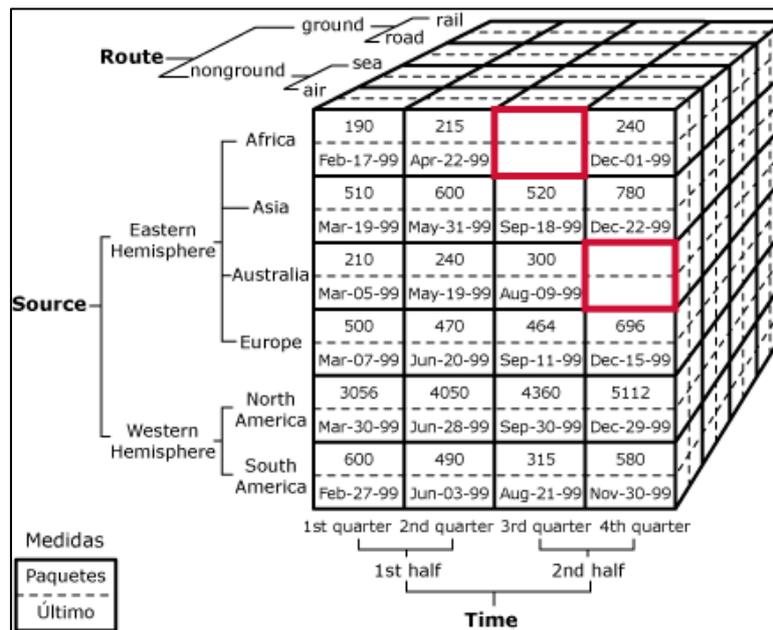
Varias tecnologías nos permiten disfrutar y analizar la información almacenada en un almacén de datos. Los cubos OLAP le permiten hacer esto para aprovechar al máximo las posibilidades que ofrecen los almacenes de datos.

Este tipo de análisis se denominan multidimensionales porque permiten analizar un evento en particular desde diferentes dimensiones, que es la mejor manera para que los tomadores de decisiones analicen la información, ya que los modelos de negocio generalmente son multidimensionales.

Las herramientas utilizadas para mostrar la información son completamente independientes de cómo fue almacenada. Las formas de acceso a las herramientas OLAP son:

- **Cliente / Servidor:** significa tener instalada en los equipos cliente la aplicación que se conectará al servidor donde se almacena la información generada por los cubos OLAP.
- **Acceso web:** En este tipo de acceso, el navegador es el que se comunica con un servidor web, el cual se comunica con el servidor OLAP.

Ilustración V – Estructura cubo OLAP



Fuente: (Microsoft Corporation, 2014)

f) Balanced scorecard o dashboard

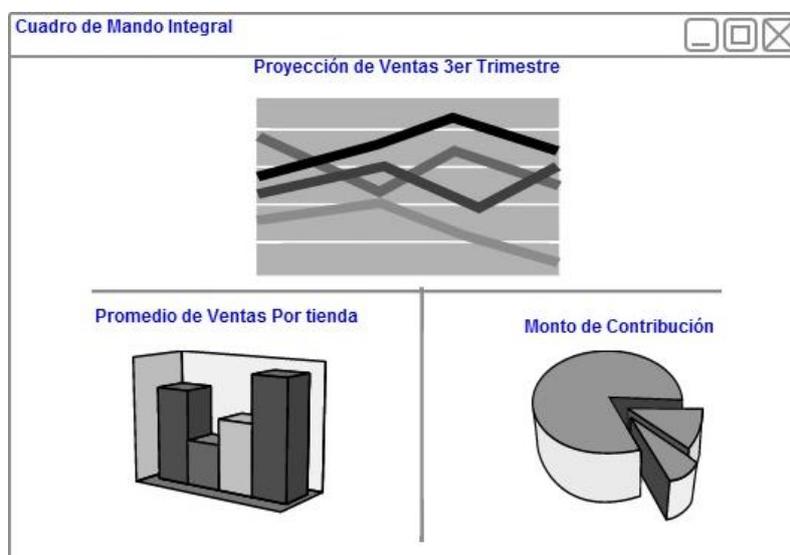
Panel de control donde se reflejan los principales indicadores de la empresa. Es una herramienta de informes utilizada por las empresas para rastrear o monitorear los factores clave que contribuyen a la operación o el éxito de su negocio.

Es importante que los datos y métricas reflejados en el tablero estén relacionados con el crecimiento de la organización, el tablero muestra la información de forma gráfica y visual, consolidándola de manera sintética, entendiendo si se logran los objetivos de la empresa. Los indicadores clave de desempeño (KPI) deben estar cercanos y bien identificados con el objetivo que representan.

Un tablero permite el monitoreo regular de la actividad y, dependiendo de su frecuencia, ofrece información sobre ciertos problemas de un sitio que permitirán

acciones correctivas. El objetivo final de un tablero es capturar una gran cantidad de información, incluida información de varias fuentes, en un espacio reducido, y que las personas que lo vean lo entiendan y puedan tomar decisiones. Los paneles mantienen informados a los usuarios sobre lo que sucede y les permiten reconocer rápidamente lo que necesita su atención. Pero ahí tienen que quedarse, no tienen que darnos toda la información que necesitamos para abordar problemas o aprovechar oportunidades que se presenten, solo tienen que mostrarnos que hay problemas u oportunidades, y si queremos ir Además, tenemos que ir a otro sitio para obtener la información, no al tablero.

Ilustración VI – DashBoard



Elaboración: los autores

g) Reportes OLAP

Son reportes dinámicos que le permite al usuario un control y flexibilidad sin precedentes, los reportes OLAP permiten tener reportes de manera inmediata, con alta disponibilidad de información, estos reportes pueden ser distribuidos eficientemente a los consumidores de información a través de la Web o mediante correo electrónico. Estas facilidades permiten compartir información importante, lo que les da a los usuarios el poder de tomar decisiones Ágiles. También pueden existir reportes predefinidos evitando la duplicación de esfuerzo.

2.2.7 Sistemas OLAP

La explotación de sistemas DataWarehouse a través de datos obtenidos directamente de sistemas transaccionales, *Online Transaction Processing* (OLTP), se basa fundamental y básicamente en estructuras agrupadas o información previamente pre calculada y procesada. La información reportada está compuesta y gestionada desde conceptos basados en datos agregados y coeficientes de gestión, que los cuadros directivos de la organización pueden definir y consultar según las dimensiones de negocio que se definan o el área a la que pertenezca. (Universidad de Cordova, 2010)

Ilustración VII - Interacción entre sistemas OLTP y OLAP



Elaboración: los autores

La implementación de un sistema de BI implica equilibrar dos extremos: la implementación independiente de las unidades organizativas y la arquitectura de almacenamiento de datos definida por el departamento de TI. Por un lado, la TI del servicio debe aceptar que no puede proporcionar información al usuario final en caso de una solicitud de emergencia sin integrar los sistemas de procesamiento analítico en línea (OLAP), por otro lado, los gerentes deben aprender a reconocer el interés de disponer de un “repositorio común” desde el que leer todos los indicadores e información terminológica que cruzan todos los datos de todos los departamentos. Solo un almacén de datos puede garantizar esta coherencia.

La información se gestiona y procesa en grandes bloques organizativos, como la estructura geográfica o académica, denominadas dimensiones, estas dimensiones empresariales a su vez se estructuran en diferentes niveles de detalle (por ejemplo, la dimensión geográfica puede estar constituida por país, comunidad autónoma, provincia, población y niveles de código postal).

Los sistemas de apoyo a la toma de decisiones que utilizan tecnologías de almacenamiento de datos se denominan sistemas OLAP. En general, estos sistemas deben:

- Soportar requisitos de análisis complejos
- Analizar datos desde diferentes ángulos
- Tener en cuenta admite análisis complejos contra un gran volumen de datos

Las principales características de los sistemas OLAP son entornos especialmente diseñados para realizar análisis multidimensionales de datos empresariales, que soportan amigablemente el análisis de cualquier usuario, así como las posibilidades de navegación, la selección de la información a obtener, permitiendo el análisis de datos segmentados y que permiten reducir el conjunto de datos reportados. Este tipo de selección da como resultado la visualización de la estructura multidimensional, a través de campos de selección que permiten elegir el nivel de agregación (jerarquía) de la dimensión, y/o la elección de datos específicos, que puede realizar el Will, entre otras cosas, las acciones de rotación, reducción de atributos, navegación, expansión o compresión de los datos visualizados.

2.2.8 Esquema MOLAP

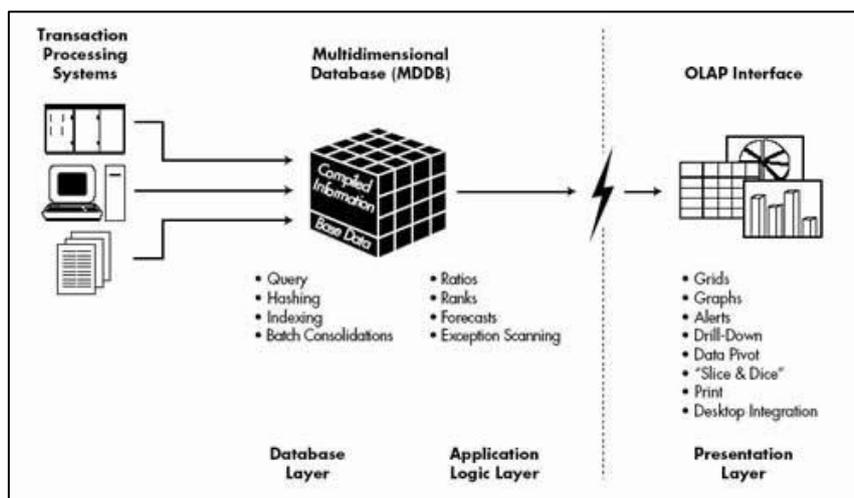
En el tipo de archivo, MOLAP se basa, con fines analíticos, en bases de datos multidimensionales. La suposición principal es ser el entorno OLAP mejor implementado y adecuado para archivar y administrar datos multidimensionales. La arquitectura y el entorno ROLAP asumen que las capacidades OLAP están completamente implementadas y reflejado en bases de datos relacionales.

Un sistema MOLAP utiliza una base de datos multidimensional, en la que la información se almacena multidimensionalmente, para ser visualizada de forma la misma forma. El sistema MOLAP utiliza una arquitectura de dos niveles: la base de datos

multidimensional que es la encargada de gestionar, acceder y obtener datos; y el motor de análisis (nivel de aplicación y presentación)

- El nivel de aplicación es responsable de cumplir con los requisitos de OLAP. El nivel de presentación se integra con el nivel de aplicación y proporciona una interfaz a través de la cual los usuarios finales ven los análisis de OLAP. Una arquitectura de cliente / servidor permite a diferentes usuarios acceder a la misma base de datos multidimensional.

Ilustración VIII - Arquitectura MOLAP



Fuente: (Universidad de Drexel, 2011)

La información procedente de los sistemas transaccionales se carga en el sistema MOLAP, mediante una serie de rutinas o procedimientos batch. Una vez cargado el dato elemental en la Base de Datos multidimensional (MDDB), se realizan una serie de cálculos en batch, para obtener los datos agregados, a través de las dimensiones de negocio, rellenando la estructura MDDB y generando los cubos multidimensionales que darán cobertura a los informes implementados en el nivel de aplicación.

Tras cargar esta estructura, se generan unos índices y algoritmos de tablas hash para mejorar los tiempos de acceso a las consultas. Una vez que el proceso de compilación ha finalizado, la MDDB está lista para su uso. Los usuarios solicitan informes a través de la interfaz, y la lógica de aplicación de la MDDB obtiene el dato.

La arquitectura MOLAP requiere unos cálculos intensivos de compilación. Lee de datos pre compilados, y tiene capacidades limitadas de crear agregaciones dinámicamente o de hallar ratios que no se hayan pre calculados y almacenados previamente.

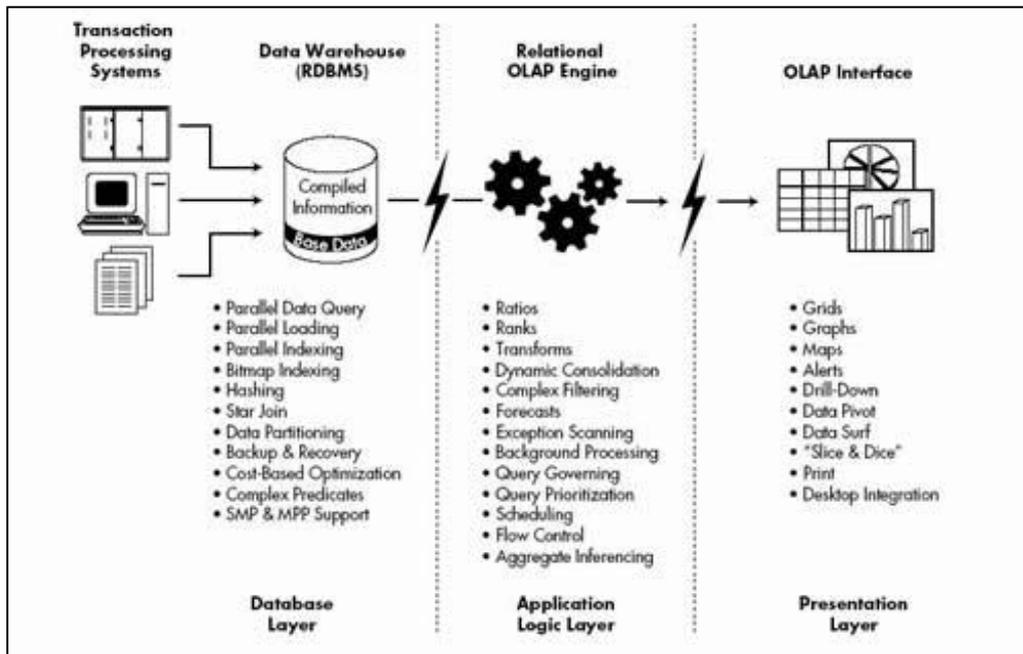
2.2.9 Esquema ROLAP

En el tipo de almacenamiento ROLAP, el sistema accede directamente a los datos almacenados en un DataWarehouse para proporcionar los análisis OLAP requeridos. El principio de estos sistemas es que las funcionalidades OLAP están mejor soportadas que las bases de datos relacionales, en lugar de implementarlas directamente en la base de datos (como en los entornos MOLAP). La esencia de estos entornos es que las acciones de filtrado y agregación equivalen a incluir una cláusula "WHERE" en una instrucción SQL.

El sistema ROLAP utiliza una arquitectura de tres niveles. La base de datos relacional gestiona los requisitos de almacenamiento de datos y el motor ROLAP proporciona funcionalidad analítica:

- El nivel de la base de datos utiliza bases de datos relacionales para la gestión, el acceso y la recuperación de datos.
- La aplicación del nivel de datos es el motor que ejecuta consultas de usuarios multidimensionales.
- ROLAP se integra con la presentación en capas, a través de la cual los usuarios realizan análisis OLAP.

Ilustración IX - Arquitectura ROLAP



Fuente: (Universidad de Drexel, 2011)

Después de definir el modelo de datos para el almacén de datos, los datos se cargan desde el sistema transaccional. Las rutinas de la base de datos se ejecutan para agregar datos si los modelos de datos lo requieren. Índice para optimizar los tiempos de acceso a las consultas. Los usuarios finales realizan sus análisis multidimensionales utilizando el motor ROLAP, que transforma dinámicamente las consultas en consultas SQL. Estas consultas SQL se ejecutan en bases de datos relacionales y sus resultados se correlacionan mediante tabulaciones cruzadas y conjuntos multidimensionales para devolver los resultados a los usuarios.

La arquitectura ROLAP puede utilizar datos pre computados, si están disponibles, o generar dinámicamente resultados a partir de datos elementales, si es necesario. Esta arquitectura accede a los datos directamente desde el almacén de datos y admite técnicas de optimización de acceso para acelerar las consultas, incluida la partición de datos a nivel de aplicación, compatibilidad con la desnormalización y uniones múltiples entre tablas.

2.2.10 Base de datos multidimensionales

Las bases de datos multidimensionales son aquellas diseñadas para desarrollar aplicaciones muy específicas, como la creación de cubos OLAP.

Básicamente, no son muy diferentes de las bases de datos relacionales (una tabla en una base de datos relacional también podría ser una tabla en una base de datos multidimensional), la diferencia es más a nivel conceptual; En las bases de datos multidimensionales, los campos o atributos de una tabla pueden ser de dos tipos, representan las dimensiones de la tabla o representan las métricas a estudiar. (Arroyo, 2009)

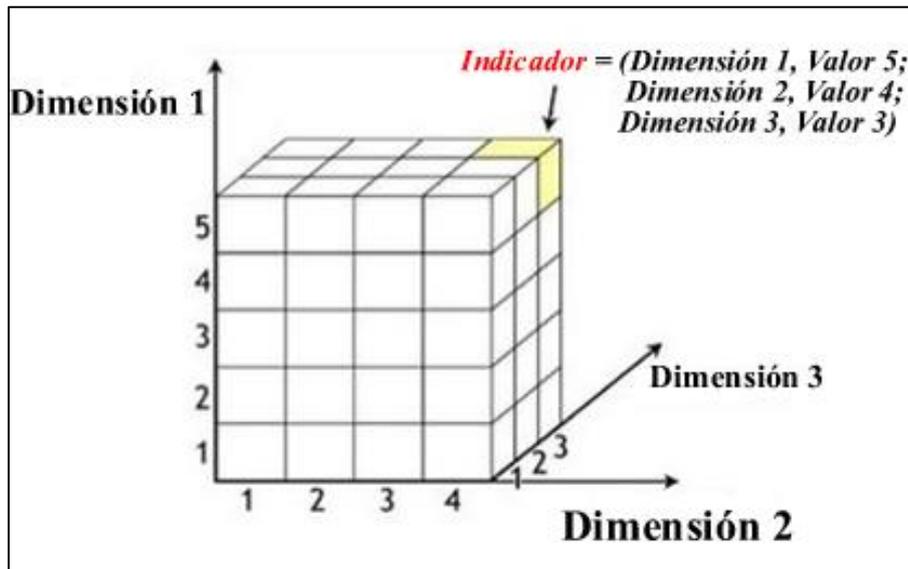
El modelado dimensional es una técnica de modelado de bases de datos sencilla y comprensible para el usuario final, la idea básica es que el usuario visualice fácilmente la relación que existe entre los diferentes componentes del modelo.

Estos tipos de bases de datos proporcionan una estructura que permite un acceso flexible a los datos, para explorar y analizar sus relaciones y los resultados resultantes. Estos pueden considerarse como un cubo multidimensional, donde existen variables relacionadas a lo largo de varios ejes o dimensiones, y la intersección de estas representa la medida, el indicador o el hecho que se evalúa.

Este modelo de datos se compone principalmente de dos tipos de elementos:

- **Tablas dimensionales:** Representan factores por los cuales se analiza un área específica del negocio, son pequeñas y generalmente desnormalizadas.
- **Tablas de hechos:** son objeto de análisis y están relacionadas con dimensiones, son tablas muy grandes y generalmente desnormalizadas, suelen incluir diferentes agregaciones como máximo, mínimo, promedio, etc.

Ilustración X - Cubo de información



Elaboración: los autores

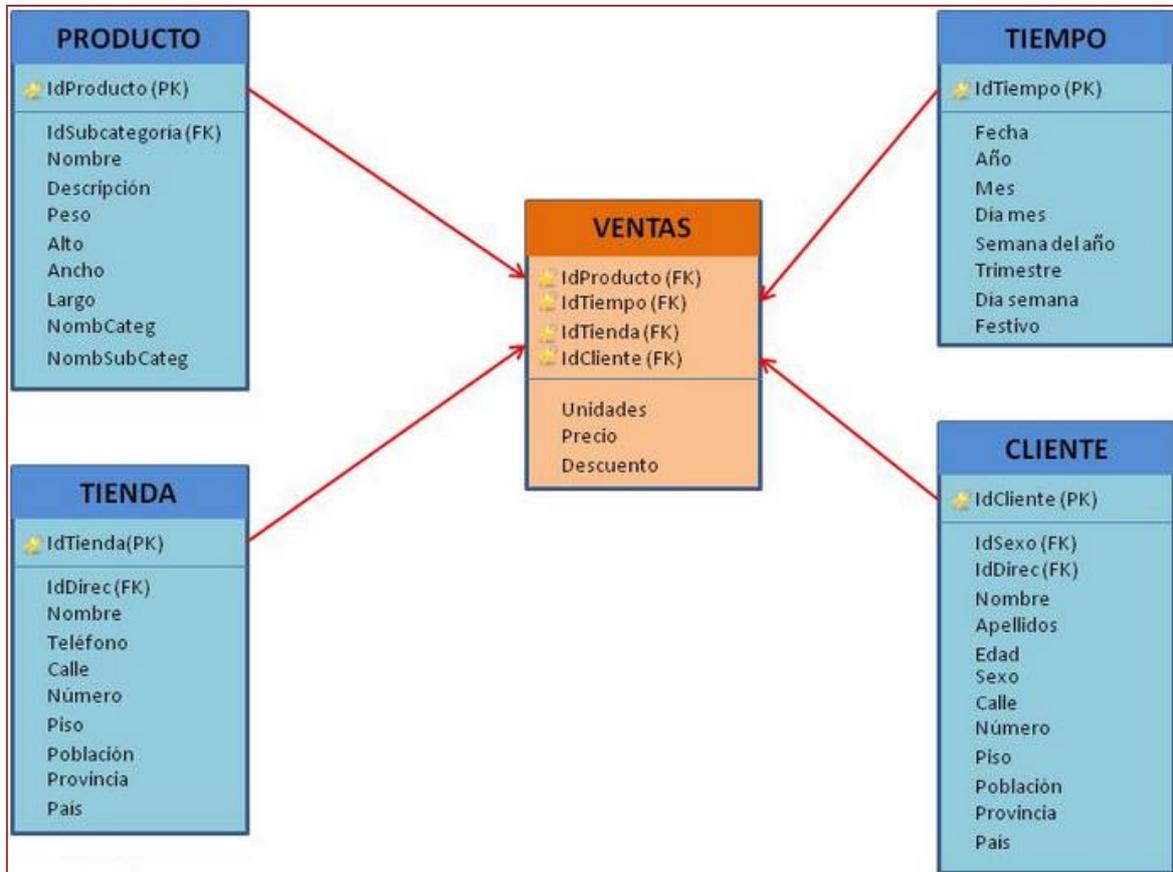
Las bases de datos multidimensionales involucran dos posibles variantes de modelado, que permiten realizar consultas de apoyo a la decisión, que son el modelo estrella y el modelo copo de nieve.

A continuación, podemos ver las definiciones de los 2 modelos acompañadas de un ejemplo que consta de la misma tabla de hechos (ventas) pero representa principalmente la diferencia entre ellos en las tablas de dimensiones a la hora de elegir aplicar cada una.

a) Modelo en estrella (Star Scheme)

Se define cuando una “Tabla de hechos” se asocia con diferentes “Tablas de dimensiones” para agilizar la búsqueda:

Ilustración XI - Modelo estrella

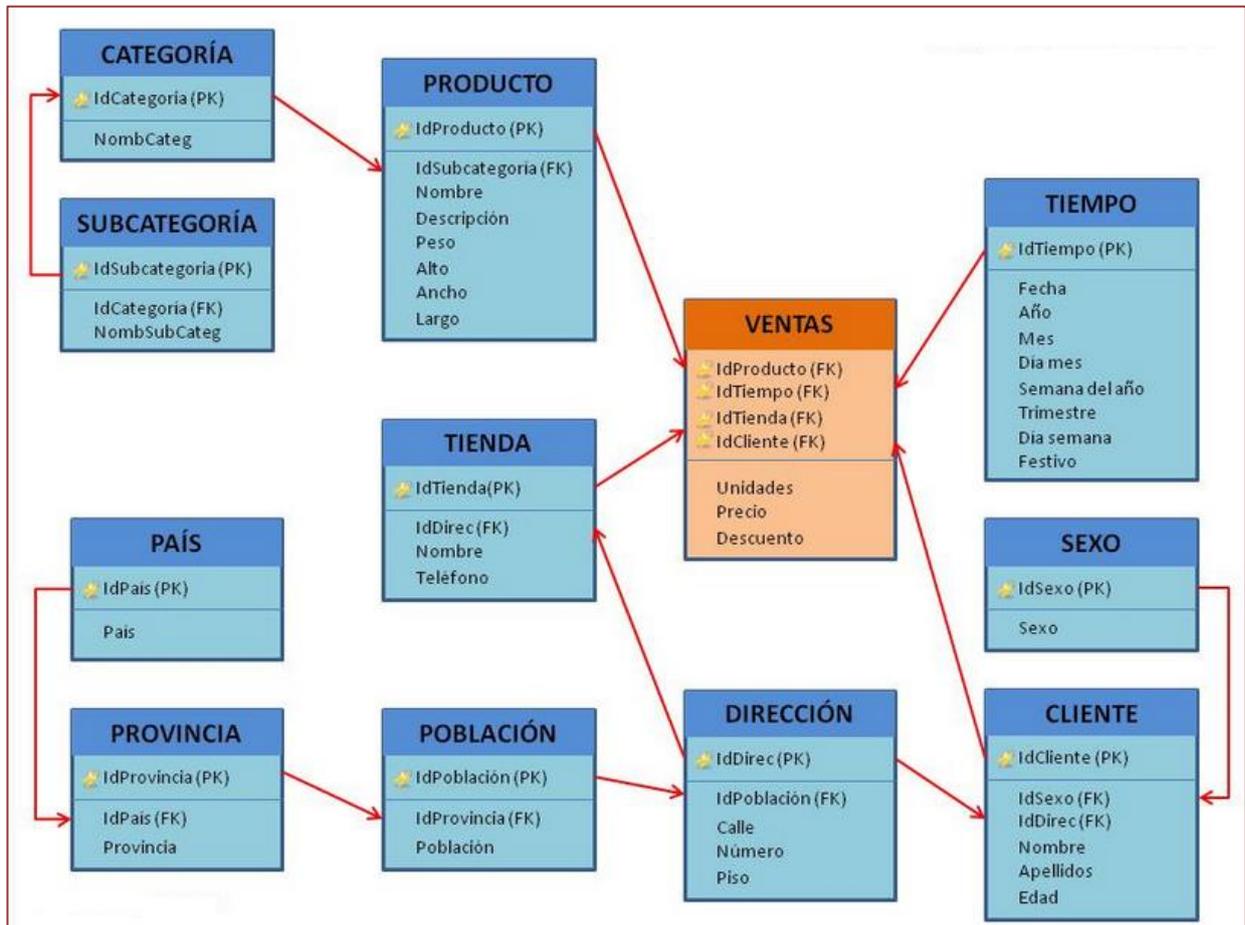


Fuente: (Politecnico Gran Colombiano, 2012)

b) Modelo copo de nieve (*Snowflake Scheme*)

Las dimensiones se pueden “romper” en sus subcomponentes, generando un número elevado de éstas; es entonces cuando se sugiere el modelo “Copo de Nieve”:

Ilustración XII - Modelo copo de nieve



Fuente: (Politecnico Gran Colombiano, 2012)

2.2.11 Metodología Ralph Kimball

Determina que un Data Warehouse es la unión de todos los Datamarts de una organización (BottomUp Approach). (Kimball Ralph, 2004), establece unos procesos para llevar a cabo el desarrollo e implementación de un Data Warehouse cuyo desarrollo incluye varias actividades que pueden realizarse en paralelo o secuencialmente.

Para concretar la realización de un almacén de datos, se deben considerar las siguientes fases que abarcan el ciclo de vida de la metodología de Ralph Kimball:

a) Planificación del proyecto

La planificación intenta identificar la definición y el alcance del desarrollo e implementación de un almacén de datos, incluidos los negocios justificaciones y evaluaciones de viabilidad La planificación de actividades se centra en recursos, perfiles, actividades, duración y secuencia.

b) Definición de requisitos comerciales

Los usuarios finales y sus requisitos siempre afectan las implementaciones realizadas en un almacén de datos. Desde la perspectiva de Kimball, los requisitos comerciales se ubican en el centro del universo de datos del almacén. "Externamente, las necesidades comerciales deben determinar el alcance del almacén de datos (qué datos debe contener, cómo debe organizarse, con qué frecuencia debe actualizarse, quién accederá a él y desde dónde, etc.) Kimball brinda consejos y técnicas para descubrir con eficacia las necesidades del negocio. Estas tácticas y estrategias se enfocan en entrevistas de investigación (diferentes tipos, preparación de entrevistas, roles a cumplir).

c) Modelo dimensional

La definición de las necesidades del negocio determina los datos necesarios para satisfacer las necesidades analíticas de los usuarios. El diseño de modelos de datos para soportar estos análisis requiere un enfoque diferente al utilizado en los sistemas operativos.

d) Diseño físico

El diseño físico de la base de datos se enfoca en seleccionar las estructuras necesarias para soportar el diseño lógico.

e) Diseño y desarrollo de ETL:

Los procesos de transformación se definen como los procesos de conversión o recodificación de datos de origen para realizar la carga real del modelo físico, mientras que los procesos de carga de datos son los procesos necesarios para poblar el almacén de datos

f) Diseño de ingeniería arquitectónica

Los entornos de almacenamiento de datos requieren la integración de muchas tecnologías, se deben considerar tres factores: requisitos comerciales, entornos técnicos actuales y direcciones técnicas estratégicas futuras anticipadas.

g) Selección de productos de instalación:

Tomando como marco la arquitectura técnica, es necesario evaluar y seleccionar componentes específicos de la arquitectura como la plataforma hardware, el motor base de datos, la herramienta ETL o su desarrollo.

h) Diseño y desarrollo de aplicaciones BI

No todos los usuarios del almacén de datos necesitan el mismo nivel de análisis, por lo que en esta fase se identifican los diferentes roles o perfiles de usuario para determinar los diferentes tipos de aplicaciones necesarias según el alcance de los diferentes perfiles (gerente, analista de negocio, proveedor, etc.)

i) Implementación (Deployment)

La implementación representa la convergencia de la tecnología. datos y aplicaciones del usuario final accesibles desde el escritorio del usuario comercial. Varios factores adicionales aseguran que todo esto funcione, incluida la capacitación, el soporte técnico y la comunicación.

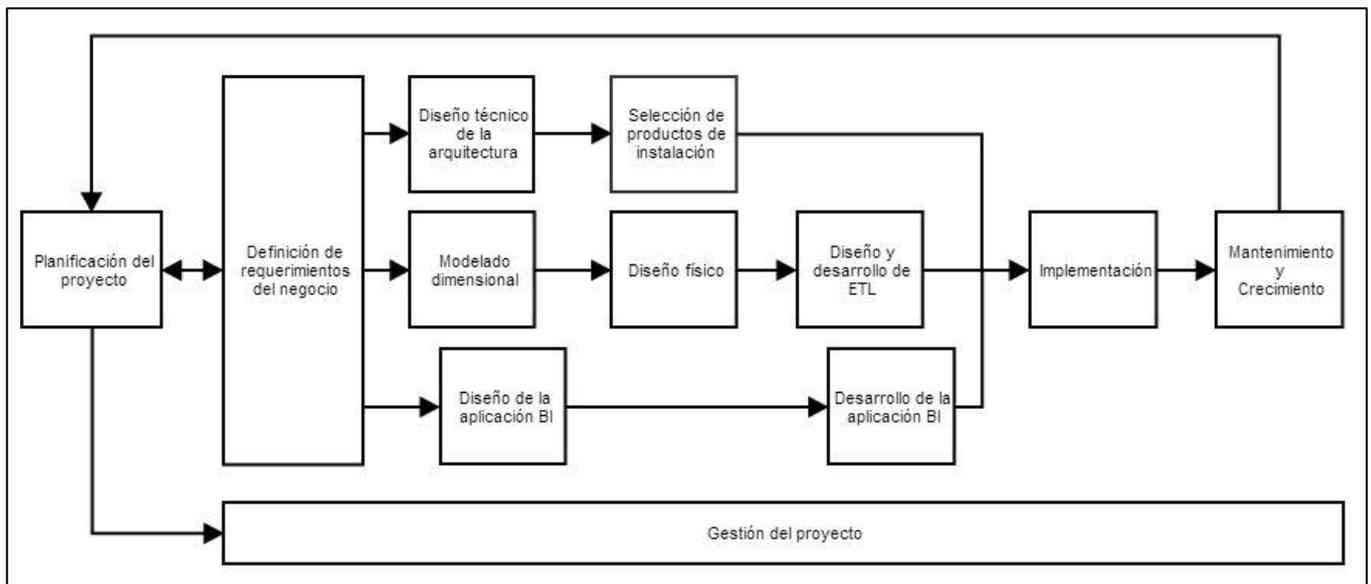
j) Mantenimiento y crecimiento:

El Data Warehouse es un proceso (con fases bien definidas, con un principio y un final, pero de carácter espiral) porque acompaña la evolución de la organización a lo largo de su historia, es necesario continuar la investigación con el fin de monitorear constantemente la evolución de los objetivos a alcanzar, se especifica que: “si se ha utilizado el ciclo de vida dimensional de la empresa, el Data Warehouse está listo para evolucionar y crecer”.

k) Gestión de proyectos

La gestión de proyectos asegura que las actividades del ciclo de vida dimensional de la empresa se realicen de manera sincronizada, como se muestra en el diagrama, la gestión acompaña todo el ciclo de vida. Sus principales actividades son el monitoreo del estado del proyecto y la comunicación entre los requisitos de la compañía y las restricciones de información para gestionar adecuadamente las expectativas en ambas direcciones.

Ilustración XIII – Modelo propuesto por Ralph Kimball



Fuente: (Kimball Ralph, 2004)

2.2.12 Metodología Bill Inmon

Para (Inmon, 2005), los Datamarts se crearán después de completar el desarrollo del almacén de datos completo de la organización (enfoque TopDown). El diseño de un almacén de datos comienza con la carga de datos, la eficiencia de los sistemas de acceso a datos depende del tamaño de esta carga útil.

Inmon a menudo se asocia con almacenes de datos de toda la empresa, que involucran todo el entorno empresarial desde el principio, sin centrarse en un incremento específico hasta la finalización del desarrollo del almacén de datos. En su concepto, un Datamart es solo uno de los niveles. del Data Warehouse y los Datamarts dependen del datawarehouse central, por lo que se construyen con posterioridad a este para evitar la aparición de situaciones imprevistas que puedan poner en peligro el proyecto, lo que permite concentrar los esfuerzos del desarrollo en curso para que sea compatible con los subsiguientes.

2.2.13 Herramientas ETL y de explotación de información

a) Pentaho BI Suite

Pentaho, creada en 2004, es el líder actual en soluciones Open Source Business Intelligence. Ofrece, con sus soluciones, el espectro completo de recursos para desarrollar, mantener y operar un proyecto BID, desde ETL con integración de datos hasta tableros con Dashboard Designer o marco de tablero de comunidad.

Cómo Pentaho construyó su B.I. integra varios proyectos existentes y de reconocida solvencia. Anteriormente, Data Integration era Kettle, de hecho, aún conserva su antiguo nombre como un nombre familiar. Mondrian es el otro componente de Pentaho que continúa manteniendo su propia entidad.

La plataforma de inteligencia empresarial de código abierto de Pentaho cubre una amplia gama de necesidades de informes y análisis de datos empresariales. Las soluciones de Pentaho están escritas en Java y cuentan con un entorno de implementación basado en Java que también es muy flexible para cubrir una amplia gama de necesidades empresariales, tanto típico y sofisticado y específico de la empresa. (Pentaho Open Source Business Intelligence, 2012)

La plataforma Pentaho cuenta con los siguientes productos:

- **Pentaho Data Integration (anteriormente Kettle):** Esta es una de las soluciones ETL más conocidas y populares del mercado, tiene una larga trayectoria y una solidez y robustez que la convierte en una herramienta muy recomendable, permite transformaciones y un manejo muy simple y manera intuitiva de trabajar. Asimismo, los proyectos realizados con integración de datos son muy fáciles de mantener. Con volúmenes y una variedad y velocidad de datos, las organizaciones necesitan formas rápidas y fáciles de aprovechar los datos y aprender de ellos. Sin embargo, uno de los mayores desafíos que enfrentan las organizaciones de TI en la actualidad es proporcionar una versión única y coherente de la verdad sobre todas las fuentes de información en un formato listo para el análisis. (Pentaho Open Source Business Intelligence, 2012)
- **Pentaho Business Analytics:** es una solución integral que incluye capacidades de inteligencia comercial, integración de datos y minería de datos para capacitar a los tecnólogos y comprender rápidamente a los usuarios. Pentaho Business Analytics permite a los usuarios comerciales acceder, descubrir y analizar de forma intuitiva sus datos, lo que les permite convertir los conocimientos en decisiones que impactan positivamente en el rendimiento de la organización. Proporciona una interfaz web interactiva fácil de usar para que los usuarios empresariales accedan a datos, creen e interactúen con informes y paneles, y analicen datos en múltiples dimensiones, sin depender de TI o desarrolladores. plataforma liviana y de alto rendimiento y se puede implementar de manera flexible, en la nube o integrarse a la perfección con otras aplicaciones de software (Pentaho Open Source Business Intelligence, 2012)

b) Microsoft SQL server 2008 R2 (SSIS – SSAS)

Microsoft SQL Server 2008 R2 le permite crear soluciones organizacionales que son más fáciles de planificar, implementar y administrar que las versiones anteriores de Microsoft SQL Server.

Esta herramienta ofrece mayor seguridad, confiabilidad y rendimiento que su antecesor MS SQL Server 2008.

MS Server 2008 R2 brinda conectividad y control a recursos locales y remotos. Esto significa que las organizaciones pueden beneficiarse de costos reducidos y mayor eficiencia a través de la gestión y el control de los recursos dentro de la empresa.

Además, funciona con SQL Server Management Studio, que es un entorno integrado para acceder, configurar, administrar y desarrollar todos los componentes de SQL Server. SQL Server Management Studio combina un gran conjunto de herramientas gráficas con un amplio conjunto de secuencias de comandos que permiten a los desarrolladores y administradores de todos los niveles para acceder a SQL Server. (Microsoft Corporation, 2014)

SQL Server Management Studio combina la funcionalidad de Enterprise Manager, Query Analyzer y Analysis Manager, herramientas incluidas en versiones anteriores de SQL Server, en un solo entorno. Además, SQL Server Management Studio funciona con todos los componentes de SQL Server.

- **Integration Services (Servicios de Integración - SSIS):** lo que permite la integración de datos provenientes de diferentes tipos de fuentes, facilitando así una visión completa de la empresa.
- **Analysis Services (Servicios de Análisis - SSAS):** permite un análisis completo de datos, preparando el marco necesario para informes, análisis OLAP, extracción de datos y desarrollo de cuadros de mando.
- **Reporting Services (Servicios de Reportes - SSRS):** permite la administración y elaboración de reportes, desde formatos predefinidos y de forma personalizada, con posibilidad de ser visualizados ya sea en PDF, en la Web o con cualquier otra herramienta de Microsoft. La herramienta desarrollada por el módulo operativo incorpora algunas de las funcionalidades de estos servicios.

Características y beneficios:

- SSIS puede extraer y transformar datos de una amplia variedad de fuentes como archivos XML, archivos planos y bases de datos relacionales para luego cargar los datos en uno o más destinos.
- SSIS posee un amplio rango de componentes como lookups, ordenamiento, agregación, combinaciones entre otros. Que son usados para el diseño del ETL.
- SSIS Provee de modo debug gráfico.
- SSAS permite un Almacenamiento en cache proactivo, además de la combinación de las actualizaciones en tiempo real con el rendimiento del uso de una arquitectura MOLAP (Procesamiento analítico en línea multidimensional). Esto permite tener actualizados la caché de datos mientras los datos se van modificando en sus orígenes de datos, lo cual permite obtener un rendimiento mayor en la realización de las consultas.
- SSAS Permite realizar análisis OLAP (Procesamiento analítico en línea). Con lo cual se accede a los datos organizados y agregados del origen de datos en una estructura multidimensional.

c) Comparación entre ambas herramientas

Se consideraron dos herramientas candidatas: la suite Pentaho BI y Microsoft SQL Server 2008 R2 (SSIS-SSAS).

Para cada requisito que debe cumplir la herramienta se ha asignado una puntuación de 0 (no cumple) a 5 (cumple completamente).

MS SQL Server 2008 R2 (SSIS-SSAS) tiene una interfaz gráfica para una fácil administración por parte de los usuarios.

Como una ventaja importante, tenemos que la herramienta de Microsoft cuenta con soporte técnico completo y documentación que respalda la herramienta en comparación con Pentaho BI Suite, que es una herramienta Open Source y no tiene soporte técnico certificado ni documentación que cubra todas las necesidades del proyecto.

A continuación, se presenta un cuadro comparativo, con las herramientas antes mencionadas, que muestra las características trascendentales de un trabajo como el que se desarrollará en esta tesis. La herramienta que obtuvo mayor puntaje, según el cuadro comparativo (Tabla I), fue SQL Server 2008 R2 (SSIS-SSAS)

Tabla I - Comparación de soluciones

Criterio	Pentaho BI Suite	SQL Server 2008 R2 (SSIS-SSAS)
Es multiplataforma	5	3
Limpieza de datos	5	5
Trabajo con múltiples fuentes de datos	4	5
Interfaz gráfica para el desarrollo del proceso ETL	5	5
Permitir el monitoreo del proceso ETL, así como el diseño de este.	4	3
Programación del proceso ETL para que se ejecute automáticamente	5	4
Trabajo con las principales bases de datos	4	5
Soporte de la herramienta.	1	5
Análisis de la información en todas las dimensiones	5	5
Generación de reportes gráficos	5	5
Exportación a archivos Excel para visualizar la información	5	5
Drill and down	4	5
Filtros y búsquedas personalizables.	4	4
Puntaje Total	56	59

Elaboración: los autores

Según el análisis realizado, las herramientas que serán utilizadas para el desarrollo de la presente tesis son las que brinda Microsoft, Microsoft SQL Server Integration

Services, Microsoft SQL Server Analysis Services y Microsoft SQL Server Reporting Services.

2.3 Definición de términos básicos

- **Área staging:** bloque de extracción, transporte, transformación, normalización, limpieza y carga de datos, es decir, el bloque responsable de la copiar de data el Data Warehouse.
- **Cuadro de mando integral (BSC):** El Cuadro de Mando Integral (BSC por sus siglas en ingles), también conocido como cuadro de mando, es una herramienta de control empresarial que ayuda a establecer y monitorear los objetivos de una empresa y sus distintas áreas o unidades.
- **Cubos de información:** subconjuntos de datos de un almacén de datos, organizados y resumidos en una estructura multidimensional. Los datos se resumen en función de factores comerciales seleccionados, lo que proporciona el mecanismo para tiempos de respuesta rápidos y consistentes a consultas complejas. Estos datos, organizados en Cubos, son la materia prima que almacena DataWare House.
- **Customer Relationship Management (CRM):** La gestión de la relación con el cliente (CRM) es un término general que describe los métodos comerciales utilizados para aprovechar la información intrínseca asociada con los clientes de una empresa. Esto incluye muchas clases de técnicas de BI que implican capturar, almacenar y analizar información del cliente.
- **Datamart:** es una base de datos departamental, especializada en el archivo de datos de un área específica de actividad, caracterizada por la estructura de datos óptima para analizar en detalle la información desde todos los puntos de vista que afecta los procesos de esta actividad. Datamart puede

alimentarse con datos de un almacén de datos o integrar una colección de diferentes fuentes de información por sí mismo.

- **Datamart OLAP:** Se basan en los populares cubos OLAP, que se construyen añadiendo, según las necesidades de cada dominio o servicio, las dimensiones e indicadores necesarios de cada cubo relacional muy heterogéneo, según la herramienta final utilizada.
- **Datamart OLTP:** pueden basarse en un simple extracto del almacén de datos, pero es habitual introducir mejoras en su rendimiento (la agregación y el filtrado suelen ser las operaciones más habituales) explotando las características propias de cada dominio de Las estructuras más habituales en este sentido son las tablas de relaciones, cuya factibilidad es reducida (que añaden las dimensiones adecuadas), y las vistas materializadas, que se construyen con la misma estructura que las anteriores, pero con el fin de aprovechar la reescritura de consultas (aunque esto solo es posible en algunos DBMS avanzados, como Oracle).
- **Data Mining (Minería de datos):** extracción de grandes volúmenes de datos para tendencias y patrones sutiles con aplicaciones de software complejas.
- **DataWarehouse (DW):** Un repositorio de datos o Data Warehouse es una base de datos corporativa que se caracteriza por la integración y depuración de información de una o más fuentes diferentes, para ser procesada posteriormente, permitiendo su análisis desde infinitas perspectivas y con gran respuesta velocidades.
- **Dimensión:** es lo que podemos utilizar para desglosar un indicador o métrica, son los datos que nos permiten filtrar, agrupar o seccionar información.
- **DrillDown o Drill Through:** es una operación para acceder a datos en cubos multidimensionales, lo que significa exponer progresivamente más detalles (dentro de un informe o consulta) y luego seleccionar elementos.

- **Enfoque BottomUp:** Es una metodología rápida basada en experimentos y prototipos. Es un método flexible que permite a la organización llegar más lejos a un menor costo. La idea es construir Datamarts independientes para evaluar los beneficios del nuevo sistema a medida que avanzamos. En él, las partes individuales se diseñan en detalle y luego se unen para formar componentes más grandes, que a su vez se unen para formar el sistema completo.
- **Enfoque TopDown:** Este es un método sistémico, que minimiza los problemas de integración, pero que es costoso, debido a la gran cantidad de datos y su falta de flexibilidad. En este método se formula un resumen del sistema, sin especificar detalles. Cada parte del sistema se refina diseñándola con más detalle. Luego, cada parte nueva se refina, cada vez con mayor detalle, hasta que la especificación completa es lo suficientemente detallada como para validar el modelo.
- **Indicador:** un indicador se utiliza para medir la evolución de una medida en un segmento de análisis.
- **Inteligencia de Negocios (BI):** La inteligencia de negocios más conocida por sus siglas en inglés BI, es la encargada del análisis de la información interna de una organización, conocida como el conjunto de estrategias y herramientas enfocadas a la administración y creación de conocimiento mediante "análisis de información existente". datos en una organización o empresa.
- **Indicador clave de desempeño (KPI):** Indicadores Clave de Desempeño: el término utilizado para indicar la información más importante que reside en una empresa. Diríamos que son estos indicadores, cálculos, informes, métricas, etc,v los que nos permiten medir los factores y aspectos críticos de un negocio.

- **Job (Trabajo):** una tarea responsable de ejecutar un conjunto de procesos ETL, que se puede programar para ejecutarse periódicamente, se denomina trabajo.
- **Medida:** Una medida es un valor que contiene, a nivel de detalle de la tabla de hechos a la que pertenece, un valor representativo. Por ejemplo, para la tabla de hechos Ventas, una métrica útil sería Cantidad vendida, que contiene la cantidad de veces que se vendió un producto a un cliente por mes.
- **Metadatos:** Los metadatos son datos que describen otros datos.
- **Modelo estrella:** es un esquema de base de datos en el que una tabla central de hechos está vinculada a tablas dimensionales relacionadas.
- **Modelo copo de nieve (Snowflake):** es un esquema de base de datos en el que una tabla de hechos central está vinculada a sus tablas dimensionales, pero estas a su vez están vinculadas a otras tablas dimensionales.
- **MOLAP:** la arquitectura MOLAP utiliza bases de datos multidimensionales para proporcionar análisis. Su principio fundamental es que OLAP se implementa mejor almacenando datos de forma multidimensional.
- **Procesamiento analítico en línea (OLAP):** Los sistemas OLAP son bases de datos orientadas al procesamiento analítico. Este análisis suele implicar la lectura de grandes cantidades de datos para extraer cierto tipo de información útil: tendencias de ventas, comportamientos de los consumidores, informes complejos, etc. El sistema es típico de Datamarts.
- **Procesamiento de transacciones en línea (OLTP):** Los sistemas OLTP son bases de datos orientadas al procesamiento de transacciones. Una transacción genera un proceso atómico (que debe ser validado por una confirmación o invalidado por una reversión) y que puede involucrar

operaciones de inserción, modificación y eliminación. datos El proceso transaccional es típico de las bases de datos operativas.

- **Proceso ETL (Extract, Transform and Load):** es el proceso por el cual los datos se extraen de la base de datos transaccional, se transforman utilizando operaciones y algoritmos complejos, y los datos ya transformados se cargan en el Data Warehouse.
- **ROLAP:** la arquitectura ROLAP accede a los datos almacenados en un almacén de datos para proporcionar análisis OLAP. La premisa de los sistemas ROLAP es que la funcionalidad OLAP está mejor soportada que las bases de datos relacionales.
- **Roll Up:** es una operación de acceso a datos en cubos multidimensionales, es el efecto contrario al drilldown, significa ver menos niveles de detalle, en la jerarquía significa generalizar o resumir, es decir, subir en el árbol jerárquico
- **Sistemas de información ejecutiva (EIS):** Un EIS es un sistema de información para gerentes que automatiza la tarea de obtener los datos más importantes de una organización, resumiéndolos y presentándolos de la manera más comprensible posible, brindándole al gerente un fácil acceso a la información. internos y externos a la empresa con el fin de monitorear los factores críticos de éxito.
- **Sistemas de soporte a la decisión (DSS):** Un Sistema de Soporte a la Decisión (DSS) es una herramienta de BI enfocada en analizar los datos de una empresa.
- **Sumarización:** Actividades para aumentar la granularidad de la información en una base de datos Resumen reduce el nivel de detalle y es muy útil para presentar datos para apoyar la toma de decisiones.
- **Tabla de hechos (Fact):** Un hecho o tabla de hechos es una entidad que contiene información integrada con cierto nivel de detalle para determinados

valores de análisis. El nivel de detalle, también llamado granularidad, está determinado por el número de dimensiones asociadas a dicho hecho. Los valores de análisis se presentan generalmente en cantidades numéricas y se denominan, según la información que contienen, medidas e indicadores. Por ejemplo, un hecho ilustrativo podría ser el hecho de ventas, relacionado con el tamaño del cliente y del producto y la dimensión de tiempo mes.

CAPÍTULO III

METODOLOGÍA

3.1 Materiales

3.1.1 Software

Las herramientas de software que se utilizaron para las diversas actividades dentro del desarrollo de la tesis se presentan a continuación.

Tabla II - Herramientas de software

Herramientas	Actividades donde se utilizan
Microsoft Office Profesional Plus 2010	Elaboración del documento de Tesis
	Elaboración de Bitácora de Revisiones.
	Elaboración del Cronograma de Actividades
	Elaboración de la presentación para la sustentación final
	Elaboración de cuestionarios de negocio.
	Elaboración de Arquitectura Técnica
CA ERwin® r7.3 Data Modeler Standard Edition	Modelamiento Dimensional
	Diseño Físico del Datamart

Microsoft SQL server 2008 R2 Integration Services (SSIS)	Diseño y elaboración y ejecución de procesos ETL
Microsoft SQL server 2008 R2 Analysis Services (SSAS)	Análisis y explotación de datos (cubos OLAP)
MS Excel	Elaboración y presentación de reportes de usuario final.
SQL Server Management Studio	Gestor de Base de Datos
Microsoft SQL server 2008 R2	Motor de base de Datos

Elaboración: los autores

3.1.2 Hardware

Debido a que la implementación del Datamart no se contempla en el alcance de la tesis, no será necesaria la utilización de hardware más que para el diseño del Datamart y para la simulación de los procesos de carga, análisis y explotación. Las herramientas detalladas anteriormente se instalarán en dos Laptops HP Pavilion Core i5 que se configurarán para dicho propósito.

3.2 Métodos

Como parte de la definición de los métodos utilizados para el desarrollo del Datamart, se vio conveniente analizar dos puntos importantes:

- Elección del Enfoque de Inteligencia de Negocios.
- Elección de la Metodología para el desarrollo del Datamart.

3.2.1 Enfoque de inteligencia de negocios

En el “Capítulo II. Marco Teórico”, se definieron los dos enfoques de Inteligencia de Negocios: el Enfoque Top-Down, presentado como el Modelo de Inmon y el Enfoque Bottom-Up, presentado como el Modelo de Kimball.

En el desarrollo de esta tesis se ha decidido utilizar el enfoque Bottom–Up, debido a que sus características son aplicadas de la siguiente manera:

- a) Los usuarios de la Oficina de Registros Académicos requieren obtener resultados rápidos. El enfoque Bottom – Up, requiere un menor tiempo de implementación porque se centra específicamente en la implementación del datamart que atiende los requerimientos de un área específica de la facultad.
- b) La presente tesis pretende realizar un modelo para la implementación de un Datamart para la Oficina de Registros Académicos y como un trabajo futuro se puede realizar la implementación de un Data Warehouse para la facultad. Esta es una gran ventaja que presenta el Enfoque Bottom – Up, debido a que plantea que las implementaciones se realicen por áreas temáticas, por Datamarts. Los cuales constituyen el Data Warehouse.
- c) El tiempo con el que se cuenta para desarrollar la tesis es corto (solo cuatro meses). El Enfoque Bottom – Up conduce a una solución completa en una cantidad de tiempo relativamente pequeña. Permite que, partiendo de cero, podamos empezar a obtener información útil en cuestión de días y después de los prototipos iniciales, comenzar el ciclo de vida normal que nos ofrezca una solución completa de BI.
- d) El volumen de información que se manejará en el desarrollo de la presente tesis es relativamente bajo. El Enfoque Bottom – Up, es ideal para dar los primeros pasos en la implementación de un Datamart, cuando la complejidad del almacenamiento de datos no es demasiado grande y donde la infraestructura del BI se encarga de los datos procedentes de un número limitado de fuentes.

3.2.2 Metodología para el desarrollo del datamart

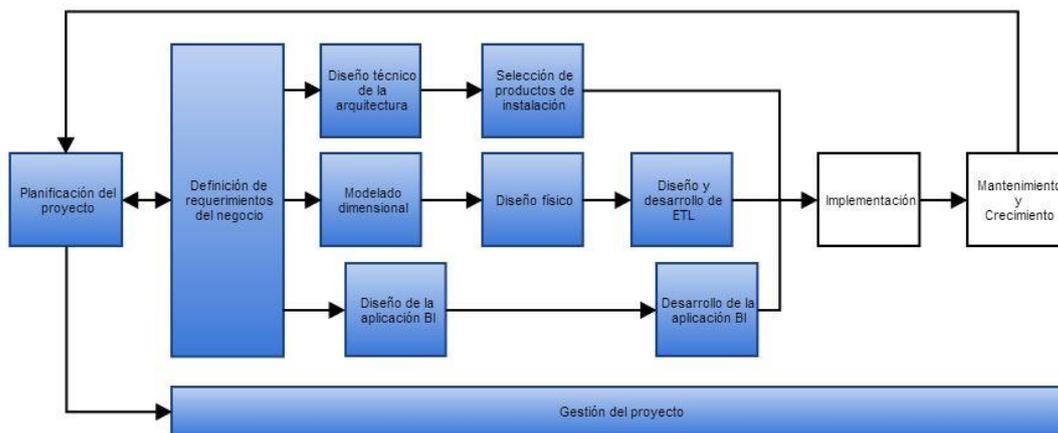
A continuación, se presenta la metodología que guía el proceso de Desarrollo del Datamart para la Oficina de Registros Académicos. Teniendo en cuenta la información de los apartados anteriores y el interés de esta tesis en centrar el estudio solo en la información brindada por la Oficina de Admisión, además de las restricciones de tiempo y recursos, la metodología de Ralph Kimball se ajusta más a lo que se quiere desarrollar al permitir la creación de un Datamart que pueda ayudar en la toma de decisiones de la Oficina de Registros Académicos de la facultad.

De esta forma, la tesis puede considerarse un primer acercamiento a la Inteligencia de Negocios en el ámbito académico a partir del estudio de un proceso de negocio concreto, el proceso de admisión. La metodología de Kimball nos permite, por tanto, ofrecer soluciones en un plazo inferior al que resultaría de abordar un desarrollo global destinado a la facultad o inclusive a toda la universidad.

Por otro lado, la metodología de Kimball ofrece una clara exposición de las fases a seguir y de las actividades propias de cada fase, así como un buen número de ejemplos documentados en los cuales apoyarse.

A continuación, se muestra nuevamente la figura del ciclo de vida de la metodología de Kimball:

Ilustración XIV - Ciclo de vida metodología de Kimball



Fuente: (Kimball Ralph, 2004)

Las actividades que están sombreadas en azul son las que se van a servir de guía para la presente tesis.

3.2.2.1 Planificación y gestión del desarrollo de la tesis

3.2.2.1.1 Cronogramas de actividades

En este apartado se definirán todas las actividades y tareas a realizar en la presente tesis. Además, se podrá tener una visión de la duración de cada tarea, así como también las dependencias entre dichas tareas. (Se anexará el documento Project)

3.2.2.1.2 Roles del equipo de trabajo BI

Se elaborará un gráfico con los roles desempeñados por el equipo de trabajo.

3.2.2.1.3 Asignación de tareas por rol

En este punto se mostrará un recuadro de asignación de tareas por rol.

3.2.2.1.4 Estimación de tiempo total

En esta sección se indicará la estimación de tiempo total en días para el desarrollo del modelo.

3.2.2.2 Definición de los requerimientos del negocio

En este punto se realizará el análisis del negocio, comenzando con entrevistas a los usuarios, para luego definir los requerimientos funcionales y los no funcionales.

3.2.2.2.1 Preparación y desarrollo de entrevistas

Para entender mejor los requerimientos se debe empezar por hablar con los usuarios clave del negocio. De acuerdo a las mejores prácticas, se elaborará un cuestionario para entrevistar a dichos usuarios, con el objetivo de comprender sus necesidades y plantear los requerimientos que serán cubiertos a lo largo de la tesis.

3.2.2.2.2 Definición de requerimientos funcionales

En este punto se presentarán aquellos requerimientos o requisitos que se relacionan directamente con las funciones específicas que proveerá el Datamart. Para este propósito se utilizará la siguiente tabla:

Ilustración XV - Requerimientos funcionales

N°	Requerimiento	Prioridad	Exigible /Deseable
1	El sistema permitirá generar un reporte en el que se muestre el dinero que ingresa a la entidad a través de un período de tiempo. Los conceptos por los cuales se recaudan los ingresos debe ser mostrado en toda su jerarquía (Categoría, Específica, Sub- específica); el período de tiempo.	Media	Exigible
2
3
4

Elaboración: los autores

3.2.2.2.3 Definición de requerimientos no Funcionales

En este punto se presentarán los requerimientos no funcionales que debe cumplir el Datamart. Los requerimientos no funcionales se encargan de analizar los factores tales como facilidades de uso, escalabilidad, portabilidad, disponibilidad, seguridad, etc. Para este propósito se utilizará la siguiente tabla:

Ilustración XVI - Requerimientos no funcionales

N°	Requerimiento	Prioridad	Dificultad
1	El sistema deberá estar desarrollado en una herramienta de software libre.	Alta	Alta
2
3
4

Elaboración: los autores

3.2.2.2.4 Análisis de fuentes de datos

En este apartado se ubican cada una de las fuentes de datos necesarias para resolver las preguntas de negocio encontradas en los puntos anteriores. Este paso requiere que se hayan designado usuarios de negocio que conozcan el detalle funcional de las fuentes de información.

Luego se elaborará un mapeo para encontrar la relación entre los conceptos de negocio y los atributos de cada fuente de datos. Para este propósito se utilizará la siguiente matriz:

Ilustración XVII - Análisis de fuentes de datos

N°	Concepto de Negocio	Fuente de Datos	Campo Fuente
1	Modalidad de admision	Admision.txt	CODADMISION
2	
3	
4	

Elaboración: los autores

3.2.2.3 Modelamiento dimensional

En esta sección se analizará el modelo dimensional lógico que responde a los requerimientos expuestos en la sección anterior. El proceso de diseño dimensional comenzará con un modelo dimensional de alto nivel obtenido a partir de los requerimientos priorizados. El proceso consta de cuatro pasos.

3.2.2.3.1 Definición del modelo de la estructura de datos

En este punto se definirá el modelo de estructura de datos para el desarrollo del Datamart. Para lo cual se realizará una comparación entre el modelo Estrella y el Copo de Nieve, describiendo las ventajas y desventajas de cada uno.

3.2.2.3.2 Establecimiento del nivel de granularidad

La granularidad significa especificar el nivel de detalle. La elección de la granularidad dependerá de los requerimientos del negocio del punto anterior y lo que es posible a partir de los datos actuales. Se comenzará a diseñar el Datamart al mayor nivel de detalle posible, ya que se podrá luego realizar agrupamientos al nivel deseado.

3.2.2.3.3 Elección y detalle de las dimensiones

Las dimensiones se determinarán en base a las variables por las cuales los usuarios realizan su análisis. Para determinar las dimensiones con las que contará el Datamart, se identificará las variables de análisis por las cuales el usuario suele elaborar sus reportes, realizando un listado de variables. Luego se agruparán las variables de análisis por la afinidad entre ellas, formando las dimensiones del Datamart.

3.2.2.3.4 Identificación de medidas y tablas de hechos

En este punto se identificarán las medidas que surgen de los procesos de negocios. Una medida es un atributo (campo) de una tabla

que se desea analizar, sumando o agrupando sus datos, usando los criterios de corte conocidos como dimensiones.

Para un mejor entendimiento, el detalle de las tablas de hechos se realizará tomando en cuenta la siguiente estructura:

- Tabla de Hechos
- Descripción
- Granularidad (Dimensiones)
- Medidas
- Indicadores

3.2.2.3.5 Desarrollo del modelo lógico dimensional

Una vez realizados los cuatro puntos anteriores, se procederá a elaborar el modelo lógico dimensional que consiste en modelamiento de las tablas de hechos con sus respectivas dimensiones.

3.2.2.4 Diseño físico

El diseño físico de la base de datos o arquitectura de base de datos se focaliza sobre la selección de las estructuras de almacenamiento necesarias para soportar el diseño lógico. En este apartado, se diseñarán las tablas que van a dar el soporte físico al modelo multidimensional, así como también el diseño de las columnas con su respectivo tipo de dato. Además, se elegirá el esquema de almacenamiento de la información. Por último, se diseñarán otras estructuras de datos como índices y particiones.

3.2.2.4.1 Definición del esquema de almacenamiento

En este punto se especificará que tipo de esquema de almacenamiento se va a utilizar para la implementación del Datamart. Esto se refiere a que, si el Datamart utilizará un esquema de tipo MOLAP, HOLAP o ROLAP. Cualquiera de estos esquemas debe contribuir al adecuado desempeño del Datamart en base a las especificaciones hechas en la etapa de definición de requerimientos.

3.2.2.4.2 Diseño de tablas y columnas

En este apartado se empleará una herramienta de modelado para realizar el diseño las tablas con sus respectivas columnas. Tomando en cuenta los tipos de atributos, longitudes, llaves primarias y secundarias entre otros detalles. El diseño físico permitirá luego la generación del script DDL para la generación de la base de datos.

3.2.2.4.3 Diseño de índices

Se diseñarán los índices en las tablas que los requieran para poder optimizar las búsquedas. Se definirán uno a más campos como índices, dependiendo de la necesidad del usuario.

3.2.2.4.4 Desarrollar diseño físico

En este punto se mostrará el diseño físico ya finalizado, con sus respectivas tablas, campos, tipo de datos y otras especificaciones. Además, se adjuntarán el DLL para la creación de las tablas.

3.2.2.5 Diseño y desarrollo de los procesos de ETL

En este paso se diseñarán los procesos para la extracción de los datos desde diversas fuentes, aplicando diferentes reglas para aumentar la calidad y consistencia de estos, consolidando la información proveniente de distintos sistemas, y finalmente cargando la información en el Datamart en un formato acorde para la utilización por parte de las herramientas de análisis.

3.2.2.5.1 Elaborar el mapeo de datos

La elaboración del mapeo de datos entre la fuente de información y las tablas destino se realizará con la ayuda de una plantilla en Excel. En ella se plasmará detalladamente la relación entre campo fuente y campo destino en las tablas del Datamart. También se estarán detallando los filtros, transformaciones u otro particular en el manejo de la información.

3.2.2.5.2 Diseño y desarrollo de los procesos de ETL

Como tarea inicial, se realizará el diseño y desarrollo de los procesos de extracción, transformación y carga de todas las dimensiones del modelo multidimensional, comenzando por aquellas dimensiones que son más simples para luego llegar a las más complejas. Luego se diseñarán y desarrollarán los procesos de extracción, transformación y carga de las tablas de hechos.

Por otro lado, se definirán las lógicas del cambio de una dimensión como también el proceso de carga histórica de la tabla de hechos.

3.2.2.5.3 Diseño de la secuencia de ejecución

Una vez terminada la labor del diseño de los procesos ETL, es necesario determinar cuál será la secuencia de ejecución de dichos procesos teniendo en cuenta las dependencias existentes en el modelo.

3.2.2.6 Diseño de la arquitectura técnica

En este punto se presentará la arquitectura de la solución. Esta constituye el diseño de una aplicación a alto nivel; es decir, se realiza el estudio de la estructura de la aplicación desde el punto de vista de componentes que interactúan entre sí. Además, se mostrarán los componentes que intervienen en el diseño de la arquitectura del Datamart y la interacción con el resto. Por último, se detallarán cada uno de los componentes de la Arquitectura elegida.

3.2.2.7 Selección de los productos e instalación

En esta sección se definirán las herramientas que se utilizarán en la implementación (simulación) del Datamart.

3.2.2.7.1 Elaboración de matriz herramienta por actividad

La matriz se desarrolló en el apartado “3.1 Materiales”, revisar dicha sección para mayor detalle.

3.2.2.7.2 Instalación de herramientas para simulación

En esta actividad se prepara el ambiente necesario para simulación del modelo.

3.2.2.8 Diseño y desarrollo de aplicaciones BI

Una parte fundamental de todo desarrollo de BI está en proporcionarles a una gran comunidad de usuarios una forma más estructurada y, por lo tanto, más fácil, de acceder al almacén de datos. Proporcionamos este acceso estructurado a través de lo que llamamos aplicaciones de inteligencia de negocios.

3.2.2.8.1 Diseño y desarrollo de reportes

La explotación de la información de Datamart se realizará a través del diseño de reportes los cuales serán estructurados con el usuario de negocio con el fin de contener la información que ellos requieren.

3.2.2.8.2 Diseño y desarrollo de cubos OLAP

En este paso se realizará el diseño de Cubos OLAP dependiendo de los requerimientos funcionales definidos inicialmente.

3.2.2.9 Implementación (Simulación)

Debido a la limitación de tiempo que existe para la elaboración de la tesis, el alcance no contempla la implementación del Datamart en la facultad. Lo que si se realizará es una simulación de la implementación en una laptop.

3.2.2.10 Mantenimiento y crecimiento

Debido a la misma razón del punto anterior, el alcance de la tesis no contempla un plan de mantenimiento ni un crecimiento a futuro.

3.3 Cronograma de la tesis

En este punto se muestra el cronograma de actividades de la tesis, donde se detallan las principales actividades en función de las fechas asignadas, duración y actividades predecesoras. **(Ver Anexo 1)**.

CAPÍTULO IV

DESARROLLO DEL MODELO

4.1 Planificación y gestión del desarrollo de la tesis

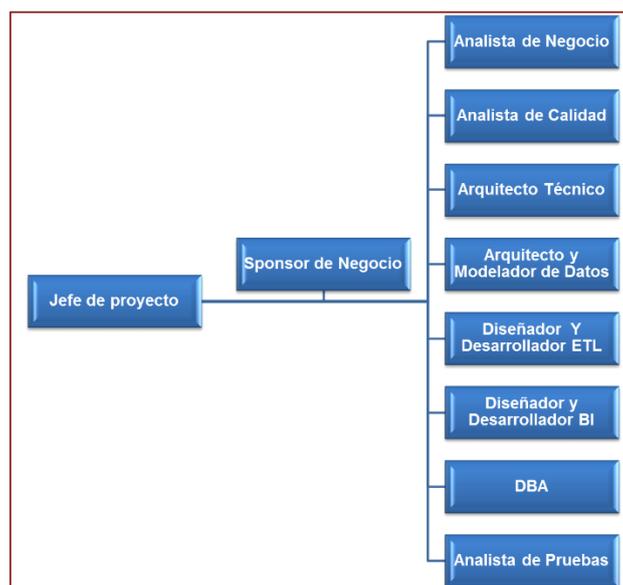
4.1.1 Cronograma de actividades

Se definieron las actividades y tareas a realizar en el desarrollo de las actividades de la tesis. Se puede observar el tiempo de duración de cada tarea y las dependencias entre ellas. **(Ver Anexo 2)**.

4.1.2 Roles del equipo de trabajo BI

A continuación, se muestra los roles que se desempeñados en el desarrollo de la tesis.

Ilustración XVIII - Roles del equipo de trabajo



Elaboración: los autores

- **Jefe de proyecto:** se encarga de planificar, gestionar y controlar los recursos y tareas necesarias para llevar a cabo el proyecto de BI. También estima los tiempos y recursos necesarios para el desarrollo de las actividades programadas. El jefe de proyecto define el perfil de cada miembro del equipo de trabajo y asigna sus responsabilidades.
- **Sponsor de negocio (Decano FIA – USMP):** es el patrocinador e interesado en la realización del desarrollo de la tesis. Además, cumple el rol de facilitador de contactos, entrevistas y Visto bueno para recabar información de los procesos de negocio e información técnica confidencial.
- **Analista de negocio:** se encarga de realizar las entrevistas y comprender los problemas y necesidades por parte del cliente. Analiza y elabora los diferentes requerimientos (funcionales – no funcionales), también participa en la fase de pruebas para velar que se cumplan los requerimientos del cliente.
- **Analista de calidad:** Valida la calidad de los datos, así como los mecanismos de conectividad de las fuentes de información, supervisa que la ejecución y desarrollo de los procesos ETL se realicen de acuerdo con la metodología y estándares establecidos, registrando todas las observaciones y comentarios técnicos para proponer los ajustes necesarios.
- **Arquitecto técnico:** Define la configuración de los componentes de la solución de inteligencia de negocios, contemplando que se utilicen los datos de manera eficiente y precisa. Adicionalmente participa en la actividad de selección de los productos de instalación.

- **Arquitecto modelador de base de datos:** se encarga de modelar los aspectos de almacenamiento de datos del sistema. Las clases persistentes, sus atributos, y sus relaciones.
- **Diseñador y desarrollador BI:** es responsable de diseñar las soluciones de BI y verificar si estas soluciones de BI son eficaces. Si existe algún problema con la solución el Desarrollador BI se encarga de averiguar cuál es el problema y solucionarlo. También es responsable del diseño y desarrollo de los cubos OLAP, trabaja con base de datos relacionales y multidimensionales. El desarrollador BI además diseña reportes, incluso asegura que los procesos operativos están funcionando correctamente.
- **Diseñador y desarrollador ETL:** es responsable de coordinar y dirigir todas las actividades relacionadas con el diseño y desarrollo de los procesos ETL para el Datamart. Esta función es responsable de trabajar con los analistas de negocio y los administradores de datos para identificar los sistemas de origen y de datos de destino. El conocimiento de las mejores prácticas de ETL es fundamental para un desempeño exitoso en este papel.
- **DBA (Administrador de base de datos):** tiene la responsabilidad de mantener y operar las bases de datos que conforman el sistema de información. Un DBA garantiza la seguridad de las bases de datos, incluyendo backups y recuperación de desastres. Analiza y reportan datos corporativos que ayuden a la toma de decisiones en la inteligencia de negocios.
- **Analista de pruebas:** es responsable de la ejecución del procedimiento de pruebas del Datamart.

4.1.3 Asignación de tareas por rol

En este punto se muestra la asignación de tareas por rol. **(Ver Anexo 3).**

4.1.4 Estimación de tiempo total

Podemos ver en la siguiente tabla la duración en días del desarrollo de la tesis según la estimación realizada (113 – aproximadamente 4 meses) y la duración del desarrollo del modelo (51 días), adicionalmente observamos las actividades principales con su respectivo tiempo asignado, duración y sus actividades predecesoras.

Tabla III – Tiempo total del desarrollo

N°	Tareas	Duración	Comienzo	Fin	Predecesoras
1	DATAMART DE REGISTROS ACADÉMICOS - FIA	113 días	sáb 08/03/14	sáb 28/06/14	
19	DESARROLLO DEL PROYECTO	51 días	dom 06/04/14	lun 26/05/14	14
20	Planificación del Proyecto	7 días	dom 06/04/14	sáb 12/04/14	
26	Definición de los requerimientos del negocio	7 días	dom 13/04/14	sáb 19/04/14	20
32	Modelamiento Dimensional	14 días	dom 20/04/14	sáb 03/05/14	26
39	Diseño Físico	7 días	dom 04/05/14	sáb 10/05/14	32
46	Diseño y desarrollo de los procesos de ETL	14 días	dom 11/05/14	sáb 24/05/14	39
51	Diseño de la arquitectura técnica	3 días	dom 20/04/14	mar 22/04/14	26
53	Selección de los productos e instalación	4 días	mié 23/04/14	sáb 26/04/14	51
57	Diseño y Desarrollo de aplicaciones BI	28 días	dom 27/04/14	sáb 24/05/14	26
63	Implementación (Simulación)	2 días	dom 25/05/14	lun 26/05/14	53,46,57

Elaboración: los autores

4.2 Definición de los requerimientos del negocio

El análisis del negocio consistió en un conjunto de entrevistas que se realizaron a los diferentes roles tanto de la Oficina de Admisión como de la Facultad de Ingeniería y Arquitectura. Las entrevistas fueron realizadas a roles pertenecientes a los tres niveles de la organización: Nivel estratégico, táctico y operativo con el fin de poder cubrir con la totalidad de necesidades.

Además, se realizó el análisis de un grupo de documentos tales como reportes en Excel, diagramas de flujo, estructura de datos, entre otros; que permitieron definir los requerimientos que deberá cumplir el Datamart, con el objetivo de satisfacer las necesidades expuestas por cada uno de los usuarios.

4.2.1 Preparación y desarrollo de entrevistas

Se elaboró un cuestionario (**Ver Anexo 4**) para las entrevistas con los usuarios claves del negocio, con el objetivo de comprender sus necesidades y plantear los requerimientos que serán cubiertos a lo largo de la tesis.

4.2.2 Definición de requerimientos funcionales

En este punto se presentan los requerimientos que se relacionan directamente con las funciones específicas que proveerá el Datamart:

Tabla IV - Definición de requerimientos funcionales

Nº	Requerimiento	Prioridad	Exigible /Deseable
1	Se desea contar con un mínimo de 7 reportes que presenten los datos cargados en el Datamart y cuyo diseño ayude a una correcta toma de decisiones en la FIA - USMP. Estos reportes se describen en la sección 3.8.1	Alta	Exigible
2	Se permitirá que el usuario elija los filtros y criterios que desea para sus reportes.	Alta	Exigible
3	Se requiere que el Datamart permita comparar la información de data histórica según seleccione el usuario.	Alta	Exigible
4	Se requiere que el Datamart permita presentar los informes de manera gráfica, para una mejor visualización de la información a través del tiempo o según criterios elegidos por el usuario.	Media	Exigible
5	Se requiere visualizar los reportes en un formato compatible con los Smarthphone.	Baja	Deseable
6	Se desea contar con Indicadores Clave de Desempeño que permitan a los usuarios tomar decisiones correctas en un periodo corto de tiempo.	Media	Exigible

Elaboración: los autores

4.2.3 Definición de requerimientos no funcionales

En este punto se presentan los requerimientos no funcionales que debe cumplir el Datamart:

Tabla V - Definición de requerimientos no funcionales

N°	Requerimiento	Prioridad	Dificultad
1	El modelo multidimensional debe ser un esquema base que permita ser implementado en el futuro en otras facultades	Alta	Alta
2	La herramienta de explotación seleccionada debe permitir realizar técnicas de consulta multidimensional a nivel de jerarquías.	Alta	Media
3	No deben existir dimensiones duplicadas.	Baja	Baja
4	Se requieren reportes predefinidos. Además, es deseable la posibilidad de exportación de archivos a formatos estándar como: Excel y Texto.	Media	Media
5	Se requiere minimizar el espacio de uso en disco para procesos temporales.	Media	Baja
6	Se requiere mantener un control de inconsistencias de información	Media	Alta

Elaboración: los autores

4.2.4 Análisis de fuentes de datos

4.2.4.1 Fuentes maestras

A continuación, se muestra la tabla del análisis que se realizó a la fuente de datos **DATOS_POSTULANTE.XLS**, que fue proporcionado por la Oficina de admisión de la Universidad. En esta tabla se puede apreciar todos los campos que dicha fuente contiene y sus respectivos conceptos de negocio. Esta fuente será utilizada por el proceso de ETL para poblar las tablas de dimensiones como las tablas de hechos del Datamart.

Tabla VI – Análisis de fuentes de datos 1 (DATOS_POSTULANTE.XLS)

N°	Concepto de Negocio	Campo Fuente
1	Código del Ingresante	CODALUMNO
2	Apellido Paterno	PATERO
3	Apellido materno	MATERNO
4	Nombres	NOMBRES
5	Sexo del ingresante	SEXO
6	Fecha de Nacimiento	FNACIMIEN
7	1er Teléfono ingresante o contacto	TELEFONO1
8	2do Teléfono ingresante o contacto	TELEFONO2
9	Modalidad de admisión	MODALIDAD
10	Nombre de la carrera de ingreso	ESPECIALIDAD
11	Puntaje de ingreso	PUNTAJE
12	Orden de mérito de ingreso	MERITO
13	Nombre del colegio de procedencia	NOMCOLEGI
14	Ubicación geográfica del Colegio	UBICOLEGI
15	Tipo de Colegio	TIPCOLEGI
16	Año de Egreso del Colegio	EGRESOCOL
17	Dirección del Colegio	DIRECCION
18	Dirección de residencia del ingresante	UBIRESIDENCIA
19	Estado civil del ingresante	ESTADOCIVI
20	Situación laboral del ingresante	TRABAJA
21	Correo electrónico	EMAIL
22	Tipo documento de identificación	TIPDOCUM
23	Nro. Documento de identificación	NUMERODOC
24	País de procedencia	PAIS
25	Situación actual de la postulación	INGRESO

Elaboración: los autores

También se cuenta con la fuente DATOS_MATRICULA.XLS que nos entrega el área de sistemas de la facultad que será utilizada para identificar que ingresantes llegaron a matricularse en el semestre académico.

Tabla VII – Análisis de fuentes de datos 2 (DATOS_MATRICULA.XLS)

N°	Concepto de Negocio	Campo Fuente
1	Código del Ingresante	CODALUMNO
2	Apellido Paterno	PATERNO
3	Apellido materno	MATERNO
4	Nombres	NOMBRES
5	Modalidad de Ingreso	MODALIDAD
6	Nombre de la carrera de ingreso	ESPECIALIDAD
7	Situación actual de la postulación	INGRESO
8	Situación actual de Matricula	MATRICULA
9	Fecha de Matricula	FECMATRICULA

Elaboración: los autores

4.2.4.2 Otras fuentes

Además de la fuente maestra que trae información de los postulantes a la facultad, existen otras fuentes que son de carga única.

Tabla VIII - Análisis de fuentes de datos 3

N°	Concepto de Negocio	Fuente de Datos	Campo Fuente
1	Nombre del Centro Educativo	COLEGIOS.XLS	NOMBRE
2	Tipo de Colegio	COLEGIOS.XLS	GESTIÓN
3	Departamento de Ubigeo	COLEGIOS.XLS	DEPARTAMENTO
4	Provincia de Ubigeo del Colegio	COLEGIOS.XLS	PROVINCIA
5	Distrito de Ubigeo del colegio	COLEGIOS.XLS	DISTRITO
6	Dirección del Colegio	COLEGIOS.XLS	DIRECCION
7	Código del Ubigeo	UBIGEOS_R.XLS	UBIGEO_R
8	Nombre del Departamento	UBIGEOS_R.XLS	NODEP
9	Nombre de la Provincia	UBIGEOS_R.XLS	NOPRV
10	Nombre del Distrito	UBIGEOS_R.XLS	NODIS
11	Código del Dpto. o Provincia Padre	UBIGEOS_R.XLS	PADRE_R
12	Tipo de Examen de la modalidad	MODALIDAD.XLS	TIPO PRUEBA
13	Tipo de Modalidad	MODALIDAD.XLS	MODALIDAD
14	Detalle del Examen	MODALIDAD.XLS	DETALLE
15	Nombre de la Sede Principal o Filial	ESPECIALIDAD.XLS	SEDE
16	Nombre de la Facultad de estudios	ESPECIALIDAD.XLS	FACULTAD
17	Nombre de la Carrera profesional	ESPECIALIDAD.XLS	CARRERA

Elaboración: los autores

4.3 Modelamiento dimensional

En esta sección se analiza el modelo dimensional lógico que responde a los requerimientos expuestos en la sección anterior.

4.3.1 Definición del modelo de la estructura de datos

La elección del modelo de estructura de datos que se utilizó en el Datamart, se eligió en base una comparación entre los Modelos Estrella y Copo de Nieve, tomando en cuenta ciertos criterios a los cuales se les asigna un valor.

Tabla IX - Comparación modelo estructura de datos

N°	Criterio	Modelo	
		Estrella	Copo de Nieve
1	Menor Complejidad de Consulta	5	3
2	Normalización	1	5
3	Menor Espacio de Almacenamiento	2	4
4	Menor Redundancia de datos	1	4
5	Mayor Rendimiento	5	2
Total		14	18

Elaboración: los autores

Observamos en la Tabla IX que el modelo de estructura de datos que mejor se ajusta para el cumplimiento de los objetivos trazados, es el modelo de Copo de Nieve.

4.3.2 Establecimiento del nivel de granularidad

Según el análisis realizado a raíz de las fuentes y medidas que se definieron, encontramos 15 entidades que nos permiten tener un nivel máximo de detalle (grano fino) y algunas con un nivel mínimo de detalle (grano grueso) y lo podemos observar en la siguiente tabla.

Tabla X - Establecimiento del nivel de granularidad

N°	Concepto	Máximo Nivel de detalle	Mínimo Nivel de detalle
1	Colegio	Colegio	-
2	Edades	Rango Edad	-
3	Egreso	Rango Egreso	-
4	Especialidad	Carrera	Sede
5	Estado Civil	Estado Civil	-
6	Letra Inicial	Letra Inicial	-
7	Merito	Rango Merito	-
8	Modalidad	Modalidad	Tipo de Prueba
9	Proceso	Semestre	Año
10	Puntaje	Rango Puntaje	-
11	Sexo	Sexo	-
12	Trabaja	Trabaja	-
13	Ubigeo	Distrito	Departamento
14	Ubigeo Colegio	Distrito	Departamento
15	Matricula	Matricula	-
16	Ingreso	Ingreso	-

Elaboración: los autores

4.3.3 Elección y detalle de las dimensiones

Para determinar las dimensiones del datamart se comenzó identificando variables de análisis por las cuales el usuario suele elaborar sus reportes, entre ellas tenemos las siguientes:

Tabla XI - Variables de análisis

Variables de Análisis		
Colegio	Carrera	Puntaje
Tipo de Colegio	Facultad	Merito
Modalidad de Ingreso	Letra Inicial	Fecha de Egreso de Colegio
Tipo de Prueba	Ubigeo del Postulante	Matricula
Sede	Estado Civil	Ingreso
Edad	Año	Semestre
Genero	Estado Laboral	Ubigeo Colegio
Dirección Colegio	-	-

Elaboración: los autores

Luego se agruparon las variables de análisis de acuerdo con la afinidad entre ellas, formando las dimensiones del Datamart y siguiendo un estándar en la nomenclatura.

- **Dimensiones:** Las tablas que representan las dimensiones llevan la nomenclatura DIM_ + NOMBRE DE DIMENSIÓN
- **Tabla de Hechos:** Esta tabla tienen como nomenclatura: FACT_ + NOMBRE DE TABLA DE HECHOS
- **Llave Primaria:** Las llaves primarias de las dimensiones llevan la nomenclatura: COD_ + NOMBRE DE DIMENSION.

A continuación, se muestran todas las dimensiones definidas en el modelo dimensional, con su respectiva descripción, campos y tipos de datos. Para ver el detalle de los dominios de campos, **(Ver Anexo 5)**

a) Dimensión Carrera

Tabla XII - Dimensión Carrera

TABLA	DIM_CARRERA
DESCRIPCIÓN	Contiene la información de las especialidades a las que se postulan

NOMBRE	DESCRIPCIÓN	TIPO DE DATO
COD_CARRERA	Código único que identifica la carrera	Numérico
CARRERA	Nombre de la Carrera	Texto (100)
CARRERA_ABREV	Abreviatura de la Carrera	Texto (50)
COD_FACULTAD	Código único que identifica la facultad	Numérico
FEC_MOD	Ultima Fecha de modificación de la tabla	Fecha

Elaboración: los autores

b) Dimensión Colegio

Tabla XIII - Dimensión Colegio

TABLA	DIM_COLEGIO
DESCRIPCIÓN	Contiene los datos principales de los colegios

NOMBRE	DESCRIPCIÓN	TIPO DE DATO
COD_COLEGIO	Código único que identifica el Colegio	Numérico
COLEGIO	Nombre del colegio	Texto (100)
COD_TIPO_COLEGIO	Código único que identifica el tipo de colegio	Numérico
DIRECCION_COLEGIO	Dirección del colegio	Texto (120)
COD_UBIGEO_COLEGIO	código único que identifica la ubicación geográfica del colegio	Numérico
FEC_MOD	Ultima Fecha de modificación de la tabla	Fecha

Elaboración: los autores

c) Dimensión Estado Civil

Tabla XIV - Dimensión Estado Civil

TABLA	DIM_ESTADO_CIVIL
DESCRIPCIÓN	Almacena los diferentes estados civiles que podría tener el postulante

NOMBRE	DESCRIPCIÓN	TIPO DE DATO
COD_ESTADO_CIVIL	Código único que identifica el estado civil del alumno	Numérico
ESTADO_CIVIL	Descripción del Estado Civil	Texto (15)
FEC_MOD	Ultima Fecha de modificación de la tabla	Fecha

Elaboración: los autores

d) Dimensión Facultad

Tabla XV - Dimensión Facultad

TABLA	DIM_FACULTAD
DESCRIPCIÓN	Contiene los nombres de las distintas facultades

NOMBRE	DESCRIPCIÓN	TIPO DE DATO
COD_FACULTAD	Código único que identifica la facultad	Numérico
FACULTAD	Nombre de la facultad	Texto (100)
FACULTAD_ABREV	Abreviatura del nombre de la facultad	Texto (50)
COD_SEDE	Código único que identifica a la Sede	Numérico
FEC_MOD	Ultima Fecha de modificación de la tabla	Fecha

Elaboración: los autores

e) Dimensión género

Tabla XVI - Dimensión Género

TABLA	DIM_GENERO
DESCRIPCIÓN	Contiene los géneros masculino y femenino

NOMBRE	DESCRIPCIÓN	TIPO DE DATO
COD_GENERO	Código único que identifica al género	Numérico
GENERO	Descripción del género	Texto (15)
FEC_MOD	Ultima Fecha de modificación de la tabla	Fecha

Elaboración: los autores

f) Dimensión Ingreso

Tabla XVII - Dimensión Ingreso

TABLA	DIM_INGRESO
DESCRIPCIÓN	Hace referencia al ingreso del postulante

NOMBRE	DESCRIPCIÓN	TIPO DE DATO
COD_INGRESO	código único que identifica el estado de ingreso del postulante	Numérico
INGRESO	Estado ingreso del postulante	Carácter (2)
FEC_MOD	Ultima fecha de modificación de la tabla	Fecha

Elaboración: los autores

g) Dimensión Letra Inicial

Tabla XVIII - Dimensión Letra Inicial

TABLA	DIM_LETRA_INICIAL_APE
DESCRIPCIÓN	Contiene todas las letras del abecedario

NOMBRE	DESCRIPCIÓN	TIPO DE DATO
COD_LETRA_INICIAL_APE	Código único que identifica la letra del abecedario	Numérico
LETRA_INICIAL_APE	Letra del Abecedario	Carácter (1)
FEC_MOD	Ultima Fecha de modificación de la tabla	Fecha

Elaboración: los autores

h) Dimensión Matricula

Tabla XIX - Dimensión Matricula

TABLA	DIM_MATRICULA
DESCRIPCIÓN	Hace referencia al estado de la matrícula del ingresante

NOMBRE	DESCRIPCIÓN	TIPO DE DATO
COD_MATRICULA	Código único que identifica el estado de la matricula	Numérico
MATRICULA	Estado de la matrícula del ingresante	Carácter (2)
FEC_MOD	Ultima Fecha de modificación de la tabla	Fecha

Elaboración: los autores

i) Dimensión Modalidad

Tabla XX - Dimensión Modalidad

TABLA	DIM_MODALIDAD
DESCRIPCIÓN	Almacena las diferentes modalidades de ingreso a la universidad

NOMBRE	DESCRIPCIÓN	TIPO DE DATO
COD_MODALIDAD	Código único que identifica la modalidad	Numérico
MODALIDAD	Nombre de la modalidad	Texto (50)
COD_TIPO_PRUEBA	Código único que identifica el tipo de prueba	Numérico
FEC_MOD	Ultima Fecha de modificación de la tabla	Fecha

Elaboración: los autores

j) Dimensión Proceso

Tabla XXI - Dimensión Proceso

TABLA	DIM_PROCESO
DESCRIPCIÓN	Hace referencia al año y semestre del proceso de admisión

NOMBRE	DESCRIPCIÓN	TIPO DE DATO
COD_PROCESO	Código único que identifica al proceso de admisión	Numérico
ANIO	Año del proceso	Numérico
SEMESTRE	Semestre del Proceso	Numérico
FEC_MOD	Ultima Fecha de modificación de la tabla	Fecha

Elaboración: los autores

k) Dimensión Rango Edad

Tabla XXII - Dimensión Rango Edad

TABLA	DIM_RANGO_EDAD
DESCRIPCIÓN	Almacena un grupo de Rangos Predeterminados de edades

NOMBRE	DESCRIPCIÓN	TIPO DE DATO
COD_RANGO_EDAD	código único que identifica el rango de edad	Numérico
RANGO_EDAD	Descripción del rango de la edad	Texto (50)
MIN_VAL	Valor mínimo del rango de edad	Numérico
MAX_VAL	Valor máximo del rango de edad	Numérico
FEC_MOD	Ultima Fecha de modificación de la tabla	Fecha

Elaboración: los autores

l) Dimensión Rango Egreso Colegio

Tabla XXIII - Dimensión Rango Egreso Colegio

TABLA	DIM_RANGO_EGRESO
DESCRIPCIÓN	Guarda un grupo de rangos predeterminados de los años en que egresaron los postulantes

NOMBRE	DESCRIPCIÓN	TIPO DE DATO
COD_RANGO_EGRESO	código único que identifica el rango de Egreso del Colegio	Numérico
RANGO_EGRESO_COLEGIO	Descripción del rango de egreso del colegio	Texto (50)
MIN_VAL	Valor mínimo del rango de egreso	Numérico
MAX_VAL	Valor máximo del rango de egreso	Numérico
FEC_MOD	Ultima Fecha de modificación de la tabla	Fecha

Elaboración: los autores

m) Dimensión Rango Mérito

Tabla XXIV - Dimensión Rango Mérito

TABLA	DIM_RANGO_MERITO
DESCRIPCIÓN	Hace referencia a los intervalos de mérito en los que se puede ingresar

NOMBRE	DESCRIPCIÓN	TIPO DE DATO
COD_RANGO_MERITO	código único que identifica el rango de merito	Numérico
RANGO_MERITO	Descripción del rango de mérito de ingreso	Texto (50)
MIN_VAL	Valor mínimo del rango de mérito	Numérico
MAX_VAL	Valor máximo del rango de mérito	Numérico
FEC_MOD	Ultima Fecha de modificación de la tabla	Fecha

Elaboración: los autores

n) Dimensión Rango puntaje

Tabla XXV - Dimensión Rango Puntaje

TABLA	DIM_RANGO_PUNTAJE
DESCRIPCIÓN	Contiene un conjunto de rangos de puntaje en la prueba de admisión

NOMBRE	DESCRIPCIÓN	TIPO DE DATO
COD_RANGO_PUNTAJE	código único que identifica el rango de puntaje de ingreso	Numérico
RANGO_PUNTAJE	Descripción del rango de puntaje del examen de ingreso	Texto (50)
MIN_VAL	Valor mínimo del rango de puntaje	Numérico
MAX_VAL	Valor máximo del rango de puntaje	Numérico
FEC_MOD	Ultima Fecha de modificación de la tabla	Fecha

Elaboración: los autores

o) Dimensión Sede

Tabla XXVI - Dimensión Sede

TABLA	DIM_SEDE
DESCRIPCIÓN	Hace referencia a la sede a la que se postula

NOMBRE	DESCRIPCIÓN	TIPO DE DATO
COD_SEDE	Código único que identifica a la Sede	Numérico
SEDE	Nombre de la Sede de la Universidad	Texto (30)
FEC_MOD	Ultima Fecha de modificación de la tabla	Fecha

Elaboración: los autores

p) Dimensión Tipo Prueba

Tabla XXVII - Dimensión Tipo Prueba

TABLA	DIM_TIPO_PRUEBA
DESCRIPCIÓN	Almacena información respecto a tipo de prueba de la modalidad de ingreso

NOMBRE	DESCRIPCIÓN	TIPO DE DATO
COD_TIPO_PRUEBA	Código único que identifica el tipo de prueba	Numérico
TIPO_PRUEBA	Tipo de examen de ingreso	Texto (50)
DETALLE_PRUEBA	Detalle del tipo de examen de ingreso	Texto (100)
FEC_MOD	Ultima Fecha de modificación de la tabla	Fecha

Elaboración: los autores

q) Dimensión Tipo Colegio

Tabla XXVIII - Dimensión Tipo Colegio

TABLA	DIM_TIPO_COLEGIO
DESCRIPCIÓN	Contiene los Tipo de Colegio Privado y Público

NOMBRE	DESCRIPCIÓN	TIPO DE DATO
COD_TIPO_COLEGIO	código único que identifica el tipo de colegio	Numérico
TIPO_COLEGIO	Tipo de Colegio de procedencia	Texto (20)
FEC_MOD	Ultima Fecha de modificación de la tabla	Fecha

Elaboración: los autores

r) Dimensión Trabaja

Tabla XXIX - Dimensión Trabaja

TABLA	DIM_TRABAJA	
DESCRIPCIÓN	Almacena la situación laboral actual del postulante	
NOMBRE	DESCRIPCIÓN	TIPO DE DATO
COD_TRABAJA	código único que identifica la situación laboral actual	Numérico
TRABAJA	Situación Laboral del postulante	Caracter(2)
FEC_MOD	Ultima Fecha de modificación de la tabla	Fecha

Elaboración: los autores

s) Dimensión Ubigeo

Tabla XXX - Dimensión Ubigeo

TABLA	DIM_UBIGEO	
DESCRIPCIÓN	Guarda las diversas ubicaciones geográficas a nivel mundial	
NOMBRE	DESCRIPCIÓN	TIPO DE DATO
COD_UBIGEO	código único que identifica la ubicación geográfica	Numérico
UBIGEO	Ubicación geográfica	Texto (50)
COD_UBIGEO_PADRE	código único que identifica la ubicación geográfica Padre del colegio	Numérico
FEC_MOD	Ultima Fecha de modificación de la tabla	Fecha

Elaboración: los autores

t) Dimensión Ubigeo Colegio

Tabla XXXI - Dimensión Ubigeo Colegio

TABLA	DIM_UBIGEO_COLEGIO
DESCRIPCIÓN	Guarda las diversas ubicaciones geográficas de los colegios a nivel nacional

NOMBRE	DESCRIPCIÓN	TIPO DE DATO
COD_UBIGEO_COLEGIO	código único que identifica la ubicación geográfica del colegio	Numérico
UBIGEO_COLEGIO	Ubicación geográfica del Colegio	Texto (50)
COD_UBI_COL_PADRE	código único que identifica la ubicación geográfica Padre del colegio	Numérico
FEC_MOD	Ultima Fecha de modificación de la tabla	Fecha

Elaboración: los autores

4.3.4 Identificación de Medidas y tabla de Hechos

Para el desarrollo del Datamart se identificó una medida clave con la cual se explotó la información de la mejor manera. La medida utilizada fue el Número de Alumnos.

Luego se definió la tabla de hechos FACT_ALUMNO, que contiene todas las características del alumno y la relación con sus dimensiones.

Tabla XXXII - Tabla de hechos

TABLA	FACT_ALUMNO	
DESCRIPCIÓN	Almacena los datos relacionados a los Alumnos	
NOMBRE	DESCRIPCIÓN	TIPO DE DATO
COD_ALUMNO	Código único que identifica al alumno	Texto (10)
NOMBRE	Nombres y Apellidos del Alumno	Texto (100)
FEC_NACIMIENTO	Fecha de Nacimiento del Alumno	Fecha
EDAD	Campo calculado de la edad del alumno	Numérico
TELEFONO_1	Teléfono de contacto 1	Texto (15)
TELEFONO_2	Teléfono de contacto 2	Texto (15)
PUNTAJE	Puntaje del Ingreso	Numérico
MERITO	Orden de mérito de ingreso	Numérico
DIRECCION	Dirección de residencia	Texto (150)
EMAIL	Email de contacto del alumno	Texto (60)
TIPO_DOCUMENTO	Tipo de Documento	Texto (40)
NUM_DOCUMENTO	Número del documento	Texto (15)
FEC_MATRICULA	Fecha de Matricula	Fecha
COD_UBIGEO	código único que identifica la ubicación geográfica	Numérico
COD_MODALIDAD	código único que identifica la modalidad	Numérico
COD_COLEGIO	código único que identifica al colegio	Numérico
COD_CARRERA	código único que identifica la carrera	Numérico
COD_GENERO	código único que identifica al género	Numérico

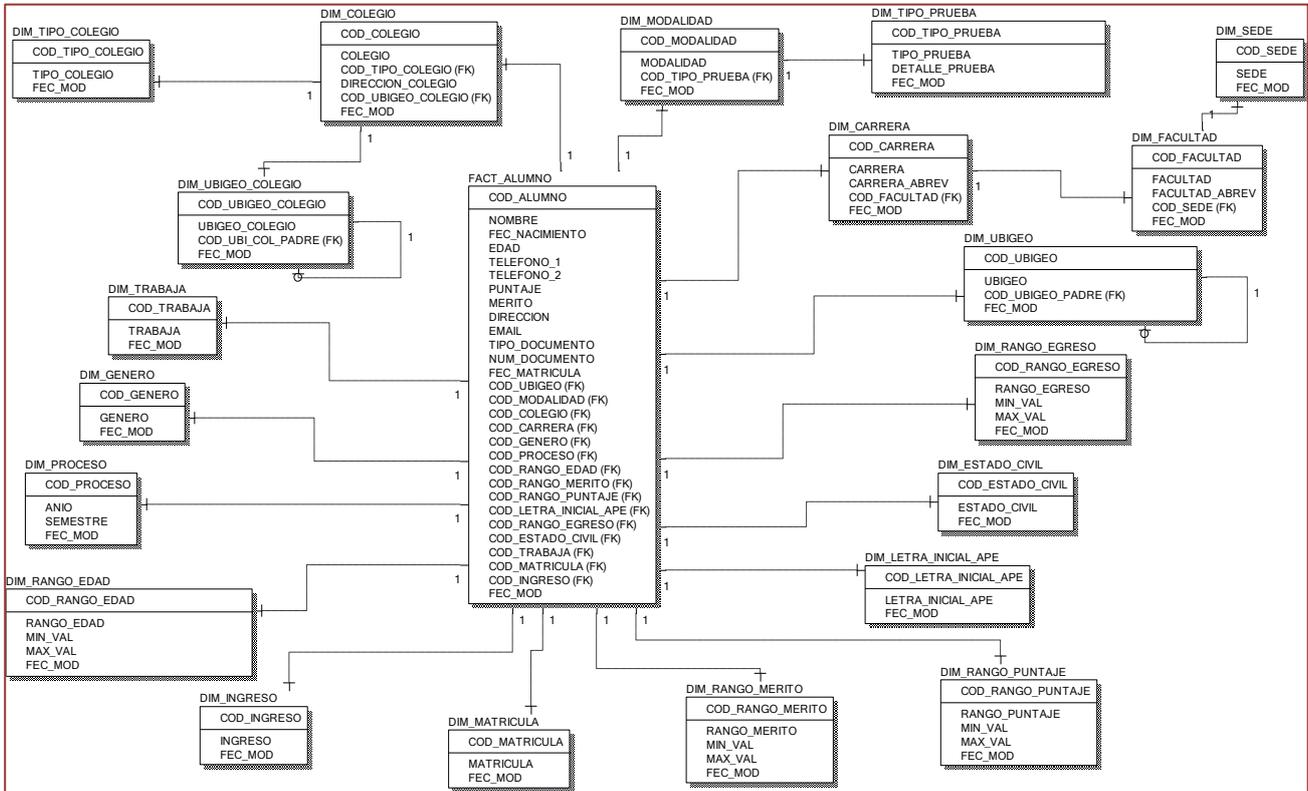
COD_PROCESO	código único que identifica al proceso de admisión	Numérico
COD_RANGO_EDAD	código único que identifica el rango de edad	Numérico
COD_RANGO_MERITO	código único que identifica el rango de mérito	Numérico
COD_RANGO_PUNTAJE	código único que identifica el rango del puntaje de ingreso	Numérico
COD_LETRA_INICIAL_APE	código único que identifica la letra inicial del apellido	Numérico
COD_RANGO_EGRESO	código único que identifica el rango de egreso del colegio	Numérico
COD_ESTADO_CIVIL	código único que identifica el estado civil	Numérico
COD_TRABAJA	código único que identifica la situación laboral actual	Numérico
COD_MATRICULA	código único que identifica el estado de la matricula	Numérico
COD_INGRESO	código único que identifica el estado de ingreso del postulante	Numérico
FEC_MOD	Ultima Fecha de modificación de la tabla	Fecha

Elaboración: los autores

4.3.5 Desarrollo del Modelo Lógico dimensional

En la Ilustración XIX se muestra el modelo lógico dimensional, utilizando la estructura de almacenamiento copo de nieve.

Ilustración XIX - Modelo lógico dimensional



Elaboración: los autores

4.4 Diseño físico

En este punto se brinda el esquema del almacenamiento que vamos a utilizar para el desarrollo del Datamart, también se muestra las tablas diseñadas con su respectivo tipo de datos, las cuales dan soporte al diseño físico.

4.4.1 Definición del esquema de almacenamiento

En esta sección, según las características de los esquemas de almacenamiento, sus ventajas y la problemática de la tesis, pasamos a definir el esquema que se utilizó en el Datamart. ROLAP se utiliza en el mercado generalmente, por compañías con grandes volúmenes de datos (millones de registros) y adicionalmente una de sus características más relevantes es que puede soportar un gran número de dimensiones. Por otro lado, tenemos que MOLAP es una solución departamental, adecuada para bajos volúmenes de información y número de dimensiones más modestos. Esto sumado a que una de sus características principales es que tiene un rendimiento superior a ROLAP.

Debido a que estamos realizando un modelo para la Oficina de Registros Académicos, que no cuenta con grandes volúmenes de información y según juicio de expertos, hemos optado por el esquema de almacenamiento MOLAP.

4.4.2 Diseño de tablas y columnas

Para el diseño de las tablas con sus respectivas columnas se utilizó CA. Erwin. El nombre que se definió para la base de datos fue BD_DATAMART_FIA.

a) Dimensión Carrera

Tabla XXXIII - Def. columnas dimensión Carrera

Tabla "DIM_CARRERA"		
Nombre	Tipo de dato	Opción Null
ID_CARRERA	smallint	IDENTITY
CARRERA	nvarchar(100)	NOT NULL
CARRERA_ABREV	nvarchar(50)	NOT NULL
ID_FACULTAD	smallint	NOT NULL
FEC_MOD	smalldatetime	NOT NULL

Elaboración: los autores

b) Dimensión Colegio

Tabla XXXIV - Def. columnas dimensión Colegio

Tabla "DIM_COLEGIO"		
Nombre	Tipo de dato	Opción Null
ID_COLEGIO	smallint	IDENTITY
COLEGIO	nvarchar(100)	NOT NULL
ID_TIPO_COLEGIO	smallint	NOT NULL
DIRECCION_COLEGIO	nvarchar(120)	NULL
ID_UBIGEO_COLEGIO	smallint	NOT NULL
FEC_MOD	smalldatetime	NOT NULL

Elaboración: los autores

c) Dimensión Estado Civil

Tabla XXXV - Def. columnas dimensión Estado Civil

Columnas de Tabla "DIM_ESTADO_CIVIL"		
Nombre	Tipo de dato	Opción Null
ID_ESTADO_CIVIL	smallint	IDENTITY
ESTADO_CIVIL	nvarchar(15)	NOT NULL
FEC_MOD	smalldatetime	NOT NULL

Elaboración: los autores

d) Dimensión Facultad

Tabla XXXVI - Def. columnas dimensión Facultad

Tabla "DIM_FACULTAD"		
Nombre	Tipo de dato	Opción Null
ID_FACULTAD	smallint	IDENTITY
FACULTAD	nvarchar(100)	NOT NULL
FACULTAD_ABREV	nvarchar(50)	NOT NULL
ID_SEDE	smallint	NOT NULL
FEC_MOD	smalldatetime	NOT NULL

Elaboración: los autores

e) Dimensión Género

Tabla XXXVII - Def. columnas dimensión Género

Tabla "DIM_GENERO"		
Nombre	Tipo de dato	Opción Null
ID_GENERO	smallint	IDENTITY
GENERO	nvarchar(15)	NOT NULL
FEC_MOD	smalldatetime	NOT NULL

Elaboración: los autores

f) Dimensión Ingreso

Tabla XXXVIII - Def. columnas dimensión Ingreso

Columnas de Tabla "DIM_INGRESO"		
Nombre	Tipo de dato	Opción Null
ID_INGRESO	smallint	IDENTITY
INGRESO	char(2)	NOT NULL
FEC_MOD	smalldatetime	NOT NULL

Elaboración: los autores

g) Dimensión Letra Inicial

Tabla XXXIX - Def. columnas dimensión Letra Inicial

Tabla "DIM_LETRA_INICIAL_APE"		
Nombre	Tipo de dato	Opción Null
ID_LETRA_INICIAL_APE	smallint	IDENTITY
LETRA_INICIAL_APE	char(1)	NOT NULL
FEC_MOD	smalldatetime	NOT NULL

Elaboración: los autores

h) Dimensión Matricula

Tabla XL - Def. columnas dimensión Matricula

Tabla "DIM_MATRICULA"		
Nombre	Tipo de dato	Opción Null
ID_MATRICULA	smallint	IDENTITY
MATRICULA	char(2)	NOT NULL
FEC_MOD	smalldatetime	NOT NULL

Elaboración: los autores

i) Dimensión Modalidad

Tabla XLI - Def. columnas dimensión Modalidad

Tabla "DIM_MODALIDAD"		
Nombre	Tipo de dato	Opción Null
ID_MODALIDAD	smallint	IDENTITY
MODALIDAD	nvarchar(50)	NOT NULL
ID_TIPO_PRUEBA	smallint	NOT NULL
FEC_MOD	smalldatetime	NOT NULL

Elaboración: los autores

j) Dimensión Proceso

Tabla XLII - Def. columnas dimensión Proceso

Tabla "DIM_PROCESO"		
Nombre	Tipo de dato	Opción Null
ID_PROCESO	smallint	IDENTITY
ANIO	smallint	NOT NULL
SEMESTRE	smallint	NOT NULL
FEC_MOD	smalldatetime	NOT NULL

Elaboración: los autores

k) Dimensión Rango Edad

Tabla XLIII - Def. columnas dimensión Rango Edad

Tabla "DIM_RANGO_EDAD"		
Nombre	Tipo de dato	Opción Null
ID_RANGO_EDAD	smallint	IDENTITY
RANGO_EDAD	nvarchar(50)	NOT NULL
MIN_VAL	smallint	NOT NULL
MAX_VAL	smallint	NOT NULL
FEC_MOD	smalldatetime	NOT NULL

Elaboración: los autores

l) Dimensión Rango Egreso Colegio

Tabla XLIV - Def. columnas dimensión Rango Egreso Colegio

Tabla "DIM_RANGO_EGRESO"		
Nombre	Tipo de dato	Opción Null
ID_RANGO_EGRESO	Smallint	IDENTITY
RANGO_EGRESO	nvarchar(50)	NOT NULL
MIN_VAL	smallint	NOT NULL
MAX_VAL	smallint	NOT NULL
FEC_MOD	smalldatetime	NOT NULL

Elaboración: los autores

m) Dimensión Rango Mérito

Tabla XLV - Def. columnas dimensión Rango Mérito

Tabla "DIM_RANGO_MERITO"		
Nombre	Tipo de dato	Opción Null
ID_RANGO_MERITO	Smallint	IDENTITY
RANGO_MERITO	nvarchar(50)	NOT NULL
MIN_VAL	smallint	NOT NULL
MAX_VAL	smallint	NOT NULL
FEC_MOD	smalldatetime	NOT NULL

Elaboración: los autores

n) Dimensión Rango puntaje

Tabla XLVI - Def. columnas dimensión Rango Puntaje

Tabla "DIM_RANGO_PUNTAJE"		
Nombre	Tipo de dato	Opción Null
ID_RANGO_PUNTAJE	smallint	IDENTITY
RANGO_PUNTAJE	nvarchar(30)	NOT NULL
MIN_VAL	smallint	NOT NULL
MAX_VAL	smallint	NOT NULL
FEC_MOD	smalldatetime	NOT NULL

Elaboración: los autores

o) Dimensión Sede

Tabla XLVII - Def. columnas dimensión Sede

Columnas de Tabla "DIM_SEDE"		
Nombre	Tipo de dato	Opción Null
ID_SEDE	smallint	IDENTITY
SEDE	nvarchar(30)	NOT NULL
FEC_MOD	smalldatetime	NOT NULL

Elaboración: los autores

p) Dimensión Tipo Prueba

Tabla XLVIII - Def. columnas dimensión Tipo Prueba

Tabla "DIM_TIPO_PRUEBA"		
Nombre	Tipo de dato	Opción Null
ID_TIPO_PRUEBA	smallint	IDENTITY
TIPO_PRUEBA	nvarchar(50)	NOT NULL
DETALLE_PRUEBA	nvarchar(100)	NULL
FEC_MOD	smalldatetime	NOT NULL

Elaboración: los autores

q) Dimensión Tipo Colegio

Tabla XLIX - Def. columnas dimensión Tipo Colegio

Tabla "DIM_TIPOCOLEGIO"		
Nombre	Tipo de dato	Opción Null
ID_TIPO_COLEGIO	smallint	IDENTITY
TIPO_COLEGIO	nvarchar(20)	NOT NULL
FEC_MOD	smalldatetime	NOT NULL

Elaboración: los autores

r) Dimensión Trabaja

Tabla L - Def. columnas dimensión Trabaja

Tabla "DIM_TRABAJA"		
Nombre	Tipo de dato	Opción Null
ID_TRABAJA	smallint	IDENTITY
TRABAJA	char(2)	NOT NULL
FEC_MOD	smalldatetime	NOT NULL

Elaboración: los autores

s) Dimensión Ubigeo

Tabla LI - Def. columnas dimensión Ubigeo

Tabla "DIM_UBIGEO"		
Nombre	Tipo de dato	Opción Null
ID_UBIGEO	smallint	NOT NULL
UBIGEO	nvarchar(50)	NOT NULL
ID_UBIGEO_PADRE	smallint	NULL
FEC_MOD	smalldatetime	NOT NULL

Elaboración: los autores

t) Dimensión Ubigeo Colegio

Tabla LII - Def. columnas dimensión Ubigeo Colegio

Tabla "DIM_UBIGEO_COLEGIO"		
Nombre	Tipo de dato	Opción Null
ID_UBIGEO_COLEGIO	smallint	NOT NULL
UBIGEO_COLEGIO	nvarchar(50)	NOT NULL
ID_UBI_COL_PADRE	smallint	NULL
FEC_MOD	smalldatetime	NOT NULL

Elaboración: los autores

u) Tabla de Hechos Alumno

Tabla LIII – Def. columnas tabla de hechos

Tabla "FACT_ALUMNO"		
Nombre	Tipo de dato	Opción Null
ID_ALUMNO	nvarchar(10)	NOT NULL
NOMBRE	nvarchar(100)	NOT NULL
FEC_NACIMIENTO	smalldatetime	NOT NULL
EDAD	smallint	NOT NULL
TELEFONO_1	nvarchar(15)	NULL
TELEFONO_2	nvarchar(15)	NULL
PUNTAJE	smallint	NOT NULL
MERITO	smallint	NOT NULL
DIRECCION	nvarchar(150)	NOT NULL
EMAIL	nvarchar(60)	NULL
TIPO_DOCUMENTO	nvarchar(40)	NOT NULL
NUM_DOCUMENTO	nvarchar(15)	NOT NULL
FEC_MATRICULA	smalldatetime	NULL
ID_UBIGEO	smallint	NOT NULL
ID_MODALIDAD	smallint	NOT NULL

ID_COLEGIO	smallint	NOT NULL
ID_CARRERA	smallint	NOT NULL
ID_GENERO	smallint	NOT NULL
ID_PROCESO	smallint	NOT NULL
ID_RANGO_EDAD	smallint	NOT NULL
ID_RANGO_MERITO	smallint	NOT NULL
ID_RANGO_PUNTAJE	smallint	NOT NULL
ID_LETRA_INICIAL_APE	smallint	NOT NULL
ID_RANGO_EGRESO	smallint	NOT NULL
ID_ESTADO_CIVIL	smallint	NOT NULL
ID_TRABAJA	smallint	NOT NULL
ID_MATRICULA	smallint	NOT NULL
ID_INGRESO	smallint	NOT NULL
FEC_MOD	smalldatetime	NOT NULL

Elaboración: los autores

4.4.3 Diseño de Índices y Funciones

Se diseñaron los índices en las tablas que los requieran para poder optimizar las búsquedas.

a) Dimensión Carrera

Tabla LIV - Índices dimensión Carrera

Tabla "DIM_CARRERA"		
Nombre	Es FK	Es PK
ID_CARRERA	No	Si
CARRERA	No	No
CARRERA_ABREV	No	No
ID_FACULTAD	Si	No
FEC_MOD	No	No

Elaboración: los autores

b) Dimensión Colegio

Tabla LV - Índices dimensión Colegio

Tabla "DIM_COLEGIO"		
Nombre	Es FK	Es PK
ID_COLEGIO	No	Si
COLEGIO	No	No
ID_TIPO_COLEGIO	Si	No
DIRECCION_COLEGIO	No	No
ID_UBIGEO_COLEGIO	Si	No
FEC_MOD	No	No

Elaboración: los autores

c) Dimensión Estado Civil

Tabla LVI - Índices dimensión Estado Civil

Tabla "DIM_ESTADO_CIVIL"		
Nombre	Es FK	Es PK
ID_ESTADO_CIVIL	No	Si
ESTADO_CIVIL	No	No
FEC_MOD	No	No

Elaboración: los autores

d) Dimensión Facultad

Tabla LVII - Índices dimensión Facultad

Tabla "DIM_FACULTAD"		
Nombre	Es FK	Es PK
ID_FACULTAD	No	Si
FACULTAD	No	No
FACULTAD_ABREV	No	No
ID_SEDE	Si	No
FEC_MOD	No	No

Elaboración: los autores

e) Dimensión Género

Tabla LVIII - Índices dimensión Género

Tabla "DIM_GENERO"		
Nombre	Es FK	Es PK
ID_GENERO	No	Si
GENERO	No	No
FEC_MOD	No	No

Elaboración: los autores

f) Dimensión Ingreso

Tabla LIX - Índices dimensión Ingreso

Tabla "DIM_INGRESO"		
Nombre	Es FK	Es PK
ID_INGRESO	No	Si
INGRESO	No	No
FEC_MOD	No	No

Elaboración: los autores

g) Dimensión Letra Inicial

Tabla LX - Índices dimensión Letra Inicial

Tabla "DIM_LETRA_INICIAL_APE"		
Nombre	Es FK	Es PK
ID_LETRA_INICIAL_APE	No	Si
LETRA_INICIAL_APE	No	No
FEC_MOD	No	No

Elaboración: los autores

h) Dimensión Matricula

Tabla LXI - Índices dimensión Matricula

Tabla "DIM_MATRICULA"		
Nombre	Es FK	Es PK
ID_MATRICULA	No	Si
MATRICULA	No	No
FEC_MOD	No	No

Elaboración: los autores

i) Dimensión Modalidad

Tabla LXII - Índices dimensión Modalidad

Tabla "DIM_MODALIDAD"		
Nombre	Es FK	Es PK
ID_MODALIDAD	No	Si
MODALIDAD	No	No
ID_TIPO_PRUEBA	Si	No
FEC_MOD	No	No

Elaboración: los autores

j) Dimensión Proceso

Tabla LXIII - Índices dimensión Proceso

Tabla "DIM_PROCESO"		
Nombre	Es FK	Es PK
ID_PROCESO	No	Si
ANIO	No	No
SEMESTRE	No	No
FEC_MOD	No	No

Elaboración: los autores

k) Dimensión Rango Edad

Tabla LXIV - Índices dimensión Rango Edad

Tabla "DIM_RANGO_EDAD"		
Nombre	Es FK	Es PK
ID_RANGO_EDAD	No	Si
RANGO_EDAD	No	No
MIN_VAL	No	No
MAX_VAL	No	No
FEC_MOD	No	No

Elaboración: los autores

I) Dimensión Rango Egreso Colegio

Tabla LXV - Índices dimensión Rango Egreso Colegio

Tabla "DIM_RANGO_EGRESO"		
Nombre	Es FK	Es PK
ID_RANGO_EGRESO	No	Si
RANGO_EGRESO	No	No
MIN_VAL	No	No
MAX_VAL	No	No
FEC_MOD	No	No

Elaboración: los autores

m) Dimensión Rango Mérito

Tabla LXVI - Índices dimensión Rango Mérito

Columnas de Tabla "DIM_RANGO_MERITO"		
Nombre	Es FK	Es PK
ID_RANGO_MERITO	No	Si
RANGO_MERITO	No	No
MIN_VAL	No	No
MAX_VAL	No	No
FEC_MOD	No	No

Elaboración: los autores

n) Dimensión Rango puntaje

Tabla LXVII - Índices dimensión Rango Puntaje

Tabla "DIM_RANGO_PUNTAJE"		
Nombre	Es FK	Es PK
ID_RANGO_PUNTAJE	No	Si
RANGO_PUNTAJE	No	No
MIN_VAL	No	No
MAX_VAL	No	No
FEC_MOD	No	No

Elaboración: los autores

o) Dimensión Sede

Tabla LXVIII - Índices dimensión Sede

Tabla "DIM_SEDE"		
Nombre	Es FK	Es PK
ID_SEDE	No	Si
SEDE	No	No
FEC_MOD	No	No

Elaboración: los autores

p) Dimensión Tipo Prueba

Tabla LXIX - Índices dimensión Tipo Prueba

Tabla "DIM_TIPO_PRUEBA"		
Nombre	Es FK	Es PK
ID_TIPO_PRUEBA	No	Si
TIPO_PRUEBA	No	No
DETALLE_PRUEBA	No	No
FEC_MOD	No	No

Elaboración: los autores

q) Dimensión Tipo Colegio

Tabla LXX - Índices dimensión Tipo Colegio

Tabla "DIM_TIPOCOLEGIO"		
Nombre	Es FK	Es PK
ID_TIPO_COLEGIO	No	Si
TIPO_COLEGIO	No	No
FEC_MOD	No	No

Elaboración: los autores

r) Dimensión Trabaja

Tabla LXXI - Índices dimensión Trabaja

Tabla "DIM_TRABAJA"		
Nombre	Es FK	Es PK
ID_TRABAJA	No	Si
TRABAJA	No	No
FEC_MOD	No	No

Elaboración: los autores

s) Dimensión Ubigeo

Tabla LXXII - Índices dimensión Ubigeo

Tabla "DIM_UBIGEO"		
Nombre	Es FK	Es PK
ID_UBIGEO	No	Si
UBIGEO	No	No
ID_UBIGEO_PADRE	Si	No
FEC_MOD	No	No

Elaboración: los autores

t) Dimensión Ubigeo Colegio

Tabla LXXIII - Índices dimensión Ubigeo Colegio

Tabla "DIM_UBIGEO_COLEGIO"		
Nombre	Es FK	Es PK
ID_UBIGEO_COLEGIO	No	Si
UBIGEO_COLEGIO	No	No
ID_UBI_COL_PADRE	Si	No
FEC_MOD	No	No

Elaboración: los autores

u) Tabla de hechos Alumno

Tabla LXXIV – Índices tabla de hechos Alumno

Tabla "FACT_ALUMNO"		
Nombre	Es FK	Es PK
ID_ALUMNO	No	Si
NOMBRE	No	No
FEC_NACIMIENTO	No	No
EDAD	No	No
TELEFONO_1	No	No
TELEFONO_2	No	No
PUNTAJE	No	No
MERITO	No	No
DIRECCION	No	No
EMAIL	No	No
TIPO_DOCUMENTO	No	No
NUM_DOCUMENTO	No	No
FEC_MATRICULA	No	No
ID_UBIGEO	Si	No
ID_MODALIDAD	Si	No
ID_COLEGIO	Si	No
ID_CARRERA	Si	No

ID_GENERO	Si	No
ID_PROCESO	Si	No
ID_RANGO_EDAD	Si	No
ID_RANGO_MERITO	Si	No
ID_RANGO_PUNTAJE	Si	No
ID_LETRA_INICIAL_APE	Si	No
ID_RANGO_EGRESO	Si	No
ID_ESTADO_CIVIL	Si	No
ID_TRABAJA	Si	No
ID_MATRICULA	Si	No
ID_INGRESO	Si	No
FEC_MOD	No	No

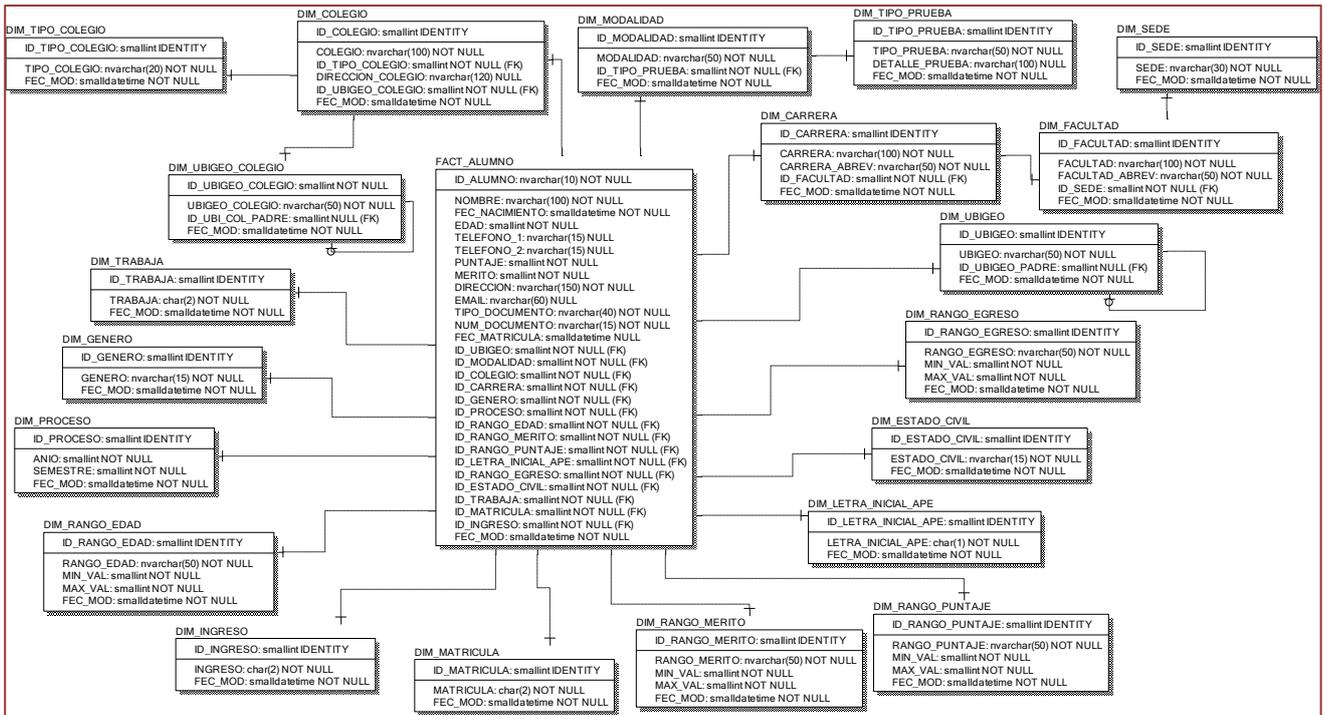
Elaboración: los autores

Se diseñaron funciones para mejorar el rendimiento de los procesos ETL que se detallaran más adelante. **(Ver Anexo 6)**

4.4.4 Desarrollar diseño físico

En la Ilustración XX podemos observar el diseño físico completo, adicionalmente se creó el Script DLL para la generación de las tablas en la base datos (Ver Anexo 7).

Ilustración XX - Diseño físico



Elaboración: los autores

4.5 Diseño y desarrollo de los procesos de ETL

Se diseñaron los procesos ETL para la extracción de los datos desde las distintas fuentes, aplicando reglas de limpieza de datos, para posteriormente cargar la información en el Datamart.

4.5.1 Elaborar el Mapeo de Datos

4.5.1.1 Mapeo Fuente COLEGIOS.XLS

La fuente COLEGIOS.XLS contiene la información de todos los colegios registrados a la fecha por la universidad. Esta fuente es utilizada para la carga de las dimensiones DIM_COLEGIO y DIM_TIPO_COLEGIO. Para mayor detalle (**Ver Anexo 8**).

4.5.1.2 Mapeo Fuente UBIGEOS_R.XLS

La fuente UBIGEOS_R.XLS contiene la información de todos departamentos, provincias y distritos del Perú. Además, se presenta información de continentes, países y ciudades en el caso de que sean ubigeos externos al Perú. Esta fuente es utilizada para la carga de las dimensiones DIM_UBIGEO y DIM_UBIGEO_COLEGIO. Para mayor detalle (**Ver Anexo 9**).

4.5.1.3 Mapeo Fuente MODALIDAD.XLS

La fuente MODALIDAD.XLS contiene la información de todas las modalidades de ingreso a la Universidad, con su respectivo tipo de prueba. Esta fuente carga las dimensiones DIM_MODALIDAD y DIM_TIPO_PRUEBA. Para mayor detalle (**Ver Anexo 10**).

4.5.1.4 Mapeo Fuente ESPECIALIDAD.XLS

La fuente ESPECIALIDAD.XLS presenta información de las diversas carreras que pertenecen a una cierta facultad y estas a su vez a una sede en particular. Esta fuente es utilizada para la carga de las dimensiones DIM_CARRERA, DIM_FACULTAD Y DIM_SEDE. Para mayor detalle (**Ver Anexo 11**).

4.5.1.5 Mapeo Fuente DATOS_MATRICULA.XLS

La fuente DATOS_MATRICULA.XLS es una fuente maestra que contiene información de todos los alumnos que se han matriculado o no en un determinado periodo académico. Esta fuente es utilizada para la carga de la dimensión DIM_MATRICULA y para la carga de la tabla de hechos FACT_ALUMNO. Para mayor detalle **(Ver Anexo 12)**.

4.5.1.6 Mapeo Fuente DATOS_POSTULANTE.XLS

La fuente DATOS_POSTULANTE.XLS es una fuente maestra que presenta los datos de todos los postulantes a las diversas carreras de la Facultad de Ingeniería y Arquitectura. Esta fuente es utilizada principalmente para la carga de la tabla de hechos FACT_ALUMNO. Además, esta fuente se utilizó para obtener el dominio de datos y luego poblar las dimensiones DIM_TRABAJA, DIM_GENERO, DIM_PROCESO, DIM_RANGO_EDAD, DIM_INGRESO, DIM_RANGO_MERITO, DIM_RANGO_PUNTAJE, DIM_LETRA_INICIAL_APE, DIM_ESTADO_CIVIL y DIM_RANGO_EGRESO. Para mayor detalle **(Ver Anexo 13 y Anexo 14)**.

4.5.2 Diseño y desarrollo de los procesos de ETL

El proceso de extracción, transformación y carga de las dimensiones y de la tabla de hechos se distribuyó en paquetes ELT, que son el estándar utilizado por la herramienta Integration Services de Microsoft. Dentro de cada paquete se encuentran contenedores de secuencia que son objetos que agrupan tareas y proporcionan estructura a los paquetes. A su vez, en cada contenedor se pueden incluir otros contenedores, así como también diversas tareas que se ejecutan en una secuencia determinada. La Tarea de Flujo de Datos es la más utilizada en el proceso ETL, ya que proporciona un control de la transformación de los datos desde que se extraen de la fuente hasta que se cargan en el destino final.

En la tesis, se utilizan tres tipos de nomenclaturas para las tareas de flujo de datos, que dependen del tratamiento de dichos datos. Las tareas de flujo de datos que tenga el prefijo “EX”, son aquellas que extraen y limpian los datos de la fuente. Las tareas que tengan un prefijo “RN”, son aquellas que aplican lógicas de negocio para luego obtener un resultado. Finalmente, las tareas con prefijo “CW”, son aquellas que toman

ese resultado y lo cargan al destino o tabla final. Existe una nomenclatura que hace referencia a tareas con funciones mixtas, como las tareas con prefijo “RN_CW”, que son las que aplican reglas y luego cargan el resultado al destino final.

El proceso ETL de las dimensiones del Datamart se separó en dos grandes paquetes, ELT_DIMENSIONES_SCD.dtsx y ETL_DIMENSIONES_NCD.dtsx. La distribución se realizó tomando como criterio la frecuencia con la que la información varía o cambia en el tiempo (Ejm: Proceso de Admisión, cada 6 meses). El primero corresponde a las dimensiones que cambian lentamente en el tiempo, es decir, dimensiones que poseen información que generalmente no se actualiza en cada proceso de admisión. En este paquete encontramos dimensiones como la tabla DIM_UBIGEO, DIM_COLEGIO, DIM_MODALIDAD, DIM_CARRERA, entre otras. Sin embargo, la frecuencia de ejecución de este paquete sí será semestral, pero solo se cargará la información diferencial, es decir, sólo los nuevos datos.

En el segundo paquete se encuentran las dimensiones que tienen información estática y que no varían en el tiempo. En este paquete encontramos dimensiones como DIM_GENERO, DIM_ESTADO_CIVIL, DIM_INGRESO, entre otras. En este paquete también se incluyeron las dimensiones que corresponden a los rangos, como por ejemplo la tabla DIM_RANGO_EDAD, DIM_RANGO_PUNTAJE, etc. La frecuencia de ejecución de este paquete es a pedido, sólo cuando se desee cargar un nuevo rango o valor en alguna dimensión.

El proceso de carga de la tabla de hechos, FACT_ALUMNO, se incluyó en el paquete ETL_FACT_ALUMNO.dtsx. La frecuencia de ejecución de este paquete es semestral, debido a que es necesario cargar la información de los postulantes en cada proceso de admisión. En este paquete también se incluyó la carga de la dimensión DIM_PROCESO, que se ejecuta después de la tarea de flujo de datos de extracción de la fuente DATOS_POSTULANTE.xls, pero antes de la tarea RN_CW (Reglas de negocio y carga) de la tabla de hechos FACT_ALUMNO. Esto se realizó así debido a que el año y semestre del proceso de admisión sólo se pueden obtener del código del postulante.

Con respecto a la limpieza y estandarización de datos en el proceso de ETL, se incluyeron objetos de limpieza de duplicados absolutos en todas las tareas de extracción de las fuentes. Estos objetos, identifican los duplicados de los registros que presentan igualdad en todos sus campos y los eliminan, dejando solo uno de ellos. En

la estandarización de datos, los valores de los campos que sean de tipo texto y que no tengan valores nulos, se convierten a mayúsculas y se eliminan los espacios en blanco. En algunos casos, dependiendo de la necesidad, los valores nulos de dichos campos se reemplazan por un punto, para que se puedan identificar luego como inconsistencia. Para el caso de otros tipos de datos, el valor nulo se trata de diversas maneras como se pudo observar en el mapeo de fuentes. Para mayor detalle ver el punto **“4.5.1 Elaborar Mapeos de Datos”**.

Finalmente, cabe resaltar que el proceso de ETL sólo carga la información diferencial, es decir, se carga solamente la información nueva que no haya sido cargada en ejecuciones anteriores. Por esta razón el proceso ETL está preparado para la ejecución normal en un determinado proceso de admisión más no para un reproceso. Si es necesario ejecutar nuevamente la carga de datos debido a algún problema en la calidad de la información, o en la ejecución misma del proceso, será necesario realizar una depuración de la tabla de hechos y de las tablas de dimensiones antes de ejecutar nuevamente el proceso. Para dicho propósito se diseñó un campo FEC_MOD que almacena la fecha en la que se cargan los datos a una dimensión o a la tabla de hechos, para así tener visibilidad de que data depurar.

4.5.2.1 Paquete ETL_DIMENSIONES_NCD

En este paquete se detalla el proceso ETL de las siguientes dimensiones:

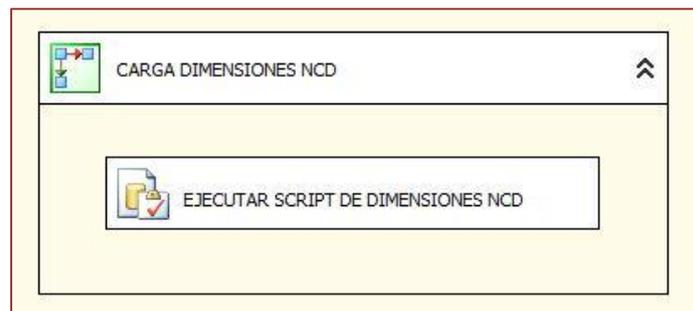
- DIM_ESTADO_CIVIL
- DIM_GENERO
- DIM_TRABAJA
- DIM_INGRESO
- DIM_MATRICULA
- DIM_LETRA_INICIAL_APE
- DIM_RANGO_MERITO

- DIM_RANGO_EDAD
- DIM_RANGO_PUNTAJE
- DIM_RANGO_EGRESO

Al ser información que no varía en el tiempo, la ejecución del paquete se realiza una única vez. Si se desea incluir nueva información en estas dimensiones, se tiene que solicitar a pedido la ejecución del paquete.

Para facilitar el desarrollo del proceso ETL y al ser información estática, se utilizó un script SQL para la carga de las tablas. El paquete ETL_DIMENSIONES_NCD, mediante una tarea, ejecuta el script de inserción hacia las tablas de dimensiones. A continuación, se muestra un gráfico del paquete ETL_DIMENSIONES_NCD.

Ilustración XXI - Proceso ELT_DIMENSIONES_NCD



Elaboración: los autores

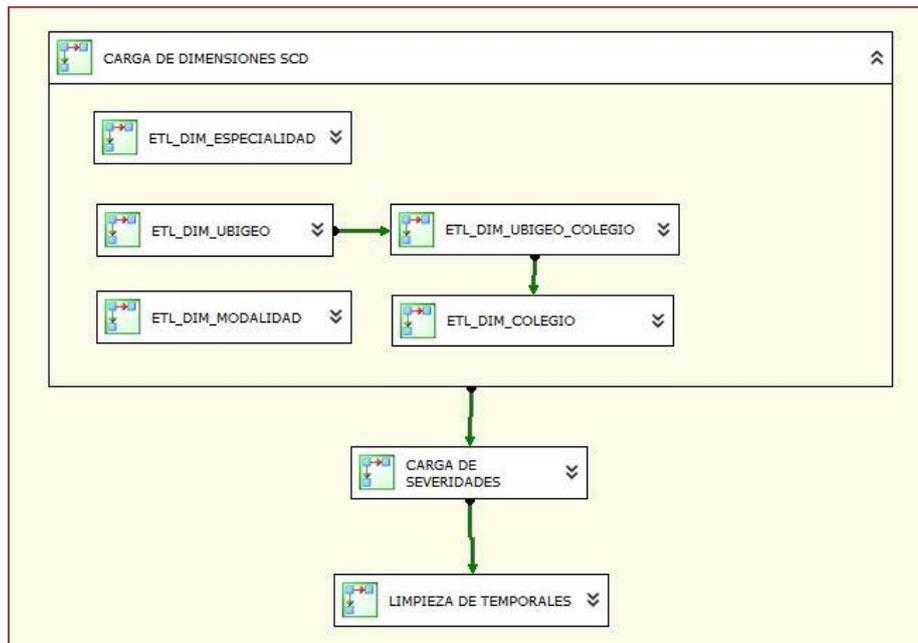
4.5.2.2 Paquete ETL_DIMENSIONES_SCD

En este paquete se detalla el proceso de ELT de las siguientes dimensiones:

- DIM_COLEGIO
- DIM_TIPO_COLEGIO
- DIM_UBIGEO
- DIM_UBIGEO_COLEGIO
- DIM_MODALIDAD
- DIM_TIPO_PRUEBA
- DIM_CARRERA
- DIM_FACULTAD
- DIM_SEDE

Dentro del paquete ETL_DIMENSIONES_SCD, se especifica un contenedor de secuencia de ejecución por cada dimensión o grupo de dimensiones según la dependencia que se tenga. Además, se incluyen otros contenedores de secuencia para la ejecución de otras tareas como Carga de Severidades y Limpieza de Temporales. A continuación, se muestra un gráfico del paquete ETL_DIMENSIONES_SCD.

Ilustración XXII - Paquete ETL_DIMENSIONES_SCD

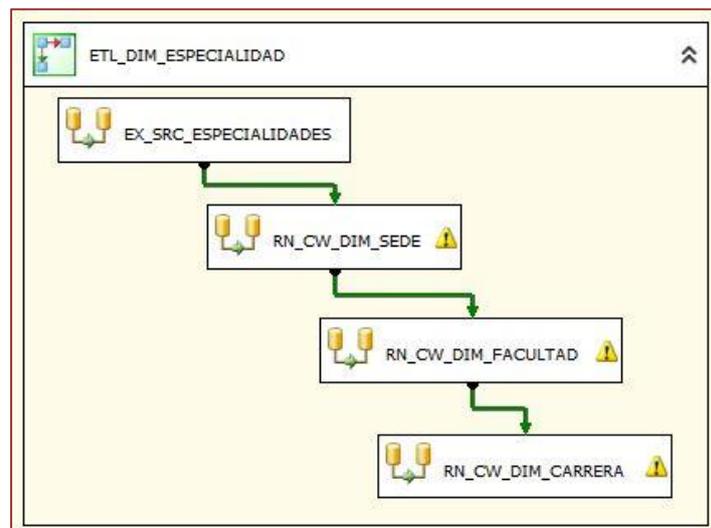


Elaboración: los autores

a) Contenedor ETL_DIM_ESPECIALIDAD

Este contenedor presenta el proceso de ELT de la dimensión DIM_SEDE, DIM_FACULTAD y DIM_CARRERA. Es independiente del resto, debido a que no consume información de ninguna otra dimensión SCD. A continuación, se muestra una imagen del proceso.

Ilustración XXIII - ELT_DIM_ESPECIALIDAD



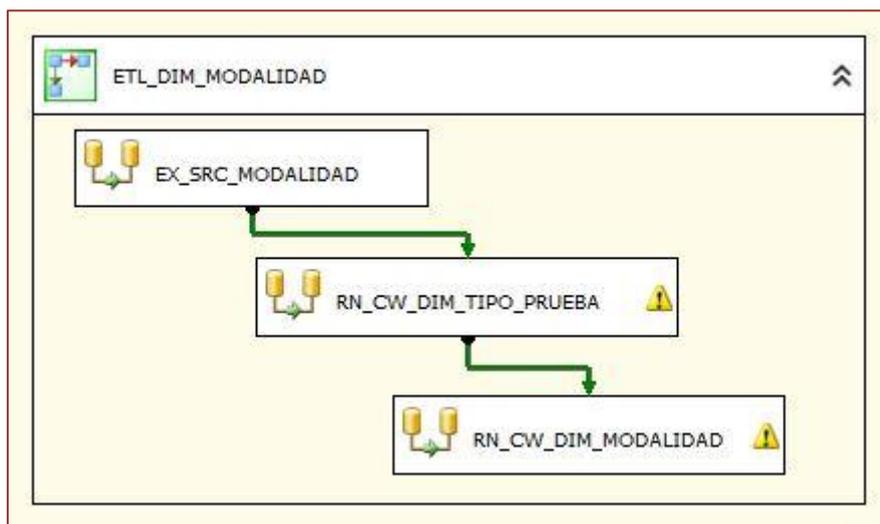
Elaboración: los autores

Como se puede apreciar, se diseñó un solo flujo de datos, EX_SRC_ESPECIALIDADES, para la extracción de fuente la fuente Especialidad.xls. Luego para cada dimensión se diseñó el flujo de datos de reglas de negocio y carga de tablas finales, RN_CW_DIM_SEDE, RN_CW_DIM_FACULTAD y RN_CW_DIM_CARRERA. Para ver el detalle de cada flujo de datos revisar **(Ver Anexo 15)**.

b) Contenedor ETL_DIM_MODALIDAD

El contenedor ETL_DIM_MODALIDAD presenta el proceso ETL de las dimensiones DIM_MODALIDAD y DIM_TIPO_PRUEBA. De la misma manera que el proceso anterior, este también es independiente, lo que significa que se ejecuta en paralelo a otros procesos dentro de un mismo contenedor. A continuación, se muestra una imagen del proceso en mención.

Ilustración XXIV - ETL_DIM_MODALIDAD



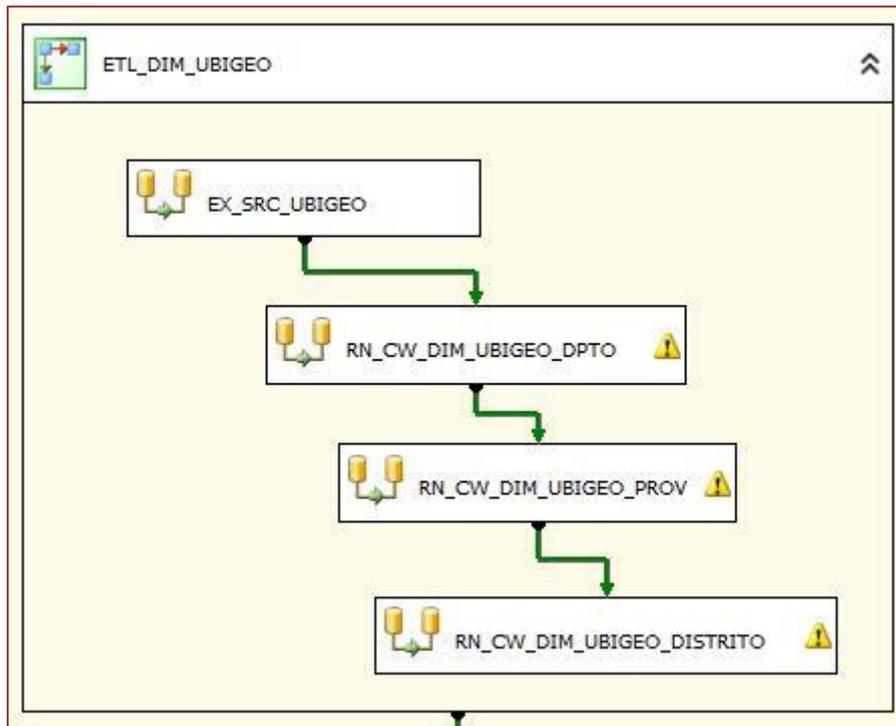
Elaboración: los autores

El primer flujo de datos EX_SRC_MODALIDAD, corresponde a la extracción de la fuente Modalidad.xls. Posteriormente se muestra el flujo de datos RN_CW_DIM_TIPO_PRUEBA donde se aplican las reglas de negocio y se carga la dimensión DIM_TIPO_PRUEBA. Por último, se muestra la tarea de flujo de datos RN_CW_DIM_MODALIDAD que se encarga de cargar la dimensión DIM_MODALIDAD. Para ver el detalle de cada flujo de datos **(Ver Anexo 15)**.

c) Contenedor ETL_DIM_UBIGEO

El contenedor ETL_DIM_UBIGEO muestra el proceso ETL encargado de poblar la dimensión DIM_UBIGEO. Este contenedor es independiente a los demás, dentro del contenedor padre. A continuación, se muestra una imagen que ilustra este proceso ETL.

Ilustración XXV - ELT_DIM_UBIGEO



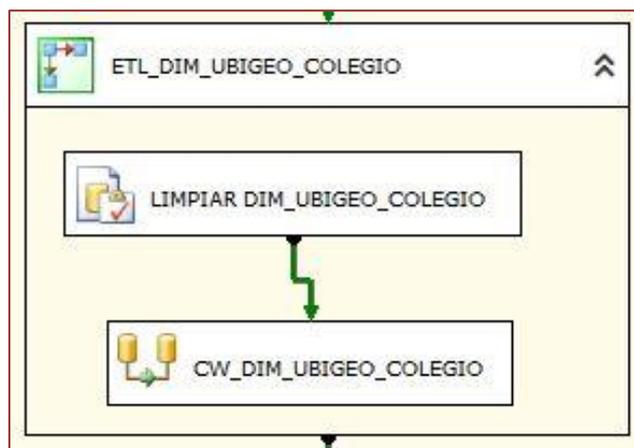
Elaboración: los autores

El flujo de datos inicial es el EX_SRC_UBIGEO, donde se extraen los datos de la fuente Ubigeo_R.xls. Posteriormente se muestran las tareas de flujo de datos RN_CW_DIM_UBIGEO_DPTO, RN_CW_DIM_UBIGEO_PROV y RN_CW_DIM_UBIGEO_DISTRITO que aplican reglas de negocio y carga la información relacionada a departamentos, provincias y distritos. Para ver el detalle de cada flujo de datos **(Ver Anexo 15)**.

d) Contenedor ETL_DIM_UBIGEO_COLEGIO

El contenedor ETL_DIM_UBIGEO_COLEGIO, presentan el proceso ETL que puebla la dimensión DIM_UBIGEO_COLEGIO. Este proceso depende de la correcta ejecución del contenedor ETL_DIM_UBIGEO. A continuación, se muestra una imagen del proceso.

Ilustración XXVI - ELT_DIM_UBIGEO_COLEGIO



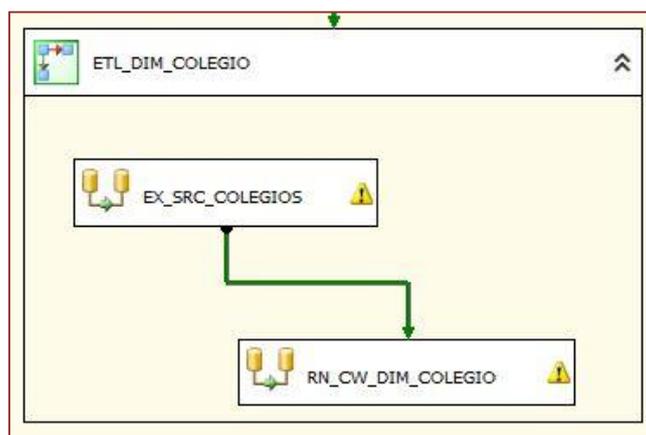
Elaboración: los autores

En la imagen anterior se muestra la tarea de ejecución de sql que limpia la dimensión DIM_UBIGEO_COLEGIO. Luego se muestra la tarea de flujo de datos que carga los ubigeos. Para ver el detalle de cada tarea (**Ver Anexo 15**).

e) Contenedor ETL_DIM_COLEGIO

El contenedor ETL_DIM_COLEGIO muestra el proceso de ETL que puebla la dimensión DIM_COLEGIO. Este contenedor depende de la correcta ejecución del contenedor ETL_DIM_UBIGEO_COLEGIO. A continuación, se muestra una imagen del proceso.

Ilustración XXVII - ETL_DIM_COLEGIO



Elaboración: los autores

La primera tarea de flujo de datos, EX_SRC_COLEGIOS, muestra el proceso de extracción de la fuente Colegios.xls. La siguiente tarea aplica reglas de negocio y carga la información de colegios a la dimensión DIM_COLEGIO. Para ver el detalle de cada flujo de datos (**Ver Anexo 15**).

f) Contenedor carga de severidades

El contenedor carga de severidades presenta la tarea de flujo de datos, que puebla la tabla de severidades del Datamart, para las inconsistencias del proceso ETL que cargan todas las dimensiones. A continuación, se muestra una imagen del proceso.

Ilustración XXVIII - Carga de severidades



Elaboración: los autores

g) Contenedor limpieza de temporales

El contenedor limpieza de temporales presenta el proceso de eliminación de archivos temporales que se hayan creado en los otros contenedores, producto de las múltiples transformaciones de los datos. Este proceso se realiza utilizando dos tareas de sistemas de archivos. A continuación, se muestra una imagen del proceso.

Ilustración XXIX - Limpieza de temporales

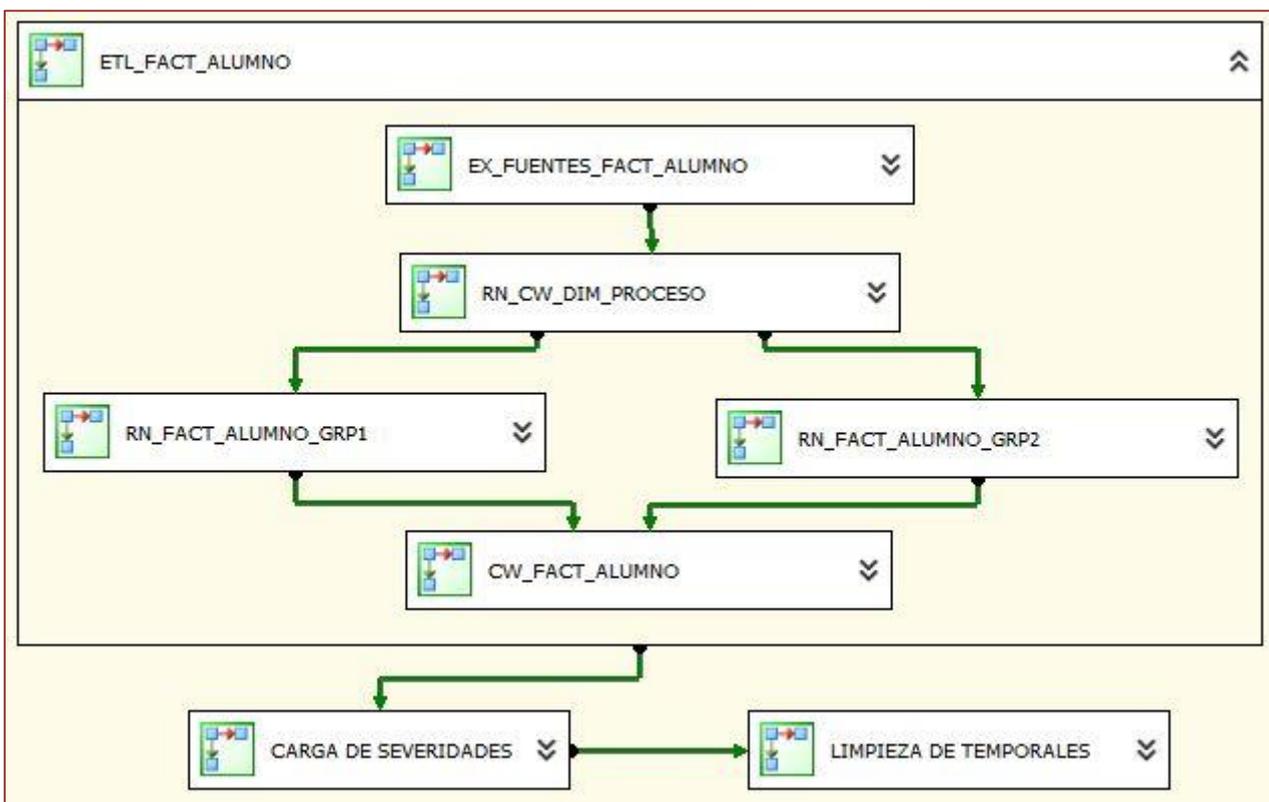


Elaboración: los autores

4.5.2.3 Paquete ETL_FACT_ALUMNO

En este paquete se detalla el proceso de ETL de la tabla de hechos FACT_ALUMNO y de la dimensión DIM_PROCESO. Dentro de este paquete se especifica un contenedor para la extracción de fuentes, otro para la carga de la dimensión DIM_PROCESO, dos contenedores de reglas de negocio y un contenedor para la carga final de la FACT_ALUMNO. Además, existe un contenedor para la Carga de Severidades y otro para la Limpieza de Temporales. A continuación, se muestra un gráfico del paquete ETL_FACT_ALUMNO.

Ilustración XXX – Paquete ETL_FACT_ALUMNO



Elaboración: los autores

a) Contenedor EX_FUENTES_FACT_ALUMNO

Este contenedor presenta el proceso de extracción de las fuentes Datos_Postulante.xls y Datos_Matricula.xls. Es independiente en términos de ejecución, es decir, no tiene relación con ningún otro anterior contenedor. A continuación, se muestra una imagen del proceso en mención.

Ilustración XXXI - EX_FUENTES_FACT_ALUMNO



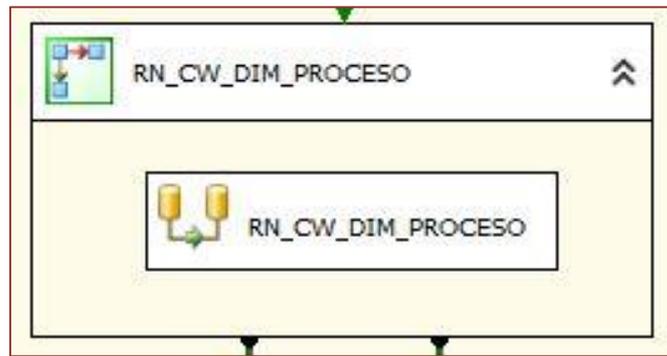
Elaboración: los autores

Se puede visualizar las dos tareas de flujos de datos EX_SRC_POSTULANTES y EX_SRC_MATRICULA que obtiene la información desde fuentes diferentes. Para ver el detalle de cada flujo de datos (**Ver Anexo 15**).

b) Contenedor RN_CW_DIM_PROCESO

Este contenedor muestra el proceso ETL de la dimensión DIM_PROCESO. Se puede observar que existe una dependencia con respecto a la extracción de las fuentes en el contenedor anterior. A continuación, se observa una imagen del proceso.

Ilustración XXXII - RN_CW_DIM_PROCESO



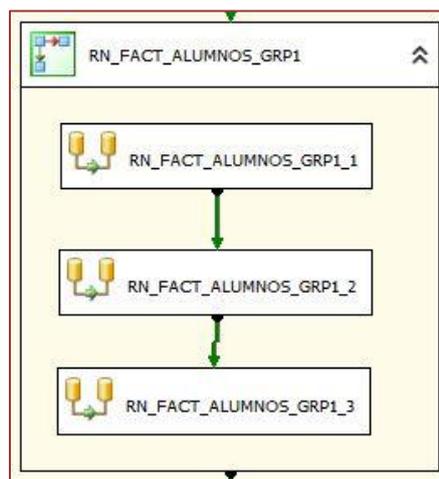
Elaboración: los autores

El contenedor presenta una tarea de flujo de datos, RN_CW_DIM_PROCESO, la cual aplica reglas de negocio a la data para luego cargarla a la dimensión DIM_PROCESO. Para ver el detalle del flujo de datos, **(Ver Anexo 15)**.

c) Contenedor RN_FACT_ALUMNO_GRP1

En este contenedor se presenta la primera parte proceso de ETL de la tabla de hechos FACT_ALUMNO. Este depende del proceso de carga de la dimensión DIM_PROCESO. A continuación, se muestra una imagen de dicho proceso.

Ilustración XXXIII - RN_FACT_ALUMNO_GRP1



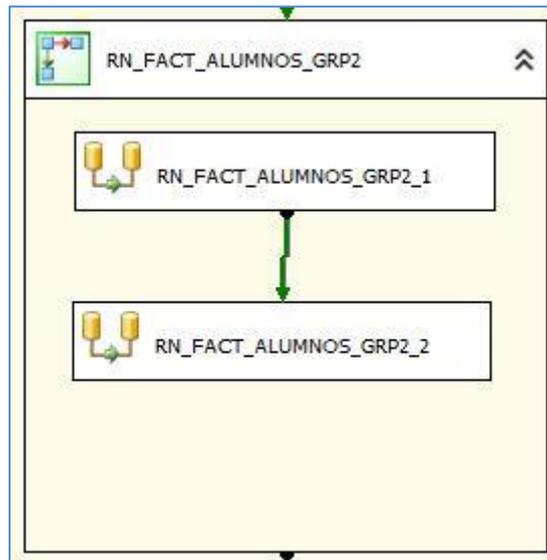
Elaboración: los autores

Se puede observar que se separó el proceso de reglas de negocio en tres tareas de flujo de datos. Para ver el detalle de cada flujo de datos, **(Ver Anexo 15)**.

d) Contenedor RN_FACT_ALUMNO_GRP2

En este contenedor se presenta la segunda parte del proceso de ETL de la tabla de hechos FACT_ALUMNO. Este depende de la correcta ejecución del contenedor anterior, RN_FACT_ALUMNO_GRP1, A continuación, se muestra una imagen de dicho proceso.

Ilustración XXXIV - RN_FACT_ALUMNO_GRP2



Elaboración: los autores

En esta imagen se muestran dos tareas de flujos de datos que aplican reglas de negocio. Para ver el detalle de cada flujo de datos, (**Ver Anexo 15**).

e) Contenedor CW_FACT_ALUMNO

Este contenedor presenta el proceso de carga de la tabla de hechos FACT_ALUMNO. Es necesario que se ejecuten y finalicen todos procesos anteriores antes de ejecutar este. A continuación, se muestra una imagen del proceso.

Ilustración XXXV - CW_FACT_ALUMNO



Elaboración: los autores

Para ver el detalle del flujo de datos, **(Ver Anexo 15)**.

f) Contenedor carga de severidades

El contenedor carga de severidades presenta la tarea de flujo de datos, que puebla la tabla de severidades del Datamart, para las inconsistencias del proceso ETL de la tabla FACT_ALUMNO. A continuación, se muestra una imagen del proceso. **(Ver el Anexo 16)**.

Ilustración XXXVI - Carga de severidades



Elaboración: los autores

Para ver el detalle del flujo de datos, **(Ver Anexo 15)**.

g) Contenedor limpieza de temporales

El contenedor limpieza de temporales presenta el proceso de eliminación de archivos temporales que se hayan creado en los otros contenedores, producto de las múltiples transformaciones de los datos. Este proceso se realiza utilizando dos tareas de sistemas de archivos. A continuación, se muestra una imagen del proceso.

Ilustración XXXVII - Limpieza de temporales



Elaboración: los autores

Para ver el detalle del flujo de datos, **(Ver Anexo 15)**.

4.5.2.4 Paquete MANTENIMIENTO_INDICES.dtsx

Este paquete está compuesto por una tarea que ejecuta un script sql, y dependiendo del estado de los índices, los recrea o los deshabilita. Cuando el paquete se ejecuta antes de iniciar el proceso ETL, los índices se eliminan, lo que permite al proceso ETL ejecutarse con un mejor rendimiento. Cuando el paquete se ejecuta después del proceso ETL, los índices se vuelven a crear dejándolos en su estado inicial. A continuación, se muestra un gráfico del paquete MANTENIMIENTO_INDICES.

Ilustración XXXVIII - Paquete mantenimiento índices

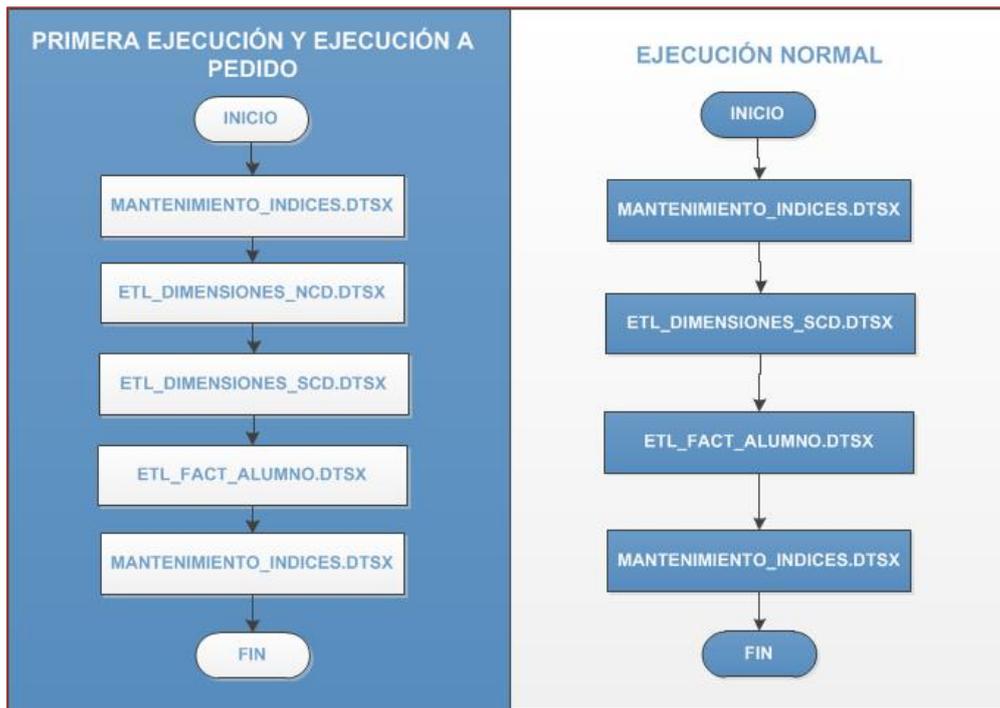


Elaboración: los autores

4.5.3 Diseño de la secuencia de ejecución

La secuencia de ejecución de cada uno de los paquetes ETL está determinada por la dependencia de entre dimensiones y la dependencia entre dimensiones y la tabla de hechos. Para la carga de información al Datamart, se consideran tres escenarios: la primera ejecución, la ejecución normal y la ejecución a pedido. A continuación, se muestra un gráfico con la secuencia de ejecución en cada situación.

Ilustración XXXIX - Secuencia de ejecución paquetes ETL



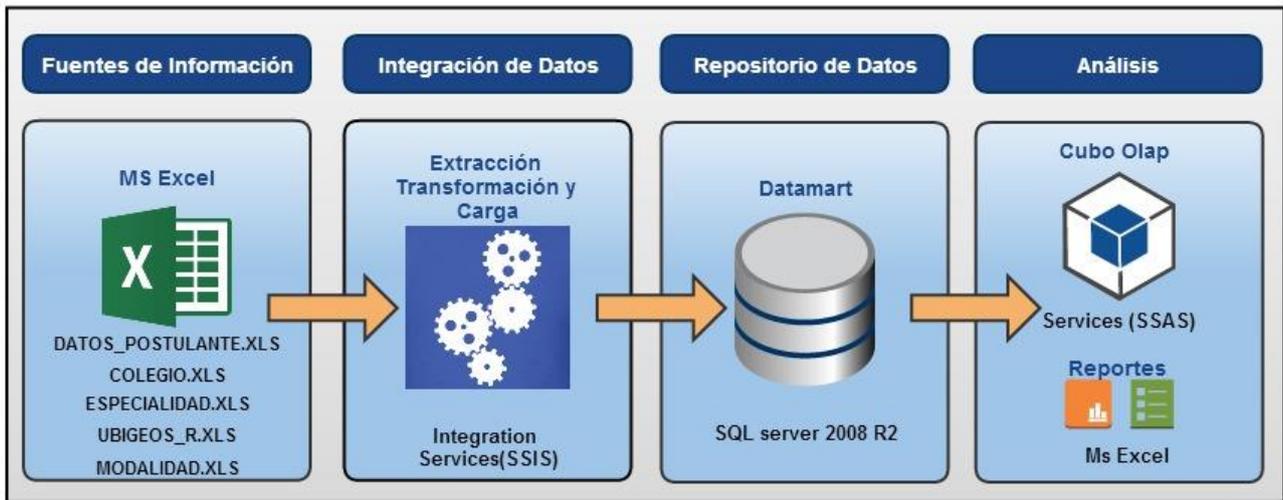
Elaboración: los autores

Cada secuencia de ejecución se agrupó en dos distintos Jobs, el DTMFIASEMESTRAL, que contiene los paquetes para la ejecución normal y el DTMFIA1ERACARGA, que contiene los paquetes para la primera ejecución y ejecución a pedido. Ambos pueden ser invocados desde el Agente SQL Server tanto de forma manual como programando una ejecución periódica. Para efectos de la presente tesis, se decidió ejecutar los jobs de manera manual ya que los calendarios de actividades de la FIA son cambiantes lo cual no nos permitió programarle una periodicidad.

4.6 Diseño de la arquitectura técnica

Evaluamos los componentes de la solución y la manera en que estos interactúan entre sí y desarrollamos la arquitectura de la solución. Podemos observar en la siguiente ilustración, cada componente con su respectiva herramienta y etapa en la que se le ubica dentro de la solución.

Ilustración XL – Arquitectura técnica



Elaboración: los autores

4.7 Selección de los productos e instalación

En esta sección se definieron las herramientas que se utilizaron en la implementación (simulación) del Datamart.

4.7.1 Elaboración de matriz herramienta por actividad

La matriz se desarrolló en el apartado “**3.1 Materiales**”, por favor revisar dicha sección para mayor detalle.

4.7.2 Instalación de herramientas para simulación

En esta actividad se preparó el ambiente necesario para simulación del modelo. Se instaló la Suite Microsoft SQL Server 2008 R2 para el manejo de la Base de Datos junto con las herramientas de Inteligencia de Negocio, SQL Server Integration Services para el ETL y SQL Server Analysis Services para la realización del cubo de información.

Además, se instaló la herramienta ERWIN para el modelamiento de la base de datos, así como también la herramienta Microsoft Excel 2010 para la explotación de los reportes.

4.8 Diseño y desarrollo de aplicaciones BI

A través de herramientas de fácil uso, se le brinda al usuario una manera más fácil de acceder al repositorio de datos, para que pueda hacer las consultas respectivas y pueda además elaborar sus propios reportes.

4.8.1 Diseño y desarrollo de reportes

La explotación se realizó a través de reportes los cuales han sido estructurados con el fin de contener la información que se requiere. Para esta sección se trabajó las especificaciones que rigen para todos los reportes que se elaboren como parte de este modelo.

4.8.1.1 Diseño de reportes

La configuración de los reportes debe cumplir los siguientes lineamientos tales como:

- a) Tamaño de Hoja: A4
- b) Orientación: Vertical u Horizontal (Depende del tipo de reporte)
- c) Márgenes: estrechos
- d) Cabecera para la impresión:

Tabla LXXV - Cabecera para la impresión

Dato	Posición	Tipo Letra/ Tamaño	Color	Formato	Observación
Datamart - FIA	Izquierda Superior	Calibri 11	Negro	Texto	Nombre General del modelo
Nombre de la Hoja	Centro	Calibri 11	Negro	-	Nombre de la Hoja a la que pertenece el reporte
Fecha y Hora	Derecha Superior	Calibri 11	Negro	dd/mm/yyyy hh:mm	Fecha y hora de impresión del reporte

Elaboración: los autores

e) Pie de página para la impresión:

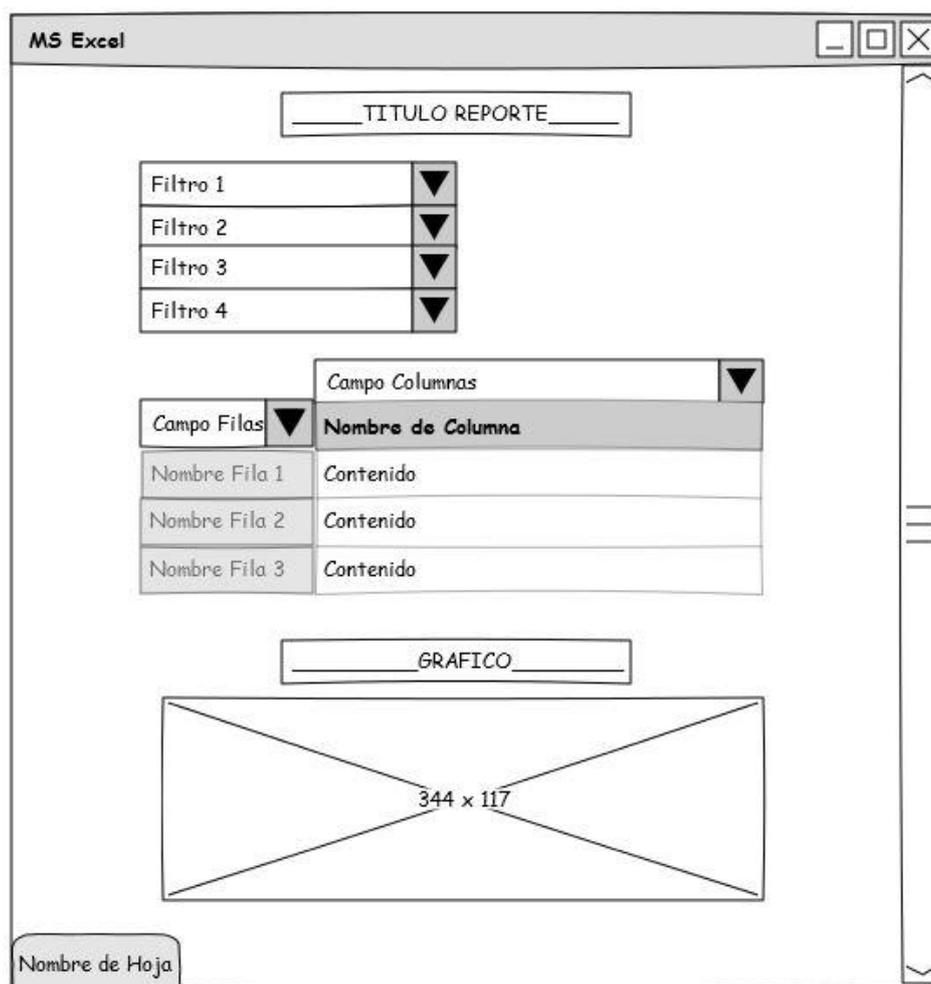
Tabla LXXVI - Pie de página para la impresión

Tipo					
Dato	Posición	Letra/ Tamaño	Color	Formato	Observación
FIA - USMP	Izquierda Inferior	Calibri 11	Negro	Texto	Facultad y Universidad
Numeración	Derecha Inferior	Calibri 11	Negro	# de página 'de' total de paginas	Numero de página del total

Elaboración: los autores

f) Cuerpo del reporte

Ilustración XLI - Cuerpo del reporte base



Elaboración: los autores

4.8.1.2 Desarrollo de reportes

Se elaboraron reportes que para el usuario serán reportes pre - establecidos, los cuales van a poder ser consultados en cualquier momento y actualizados según la carga del cubo.

Tabla LXXVII - Reportes pre-establecidos

REPORTES PRE-ESTABLECIDOS		
Nombre Reporte	Objetivo	Nombre de Hoja
Porcentaje Ingresantes Matriculados y No Matriculados	Medir la cantidad de alumnos que se han matriculado y no matriculado en el presente proceso de admisión	% MT y NMT
Porcentaje Ingresantes Matriculados y No Matriculados por Carrera	Comparar a los ingresantes que concretaron su matrícula vs lo que no se llegaron a matricular por cada carrera + Comparar carreras (Vista Global)	(MT Vs NMT) x Carreras
Cantidad y Porcentajes de alumnos matriculados por Carrera	Medir la Cantidad y Distribución de porcentaje de alumnos matriculados por carrera.	(Nro,% MT) x Carrera
Porcentaje de Matriculados y No Matriculados por Carrera	Medir la Distribución de porcentaje de alumnos matriculados y no matriculados por carrera	(% MT y NMT) x Carrera
Distribución Geográfica de Postulantes Global	Medir la distribución de los postulantes a la facultad por su ubigeo de residencia a Nivel Nacional	Distribución Postulante Nacional
Distribución Geográfica de Postulantes Lima	Medir la distribución de los postulantes a la facultad por su ubigeo de residencia en la Región Lima	Distribución Postulantes Lima
Rangos de Merito VS Rangos de Puntaje	Medir la distribución según el rango de mérito de ingreso con relación a los rangos de los puntajes.	Merito VS Puntaje

Elaboración: los autores

4.8.1.3 Diseño de indicadores de rendimiento (KPI)

A continuación, se muestra el detalle de la especificación de diseño para los KPI.

Tabla LXXVIII - Diseño de indicadores (KPI)

Nombre	Descripción	Formula
Eficiencia_Matricula	% de Alumnos matriculados	Numero de Postulantes (Ingreso:SI, Matricula:SI) /Numero de Postulantes (Ingreso:SI)
Eficiencia_Matricula_Arquitectura	% de Alumnos matriculados en Arquitectura	Numero de Postulantes (Ingreso:SI, Matricula:SI, Carrera:Arquitectura) / Numero de Postulantes (Ingreso:SI,Carrera:Arquitectura)
Eficiencia_Matricula_Aeronauticas	% de Alumnos matriculados en Aeronáuticas	Numero de Postulantes (Ingreso:SI, Matricula:SI, Carrera:Aeronauticas) / Numero de Postulantes (Ingreso:SI,Carrera:Aeronauticas)
Eficiencia_Matricula_Alimentarias	% de Alumnos matriculados en Alimentarias	Numero de Postulantes (Ingreso:SI, Matricula:SI, Carrera:Alimentarias)/Numero de Postulantes(Ingreso:SI,Carrera:Alimentarias)
Eficiencia_Matricula_Civil	% de Alumnos matriculados en Civil	Numero de Postulantes (Ingreso:SI, Matricula:SI, Carrera:Civil)/Numero de Postulantes(Ingreso:SI,Carrera:Civil)
Eficiencia_Matricula_Electronica	% de Alumnos matriculados en Electrónica	Numero de Postulantes (Ingreso:SI, Matricula:SI, Carrera:Electronica)/Numero de Postulantes(Ingreso:SI,Carrera:Electronica)
Eficiencia_Matricula_Industrial	% de Alumnos matriculados en Industrial	Numero de Postulantes (Ingreso:SI, Matricula:SI, Carrera:Industrial)/Numero de Postulantes(Ingreso:SI,Carrera:Industrial)
Eficiencia_Matricula_Sistemas	% de Alumnos matriculados en Sistemas	Numero de Postulantes (Ingreso:SI, Matricula:SI, Carrera:Sistemas)/Numero de Postulantes(Ingreso:SI,Carrera:Sistemas)

AlumnosMatriculados	# de alumnos matriculados	Numero de Postulantes (Ingreso:SI, Matricula:SI)
Matriculados_Arquitectura	# de alumnos matriculados en Arquitectura	Numero de Postulantes (Ingreso:SI, Matricula:SI, Carrera:Arquitectura)
Matriculados_Aeronauticas	# de alumnos matriculados Aeronáuticas	Numero de Postulantes (Ingreso:SI, Matricula:SI, Carrera:Aeronauticas)
Matriculados_Alimentarias	# de alumnos matriculados Alimentarias	Numero de Postulantes (Ingreso:SI, Matricula:SI, Carrera:Alimentarias)
Matriculados_Civil	# de alumnos matriculados Civil	Numero de Postulantes (Ingreso:SI, Matricula:SI, Carrera:Civil)
Matriculados_Electronica	# de alumnos matriculados Electrónica	Numero de Postulantes (Ingreso:SI, Matricula:SI, Carrera:Electronica)
Matriculados_Industrial	# de alumnos matriculados Industrial	Numero de Postulantes (Ingreso:SI, Matricula:SI, Carrera:Industrial)
Matriculados_Sistemas	# de alumnos matriculados Sistemas	Numero de Postulantes (Ingreso:SI, Matricula:SI, Carrera:Sistemas)
Alumnos_NO_Matriculados	# de alumnos que ingresaron y no se matricularon	Numero de Postulantes (Ingreso:SI, Matricula:NO)
Arquitectura_NO_Matriculados	# de alumnos que ingresaron y no se matricularon Arquitectura	Numero de Postulantes (Ingreso:SI, Matricula:NO, Carrera:Arquitectura)
Aeronauticas_NO_Matriculados	# de alumnos que ingresaron y no se matricularon Aeronáuticas	Numero de Postulantes (Ingreso:SI, Matricula:NO, Carrera:Aeronauticas)
Alimentarias_NO_Matriculados	# de alumnos que ingresaron y no	Numero de Postulantes (Ingreso:SI, Matricula:NO, Carrera:Alimentarias)

	se matricularon Alimentarias	
Civil_NO_Matriculados	# de alumnos que ingresaron y no se matricularon Civil	Numero de Postulantes (Ingreso:SI, Matricula:NO, Carrera:Civil)
Electronica_NO_Matriculados	# de alumnos que ingresaron y no se matricularon Electrónica	Numero de Postulantes (Ingreso:SI, Matricula:NO, Carrera:Electronica)
Industrial_NO_Matriculados	# de alumnos que ingresaron y no se matricularon Industrial	Numero de Postulantes (Ingreso:SI, Matricula:NO, Carrera:Industrial)
Sistemas_NO_Matriculados	# de alumnos que ingresaron y no se matricularon Sistemas	Numero de Postulantes (Ingreso:SI, Matricula:NO, Carrera:Sistemas)

Elaboración: los autores

4.8.1.4 Diseño y Desarrollo de Cubos OLAP

A continuación, se detalla la especificación de diseño para el cubo que implementa los requerimientos de usuario.

Tabla LXXIX - Especificación cubo postulantes

Cubo Postulantes			
Data source:		BD_DATAMART_FIA	
Tablas fuente:	DIM_COLEGIO		
	DIM_TIPO_COLEGIO		
	DIM_CARRERA		
	DIM_FACULTAD		
	DIM_SEDE		
	DIM_LETRA_INICIAL_APE		
	DIM_MATRICULA		
	DIM_RANGO_EDAD		
	DIM_RANGO_EGRESO		
	DIM_RANGO_PUNTAJE		
	DIM_RANGO_EDAD		
	DIM_GENERO		
	DIM_INGRESO		
	DIM_MODALIDAD		
	DIM_TIPO_PRUEBA		
	DIM_UBGEO_COLEGIO		
	DIM_UBGEO		
	DIM_TRABAJA		
	DIM_ESTADO_CIVIL		
	DIM_PROCESO		
FACT_ALUMNO			
Dimensiones	Nombre	Nombre del nivel todos	Tabla fuente
	Colegio	Todos los colegios	DIM_COLEGIO
			DIM_TIPO_COLEGIO

	Especialidad	Todas las especialidades	DIM_CARRERA
			DIM_FACULTAD
			DIM_SEDE
	Letra Inicial Apellido	Todos los tipos de letras	DIM_LETRA_INICIAL_APE
	Matricula	Situación matricula	DIM_MATRICULA
	Rango Edad	Todos los rangos de edad	DIM_RANGO_EDAD
	Rango Egreso	Todos los rangos de egreso	DIM_RANGO_EGRESO
	Rango Puntaje	Todos los rangos de puntaje	DIM_RANGO_PUNTAJE
	Rango Merito	Todos los rangos de merito	DIM_RANGO_MERITO
	Genero	Todos los géneros	DIM_GENERO
	Ingreso	Todos los tipos de ingreso	DIM_INGRESO
	Modalidad	Todas las modalidades	DIM_MODALIDAD
			DIM_TIPO_PRUEBA
	Ubigeo Colegio	Todos los ubigeos de colegio	DIM_UBIGEO_COLEGIO
	Ubigeo	Todos los ubigeos	DIM_UBIGEO
	Trabaja	Todas las condiciones	DIM_TRABAJA
	Estado Civil	Todos los estados civiles	DIM_ESTADO_CIVIL
Proceso de Admisión	Todos los procesos	DIM_PROCESO	
Alumno	-	FACT_ALUMNO	
Medidas	Nombre	Campo fuente	Función
	Máxima edad	FACT_ALUMNO.EDAD	Valor máx.
	Máximo puntaje	FACT_ALUMNO.PUNTAJE	Valor máx.
	Máximo merito	FACT_ALUMNO.MERITO	Valor máx.
	Numero de postulantes	FACT_ALUMNO.ID_ALUMNO	Contar
	Mínima edad	FACT_ALUMNO.EDAD	Valor min.
	Mínima puntaje	FACT_ALUMNO.PUNTAJE	Valor min.
Tipo almacenamiento:	MOLAP		
Lectura/Escritura:	Lectura		

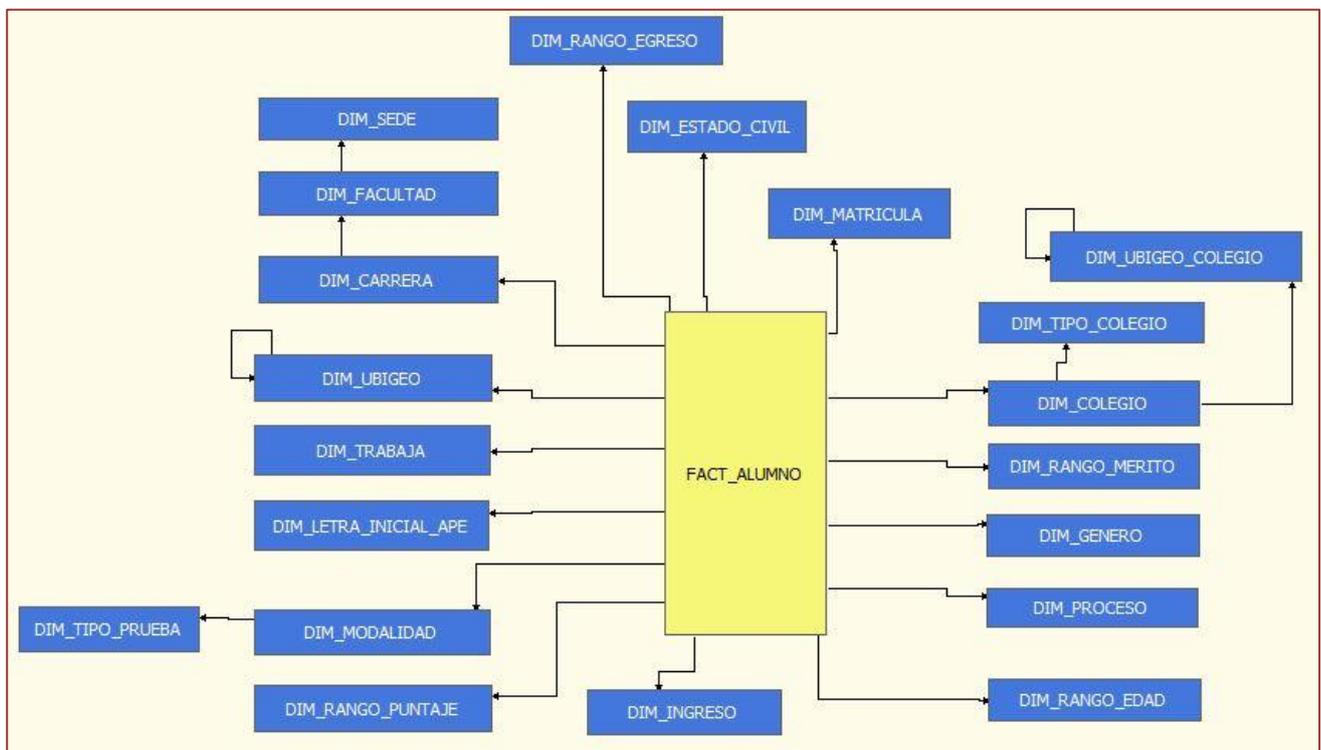
Elaboración: los autores

La Dimensión Alumno (DIM_ALUMNO), se creó únicamente para realizar el *drill through*, y poder acceder a los detalles de las consultas. En la explotación de información existe también el concepto de ir al detalle de la información, es decir, partiendo de los cruces de varias dimensiones en cierto nivel de un cubo, y tomándolas como filtros se puede ir a la información relacional que sustenta los agregados.

Para ello definimos los datos del alumno que se desean obtener de la base de datos como resultado de la consulta y su relación con las dimensiones del Datamart. Para ello a la propiedad *AttributeHierarchyVisible* de todos los atributos de la dimensión Alumno se le colocó False, menos al campo KEY que se le deja habilitado para poder obtener a través de él los datos.

El diseño del cubo nos brindó los lineamientos para el desarrollo de este, las relaciones entre las dimensiones se mantienen. Si observamos detenidamente en la Ilustración XLI podemos darnos cuenta de que la dimensión Alumno (DIM_ALUMNO), no se encuentra, y el motivo es que esta dimensión, como se explicó anteriormente, es utilizada para realizar el detalle de las consultas, por consiguiente, es una dimensión invisible en el modelo.

Ilustración XLII - Desarrollo del cubo OLAP



Elaboración: los autores

4.9 Implementación (simulación)

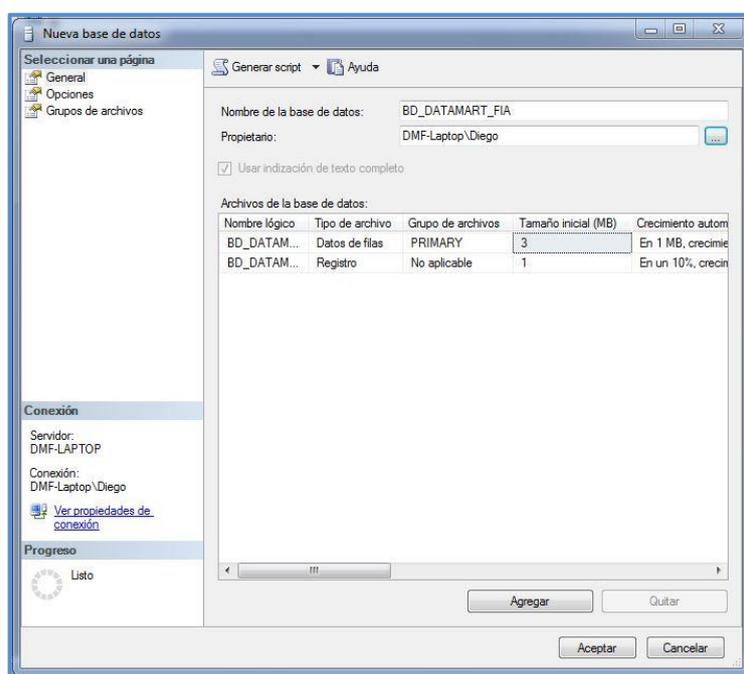
A continuación, se presenta la simulación de la implementación del datamart que se llevó a cabo utilizando una laptop, donde se instalaron las herramientas necesarias para crear la base de datos, ejecutar los procesos ETL, crear el cubo de información y generar los reportes e indicadores.

La simulación se realizó tomando como input los datos de postulantes, ingresantes y matriculados para el periodo 2014-I. Debido a políticas de confidencialidad sobre la información académica de la FIA, se tuvieron que recrear los datos con el objetivo de poder simular la carga al datamart.

4.9.1 Creación de Base de Datos

La base de datos del datamart se creó utilizando la herramienta SQL Server Management Studio, bajo el nombre de BD_DATAMART_FIA, como se puede apreciar en la siguiente imagen.

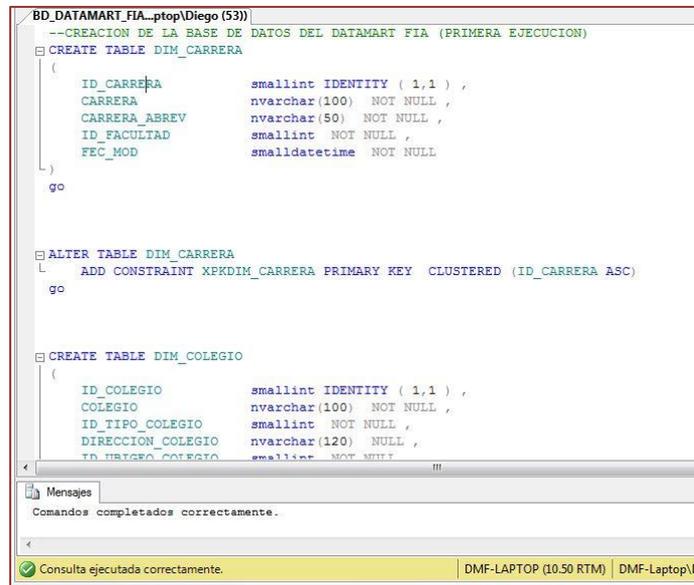
Ilustración XLIII - Creación de base de datos



Elaboración: los autores

Luego se ejecutó el script BD_DATAMART_FIA_1RA_EJEC.SQL para la creación de tablas, índices, restricciones, funciones, etc. Además, se ejecuta el script SEVERIDADES.SQL para la creación de las tablas de severidades.

Ilustración XLIV - Ejecución del script BD_DATAMART_FIA_1RA_EJEC.SQL



```
BD_DATAMART_FIA_ptop\Diego (53)
--CREACION DE LA BASE DE DATOS DEL DATAMART FIA (PRIMERA EJECUCION)
CREATE TABLE DIM_CARRERA
(
    ID_CARRERA          smallint IDENTITY ( 1,1 ) ,
    CARRERA             nvarchar(100) NOT NULL ,
    CARRERA_ABREV       nvarchar(50)  NOT NULL ,
    ID_FACULTAD         smallint NOT NULL ,
    FEC_MOD             smalldatetime NOT NULL
)
go

ALTER TABLE DIM_CARRERA
ADD CONSTRAINT XPKDIM_CARRERA PRIMARY KEY CLUSTERED (ID_CARRERA ASC)
go

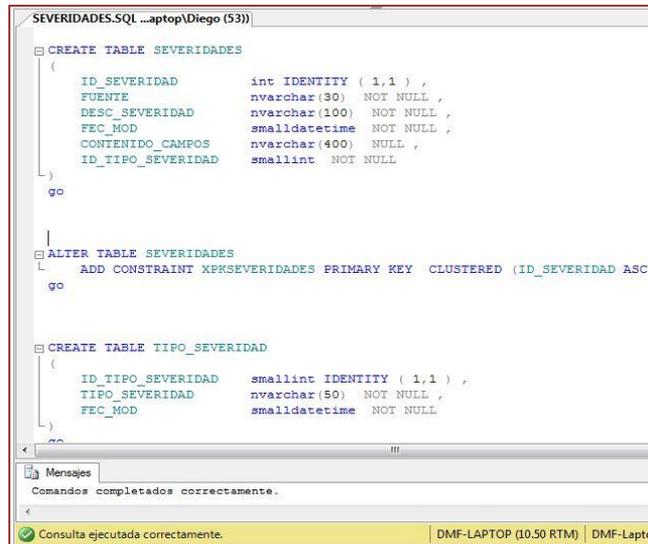
CREATE TABLE DIM_COLEGIO
(
    ID_COLEGIO          smallint IDENTITY ( 1,1 ) ,
    COLEGIO             nvarchar(100) NOT NULL ,
    ID_TIPO_COLEGIO     smallint NOT NULL ,
    DIRECCION_COLEGIO  nvarchar(120) NULL ,
    ID_ORIGEN_COLEGIO   smallint NOT NULL
)
go
```

Mensajes
Comandos completados correctamente.

Consulta ejecutada correctamente. DMF-LAPTOP (10.50 RTM) DMF-LaptopD

Elaboración: los autores

Ilustración XLV - Ejecución del script SEVERIDADES.SQL



```
SEVERIDADES.SQL _aptop\Diego (53)
CREATE TABLE SEVERIDADES
(
    ID_SEVERIDAD        int IDENTITY ( 1,1 ) ,
    FUENTE              nvarchar(30) NOT NULL ,
    DESC_SEVERIDAD      nvarchar(100) NOT NULL ,
    FEC_MOD             smalldatetime NOT NULL ,
    CONTENIDO_CAMPOS   nvarchar(400) NULL ,
    ID_TIPO_SEVERIDAD  smallint NOT NULL
)
go

ALTER TABLE SEVERIDADES
ADD CONSTRAINT XPKSEVERIDADES PRIMARY KEY CLUSTERED (ID_SEVERIDAD ASC)
go

CREATE TABLE TIPO_SEVERIDAD
(
    ID_TIPO_SEVERIDAD  smallint IDENTITY ( 1,1 ) ,
    TIPO_SEVERIDAD     nvarchar(50) NOT NULL ,
    FEC_MOD            smalldatetime NOT NULL
)
go
```

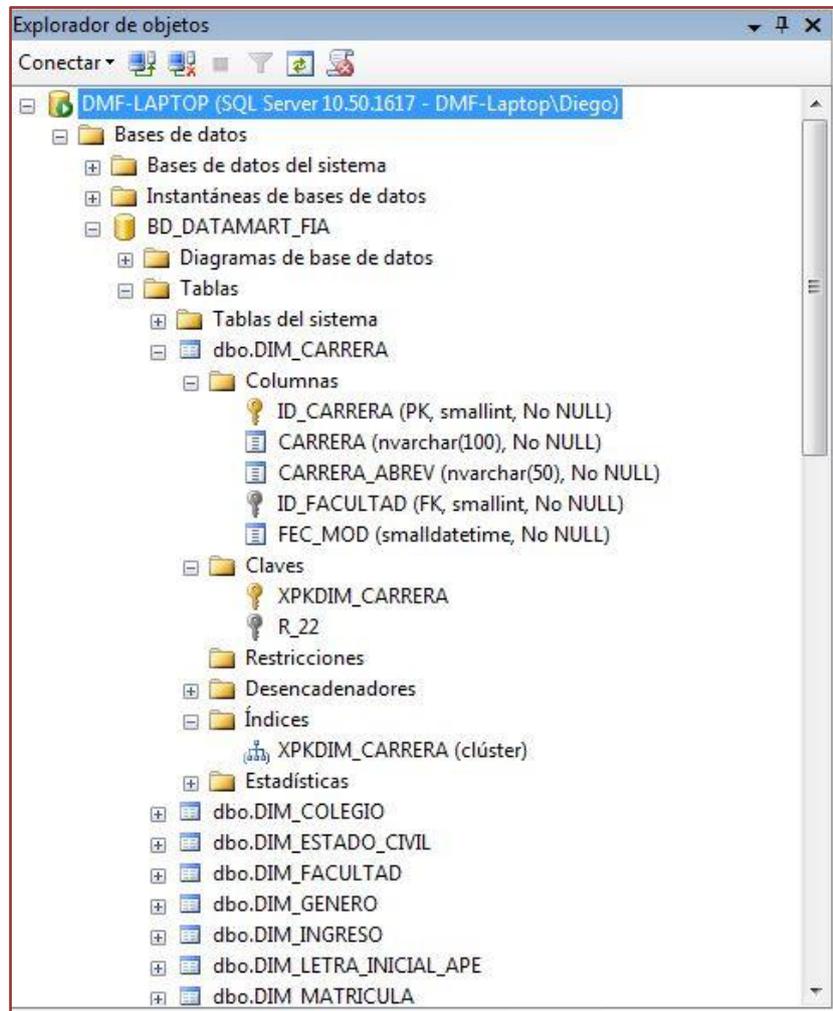
Mensajes
Comandos completados correctamente.

Consulta ejecutada correctamente. DMF-LAPTOP (10.50 RTM) DMF-LaptopD

Elaboración: los autores

Una vez ejecutado el script, se verifica la creación de tablas y columnas para comprobar la correcta creación de los objetos de base de datos.

Ilustración XLVI - Explorador de objetos



Elaboración: los autores

Finalmente, la base de datos, BD_DATAMART_FIA, ya está lista para comenzar con los procesos de extracción transformación y carga.

4.9.2 Ejecución de procesos ETL

La creación y ejecución de los procesos ETL se realizó mediante la herramienta SQL Server Integration Services. El proceso se dividió en cuatro paquetes ETL que se ejecutaron de manera secuencial. Cada paquete tiene asignado un conjunto de contenedores que a su vez contienen las diversas tareas que ejecutaron.

El primer paquete que se ejecutó fue el MANTENIMIENTO_INDICES.dtsx. Este paquete contiene una tarea que ejecuta el script SQL MANTENIMIENTO_INDICES.SQL. La ejecución de este script elimina los índices de todas las tablas de la base de datos, con el objetivo de mejorar el rendimiento de la inserción de datos a las tablas finales. A continuación, se muestra una imagen del paquete que se ejecutó en la simulación.

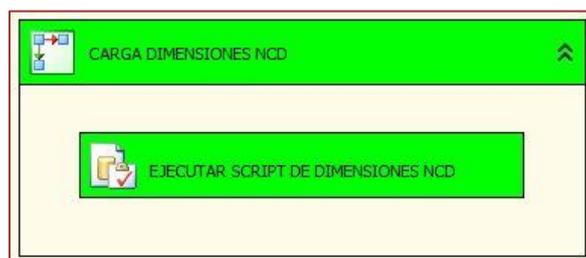
Ilustración XLVII - Mantenimiento índices



Elaboración: los autores

El siguiente paso fue ejecutar el paquete ELT_DIMENSIONES_NCD.dtsx para cargar las dimensiones NCD. De la misma manera, ejecuta una tarea que llama al script CARGA_DIMENSIONES_NCD.SQL.

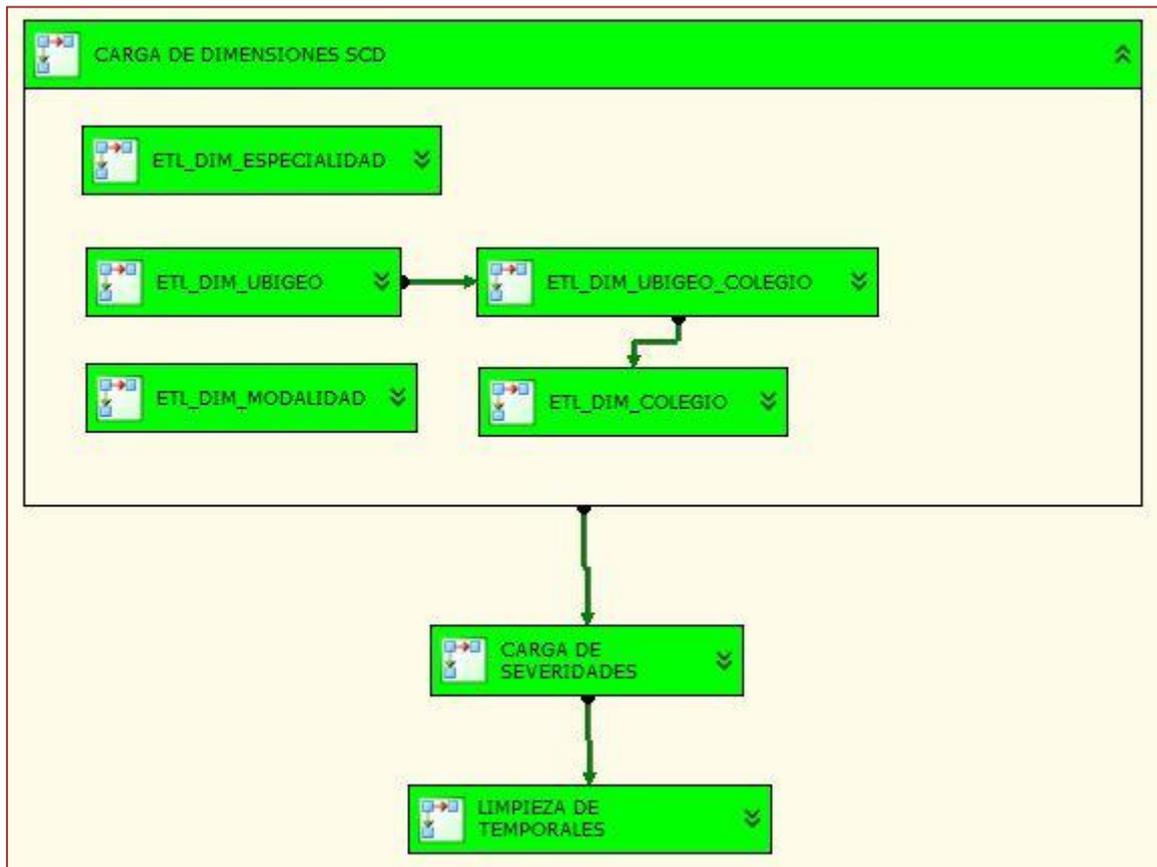
Ilustración XLVIII - Ejecución de dimensiones NCD



Elaboración: los autores

Una vez que se ejecutaron los dos paquetes anteriores, se ejecutó el paquete ELT_DIMENSIONES_SCD.dtsx. Este paquete contiene el proceso ETL para poblar diversas dimensiones como DIM_ESPECIALIDAD, DIM_COLEGIO, etc. Además, contiene un proceso de carga de severidades y uno de limpieza de temporales. En la imagen siguiente se puede observar la correcta ejecución del proceso.

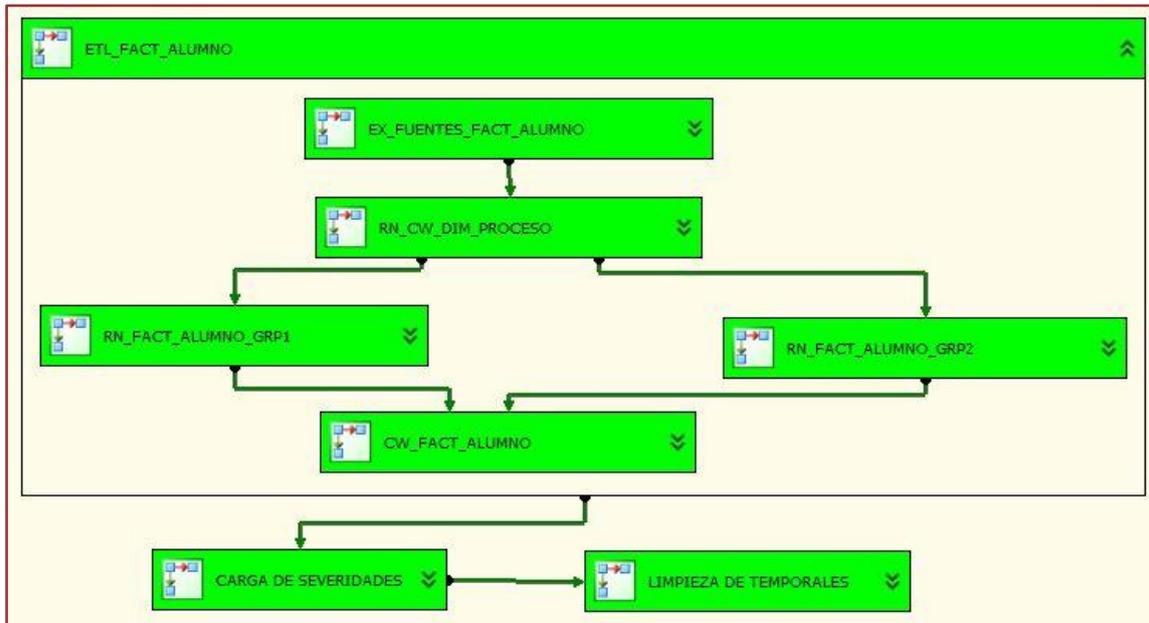
Ilustración XLIX - ETL dimensiones SCD



Elaboración: los autores

Posterior a la carga de todas las dimensiones, es decir las tablas independientes del modelo, se ejecutó el paquete ETL_FACT_ALUMNO que contiene el proceso de carga de la tabla de hechos FACT_ALUMNO y la DIM_PROCESO. Este paquete también contiene el proceso de carga de severidades y de limpieza de temporales.

Ilustración L - ETL FACT_ALUMNO



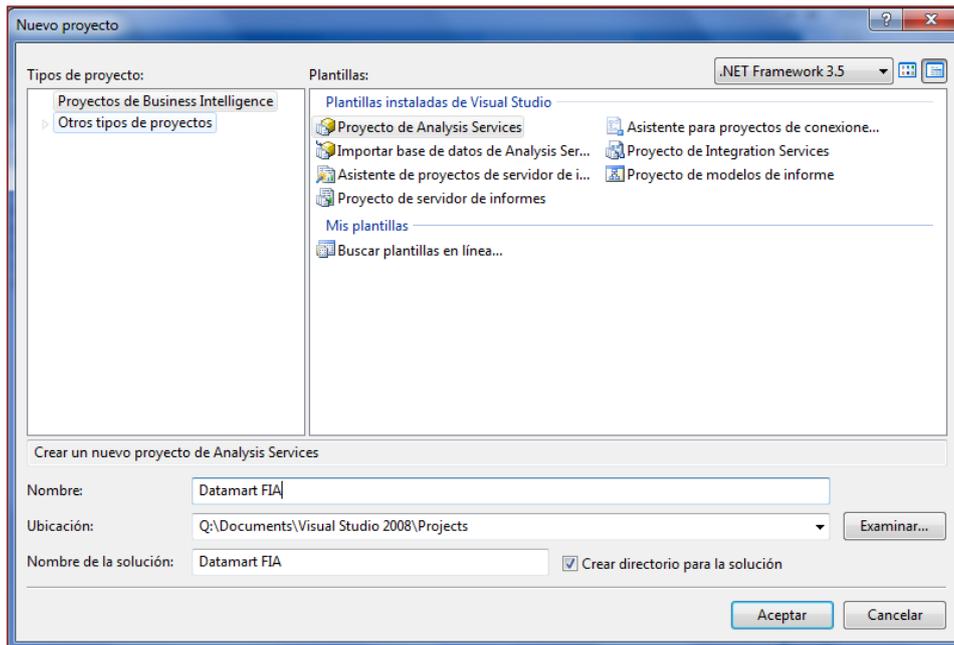
Elaboración: los autores

En este punto ya se cargaron las dimensiones y la tabla de hechos. La base de datos esta lista para ser consumida por el cubo de información que se crea a continuación.

4.9.3 Creación del cubo de información

Para la creación del Cubo utilizamos la herramienta MS Visual Studio diseñado en el punto 3.8.2 Diseño y Desarrollo de Cubos OLAP de la presente tesis. Seleccionamos Nuevo proyecto y elegimos la opción Proyecto de Analysis Services ubicado en el contenedor Plantillas, y le colocamos el nombre del modelo Datamart FIA.

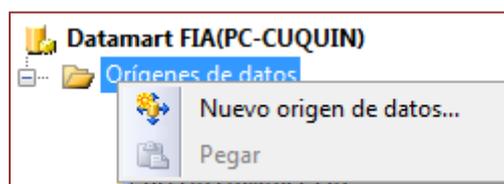
Ilustración LI - Nuevo proyecto SSAS



Elaboración: los autores

Luego de haber creado el proyecto, damos clic derecho en la carpeta origen de datos y seleccionamos nuevo origen de datos, esta conexión, nos va a permitir conectarnos con la base datos

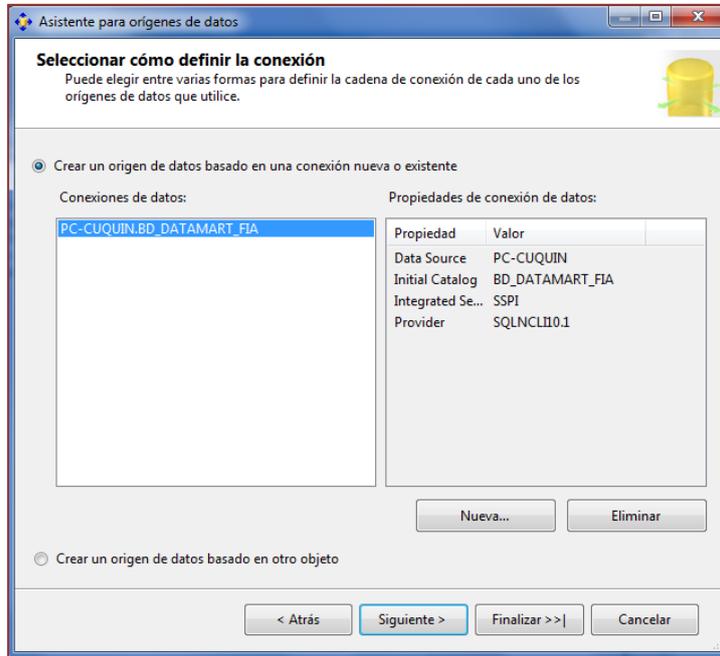
Ilustración LII – Nuevo origen de datos



Elaboración: los autores

Seleccionamos el origen de datos, la conexión existente, que es la conexión a la base de datos BD_DATAMART_FIA.

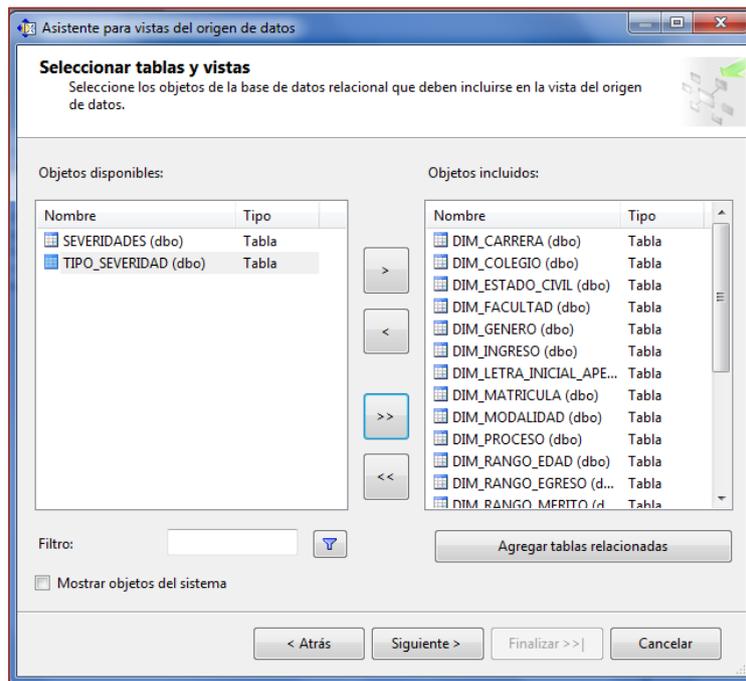
Ilustración LIII - Definir conexión



Elaboración: los autores

Lo siguiente que hicimos fue crear una vista de origen de datos. Una vista de origen de datos representa un subconjunto de datos de un origen de datos, para lo cual damos clic derecho en vistas de origen de datos y seleccionamos crear. Posteriormente seleccionamos todas las tablas que son consultadas por el cubo, las colocamos en el contenedor objetos incluidos y agregamos la relaciones a las tablas con el botón que se ubica en la parte inferior.

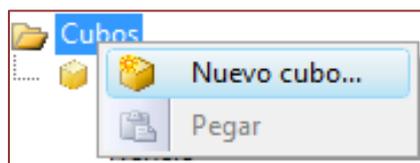
Ilustración LIV - Seleccionar tablas y vistas



Elaboración: los autores

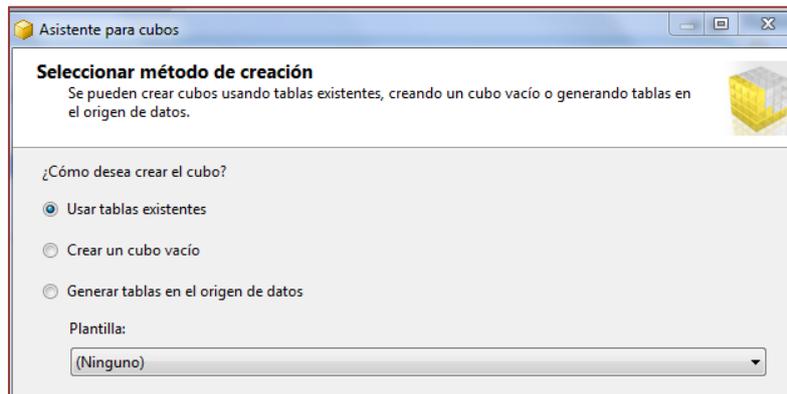
El cubo fue llamado POSTULANTES, para la creación del cubo se utilizaron, tablas previamente creadas en el la base de datos, para lo cual se seleccionó la opción tablas existentes y posteriormente las respectivas tablas que se utilizarían en el cubo.

Ilustración LV - Nuevo cubo



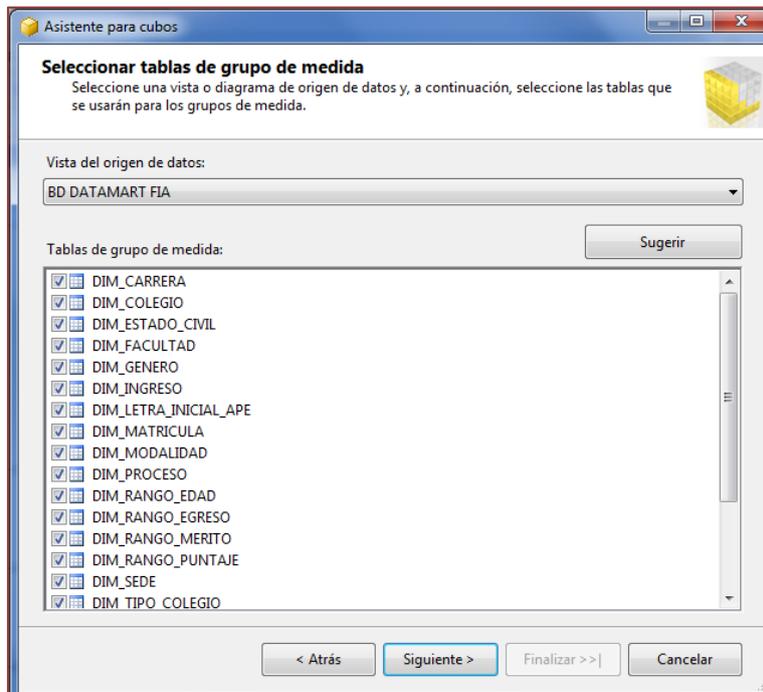
Elaboración: los autores

Ilustración LVI – Seleccionar método de creación de cubo



Elaboración: los autores

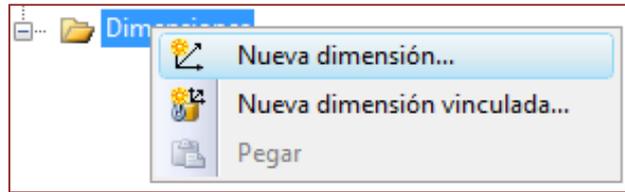
Ilustración LVII – Seleccionar tablas para el cubo



Elaboración: los autores

El siguiente paso fue crear las dimensiones que ya habían sido definidas en el punto “4.3.3. Elección y detalle de las dimensiones”.

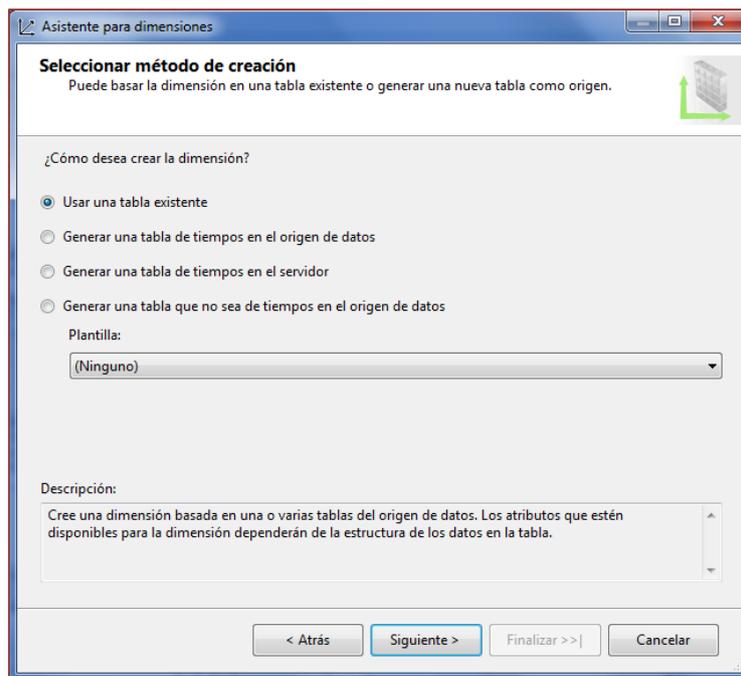
Ilustración LVIII – Crear dimensión



Elaboración: los autores

Como método de creación de las dimensiones Utilizamos las tablas existentes de la Base de datos.

Ilustración LIX – Seleccionar método de creación de dimensión



Elaboración: los autores

Seleccionamos la tabla que nos servirá como origen de datos para la dimensión creamos, en este caso hemos seleccionado la tabla DIM_CARRERA, y elegimos el campo clave de la dimensión ID_CARRERA, luego en el siguiente cuadro se mostró las tablas relacionadas con la tabla DIM_CARRERA.

Ilustración LX – Especificar información de origen

Asistente para dimensiones

Especificar información de origen
Seleccione un origen de datos y especifique cómo se enlaza la dimensión al mismo.

Vista del origen de datos:
BD DATAMART FIA

Tabla principal:
DIM_CARRERA

Columnas de clave:
ID_CARRERA
(Agregar columna de clave)

Columna de nombre:
ID_CARRERA

< Atrás Siguiente > Finalizar >> | Cancelar

Elaboración: los autores

Ilustración LXI – Tablas relacionadas

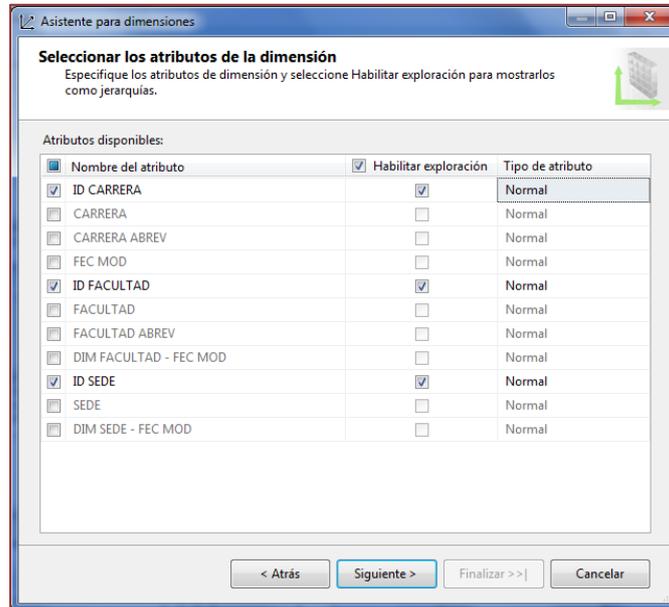
Tablas relacionadas:

- DIM_FACULTAD
- DIM_SEDE

Elaboración: los autores

Especificamos los atributos de la dimensión que estamos creando para que se muestren como jerarquías.

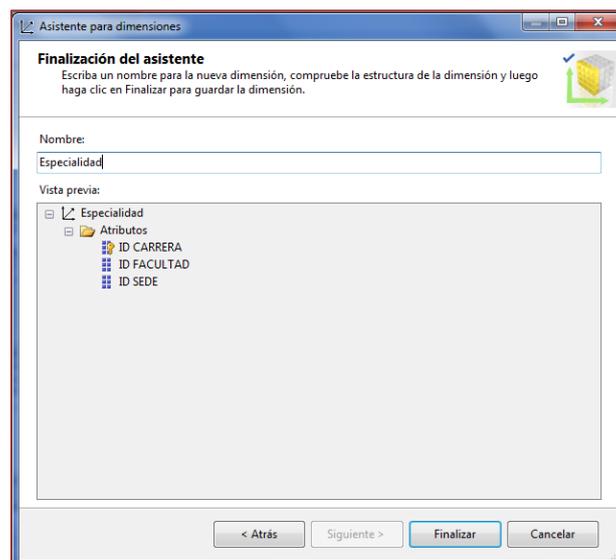
Ilustración LXII – Seleccionar atributos de la dimensión



Elaboración: los autores

Finalmente, el asistente muestra la estructura de la dimensión, antes de ser creada le cambiamos el nombre a la dimensión y le colocamos el nombre Especialidad.

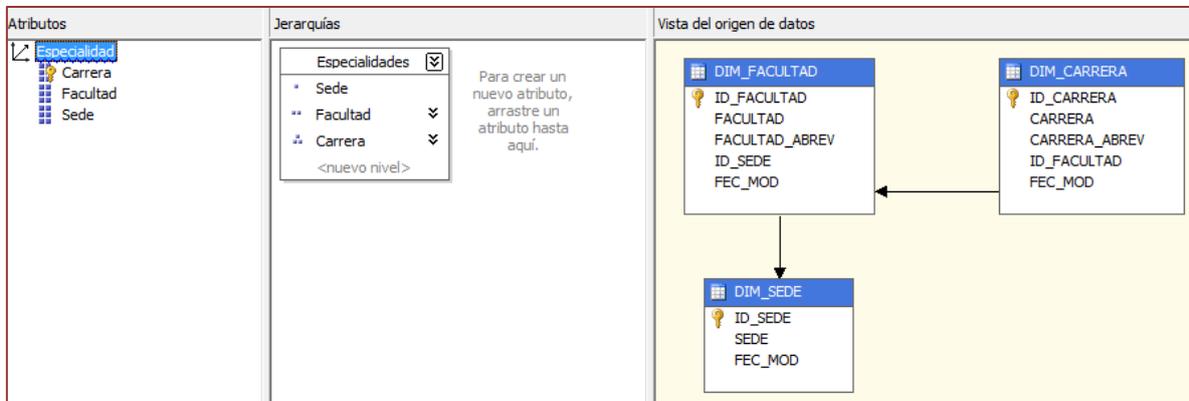
Ilustración LXIII – Estructura de la dimensión



Elaboración: los autores

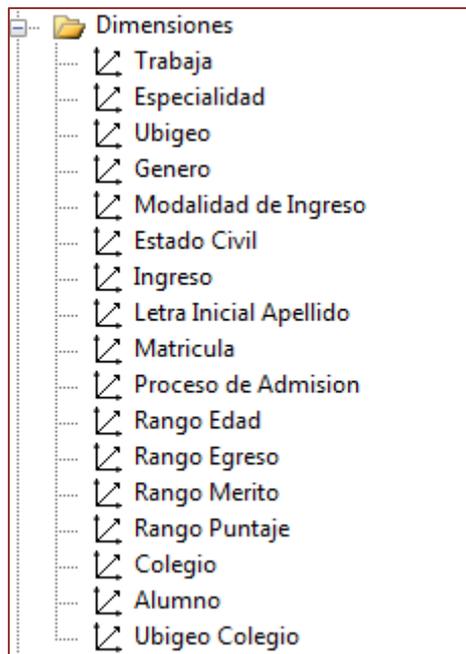
Al finalizar la creación de cada dimensión se muestran los atributos y la vista de origen donde también vemos las relaciones de las tablas, luego pasamos a elaborar las respectivas jerarquías para la dimensión.

Ilustración LXIV – Vista de dimensión



Elaboración: los autores

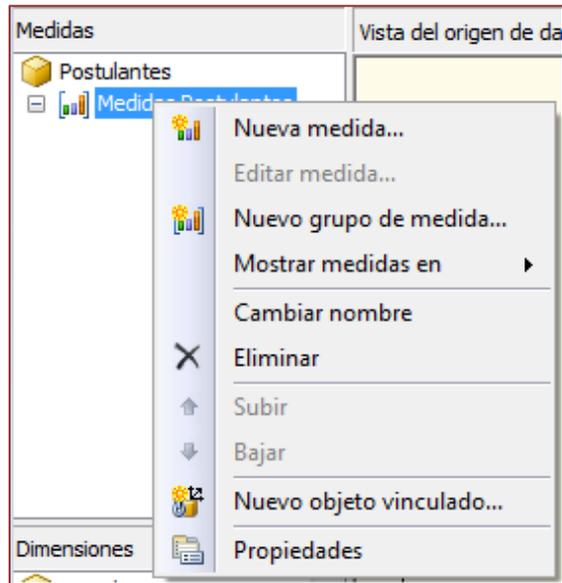
Ilustración LXV – Vista general de dimensiones



Elaboración: los autores

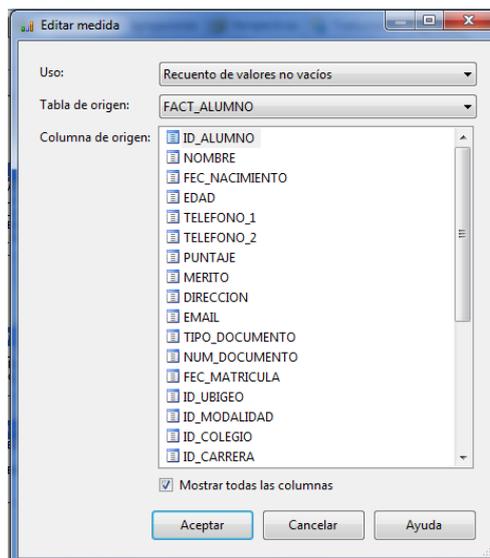
Para crear las medidas del cubo, seleccionamos la ficha estructura de cubo del diseñador de cubos, luego clic con el botón secundario en medidas de postulantes y nueva medida, en este caso mostramos la creación de la medida número de postulantes, y luego la edición de esta.

Ilustración LXVI – Nueva medida



Elaboración: los autores

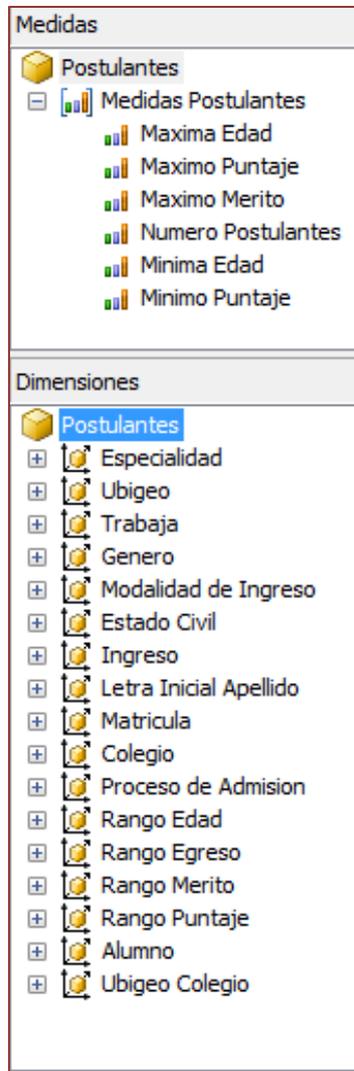
Ilustración LXVII – Editar medida



Elaboración: los autores

Luego de que hemos creado las diferentes medidas, podemos ver en la siguiente ilustración la estructura del cubo en conjunto con las medidas y las dimensiones.

Ilustración LXVIII – Estructura del cubo



Elaboración: los autores

Luego de creados los cubos, se debe relacionar cada tabla de hechos con sus dimensiones respectivas. Abrir el detalle del cubo, ubicarse en la ficha “Dimension Usage”, dar clic sobre el cuadrado plomo y agregar la relación entre la dimensión y tabla de hechos. Realizar lo mismo para todos los cubos

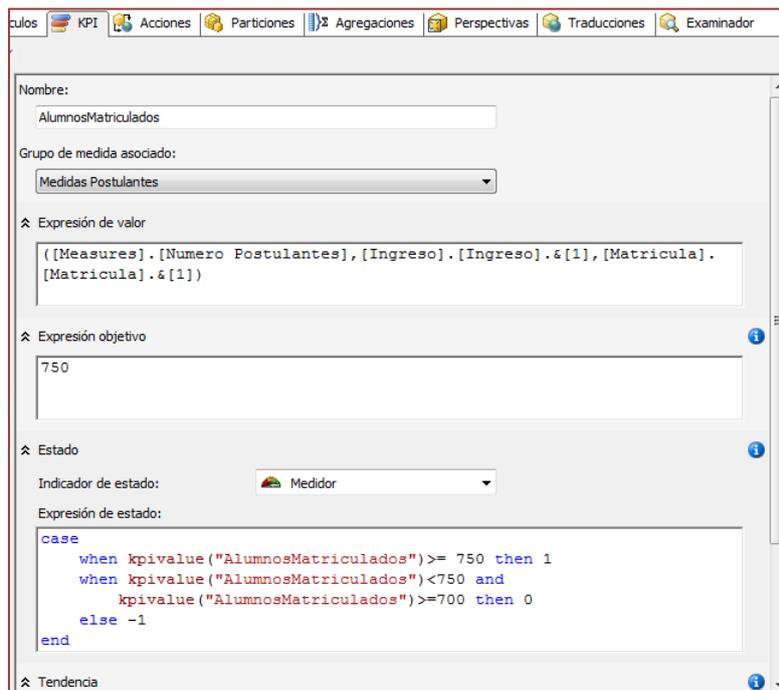
Ilustración LXIX - Relacionar dimensiones con tablas



Elaboración: los autores

Para empezar a crear los KPI, seleccionamos la ficha KPI del diseñador de cubos, clic en nuevo KPI, y colocamos el nombre, el grupo de medidas asociado, la expresión de valor, la expresión objetivo si es que la tiene y la expresión de valor del estado y la expresión de estado que nos permitirá ver los indicadores. Para mayor detalle de los indicadores (**Ver Anexo 17**).

Ilustración LXXI – Creación de indicadores (KPI)



Elaboración: los autores

Ilustración LXXII – KPI implementados

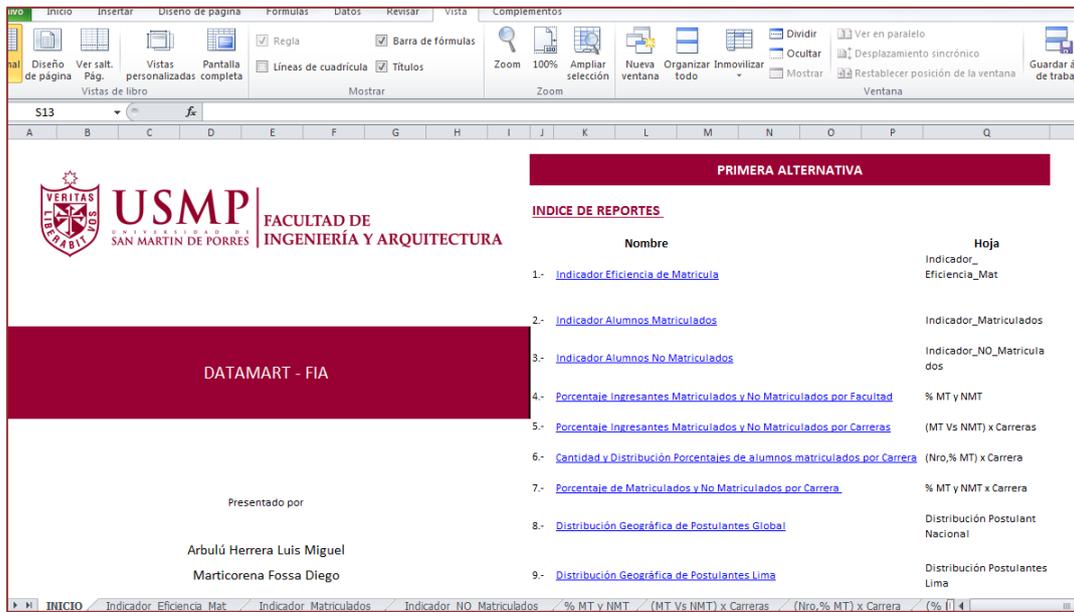
Mostrar estructura	Valor	Objetivo	Estado
KPI			
Aeronauticas_NO_Matriculados	11		
Alimentarias_NO_Matriculados	24		
Alumnos_NO_Matriculados	120	100	
AlumnosMatriculados	730	750	
Arquitectura_NO_Matriculados	23		
Civil_NO_Matriculados	17		
Eficiencia_Matricula	0,86	0,85	
Eficiencia_Matricula_Aeronauticas	0,9		
Eficiencia_Matricula_Alimentarias	0,78		
Eficiencia_Matricula_Arquitectura	0,82		
Eficiencia_Matricula_Civil	0,87		
Eficiencia_Matricula_Electronica	0,86		
Eficiencia_Matricula_Industrial	0,89		
Eficiencia_Matricula_Sistemas	0,88		
Electronica_NO_Matriculados	15		

Elaboración: los autores

4.9.4 Generación de reportes e indicadores

Para poder generar reportes, ingresamos al MS Excel, en este caso al archivo diseñado que tiene el nombre Reporte_Primer_Alt. Observamos la página de inicio de los reportes, donde se encuentra el índice de reportes.

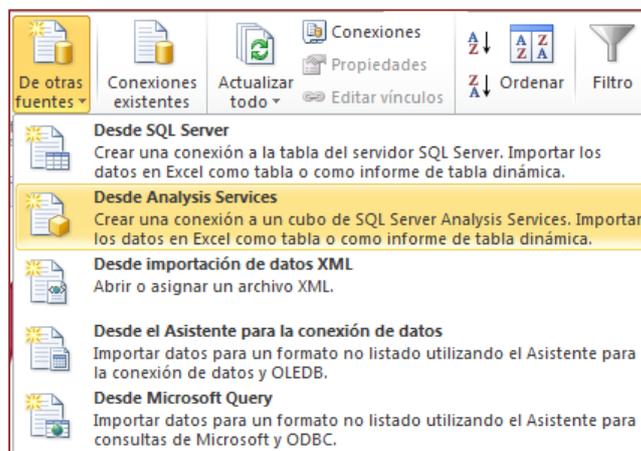
Ilustración LXXIII – Pagina inicial de reportes



Elaboración: los autores

Previamente para poder crear los reportes se tuvo que establecer conexión con el cubo de información del servidor, esta conexión fue creada en la opción Analysis Services de otras fuentes de la ficha datos

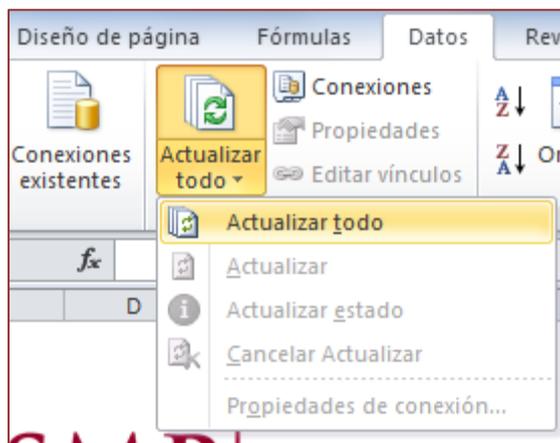
Ilustración LXXIV – Crear conexión con cubo OLAP



Elaboración: los autores

Si estamos ingresando al Excel para la consulta de reportes, antes de actualizar los reportes, debemos ingresar a la ficha datos para establecer nuevamente la conexión con la base de datos y cubo de información para actualizar los reportes, y luego de ello vamos a alguna de las pestañas que contenga reportes.

Ilustración LXXV – Actualizar conexión con cubo OLAP



Elaboración: los autores

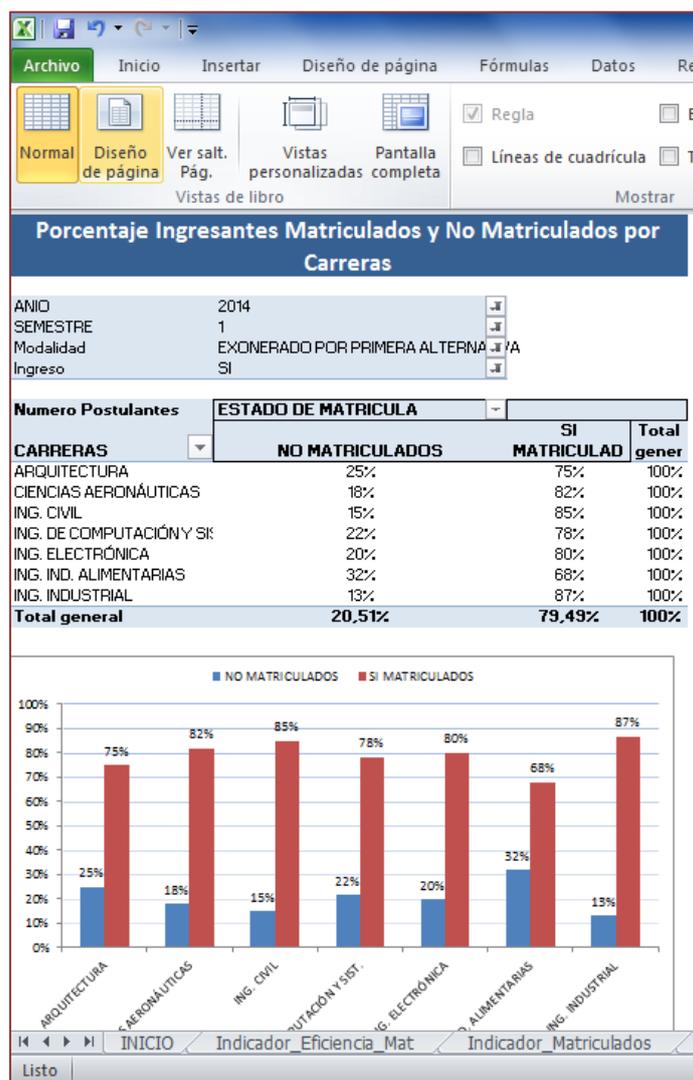
Estando en el reporte, seleccionamos el filtro de fecha que necesitamos (se ubican en la parte superior a los reportes) y la tabla con la conexión al cubo de información, refrescara los datos y el gráfico correspondiente al reporte de la hoja donde nos encontremos. En esta ocasión utilizamos el filtro 2014 – I.

Ilustración LXXVI - Filtros de reportes

Porcentaje Ingresantes Matriculados y No Matriculados por Carreras		
ANIO	2014	[Filtro]
SEMESTRE	1	[Filtro]
Modalidad	EXONERADO POR PRIMERA ALTERNANZA	[Filtro]
Ingreso	SI	[Filtro]

Elaboración: los autores

Ilustración LXXVII – Reporte ejemplo



Elaboración: los autores

En el caso de los indicadores solo necesitan actualizar la fecha de la hoja Indicador_Eficiencia_Mat para poder refrescar los estados de las dos hojas de indicadores automáticamente.

Ilustración LXXVIII – Filtro de KPI's

EFICIENCIA DE MATRICULA		
AÑO	2014	<input type="button" value="▼"/>
SEMESTRE	1	<input type="button" value="▼"/>
Eficiencia_Matricula	Objetivo Eficiencia_Matricul:	Estado Eficiencia_Matricula FIA
86%	85%	↑
EFICIENCIA DE MATRICULA	Arquitectura	Estado Eficiencia_Matricula_Arquitectura
	82%	▲
	Ciencias Aeronauticas	Estado Eficiencia_Matricula_Aeronauticas
	90%	●
	Ing. Civil	Estado Eficiencia_Matricula_Civil
	87%	●
	Ing. Sistemas	Estado Eficiencia_Matricula_Sistemas
	88%	●
Ing. Electronica	Estado Eficiencia_Matricula_Electronica	
86%	●	
Ing. Industrial	Estado Eficiencia_Matricula_Industrial	
89%	●	
Ing. Industrias Alimentarias	Estado Eficiencia_Matricula_Alimentarias	
78%	▲	

Elaboración: los autores

Como observamos en la siguiente ilustración la pestaña que contiene a los indicadores de matriculados, nos muestra filtros de fechas, y esto se debe a que se actualiza como se menciona anteriormente con la actualización de fecha de la hoja Indicador_Eficiencia_Mat.

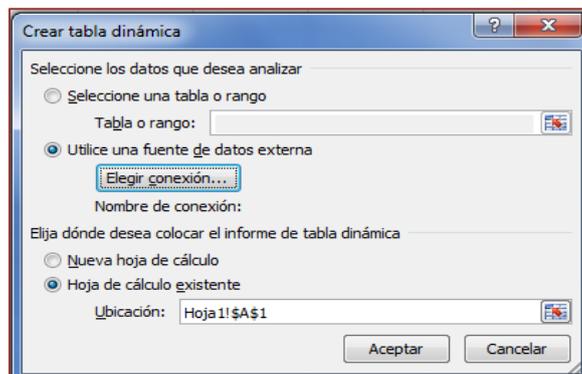
Ilustración LXXIX – Actualización de KPI's

ALUMNOS MATRICULADOS		
AlumnosMatriculados	Objetivo AlumnosMatriculados	Estado AlumnosMatriculados
730	752	→
ALUMNOS MATRICULADOS	Arquitectura	Estado Matriculados_Arquitectur
	107	●
	Ciencias Aeronauticas	Estado Matriculados_Aeronautic
	99	▲
	Ing. Civil	Estado Matriculados_Civil
	113	●
	Ing. Sistemas	Estado Matriculados_Sistemas
	114	●
Ing. Electronica	Estado Matriculados_Electronica	
95	▲	
Ing. Industrias Alimentarias	Estado Matriculados_Alimentari	
86	◆	
Ing. Industrial	Estado Matriculados_Industrial	
116	●	

Elaboración: los autores

Para realizar reportes personalizados lo único que se tiene que hacer es ir a una hoja nueva y desde la ficha insertar, agregar una tabla o gráfico dinámico y seleccionamos el cubo de información dentro de la opción utilice una fuente de datos externa.

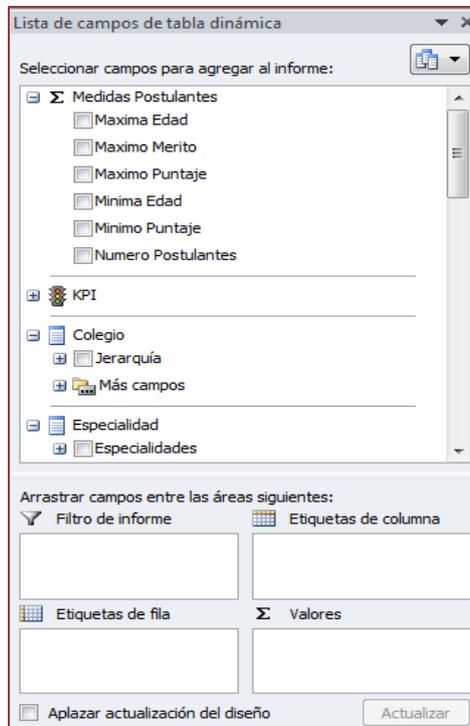
Ilustración LXXX – Creación de tabla dinámica



Elaboración: los autores

Ya con la tabla dinámica podemos realizar el reporte que deseamos, nos aparecerán en la parte derecha todas las dimensiones con sus respectivas jerarquías, medidas y KPI's que se encuentran el cubo.

Ilustración LXXXI – Área de trabajo tabla o gráfico dinámico



Elaboración: los autores

CAPÍTULO V

PRUEBAS Y RESULTADOS

5.1 Caso de prueba #1 – ETL_FACT_ALUMNO

Tabla LXXX - Caso prueba #1

CASO DE PRUEBA #1 - ETL_FACT_ALUMNO				
OBJETO A ANALIZAR	Proceso ETL_FACT_ALUMNO			
OBJETIVO	Verificar que la tabla FACT_ALUMNO sea cargada correctamente por medio del proceso ETL_FACT_ALUMNO, teniendo como fuente el archivo Excel DATOS_POSTULANTE.XLS y el Excel DATOS_MARICULA.XLS			
PRE-REQUISITOS	<ul style="list-style-type: none"> Disponibilidad de la fuente DATOS_POSTULANTE.XLS Disponibilidad de la fuente DATOS_MARICULA.XLS Disponibilidad de la base de datos BD_DATAMART_FIA Dimensiones del Datamart cargadas y operativas 			
DATOS DE ENTRADA	DATOS_POSTULANTE.XLS DATOS_MARICULA.XLS			
FUNCIONALIDADES A PROBAR				
#	FUNCIÓN A PROBAR	ACCIÓN	RESULTADO ESPERADO	RESULTADO OBTENIDO

1	Cantidad de registros de la fuente, contra la cantidad de registros en la FACT_ALUMNOS	Ejecutar proceso ETL_FACT_ALUM NO	La cantidad de registros no duplicados de la tabla FACT_ALUMNO debería ser igual a la cantidad de registros de la fuente.	Se cargaron 10 registros correctamente.
2	Limpieza de duplicados por todos los campos.	Ejecutar proceso ETL_FACT_ALUM NO	Eliminar los registros duplicados por todos campos.	Se eliminaron 5 registros duplicados.
3	Severidad Tipo 2: Número de documento nulo o vacío.	Ejecutar proceso ETL_FACT_ALUM NO	Activar la severidad tipo 2 y marcar con punto, el campo Número de Documento.	Se activó severidad tipo 2. Se marcó con punto el campo número de documento para dos registros.
4	Severidad Tipo 1: Modalidad no identificada.	Ejecutar proceso ETL_FACT_ALUM NO	Activar la severidad Tipo 1 y filtrar registros afectados.	Se activó la severidad Tipo 1 y se filtraron 2 registros con modalidad nula.
5	Limpieza de temporales	Ejecutar proceso ETL_FACT_ALUM NO	Los archivos temporales creados para realizar la carga deben eliminarse al finalizar el proceso.	Los archivos temporales fueron eliminados.

Elaboración: los autores

5.1.1 Ejecución del caso de prueba #1 - ETL_FACT_ALUMNO

5.1.1.1 Funcionalidad #1: Cantidad de Registros

a) Datos de entrada

- **Archivo Datos Postulante.XLS:** Se tiene una base de 10 postulantes para verificar esta funcionalidad.

Tabla LXXXI – Datos del postulante

CODALUMNO	PATERNO	MATERNO	NOMBRES	...	PAÍS	INGRESO
2014102013	CCONISLLA	ARIAS	ANDRES WILLIAMS	...	PERÚ	SI
2014102014	MANDUJANO	REYES	MIJAEL JOSE	...	PERÚ	SI
2014102015	TAMARACCA	TAPIA	AXEL GIANFRANCO	...	PERÚ	SI
2014102016	SUSANIBAR	ROQUE	MADELEYNE	...	PERÚ	SI
2014102017	PRUDENCIO	PARRA	JUNIOR GABRIEL	...	PERÚ	SI
2014102018	BENAVIDES	GARRO	MERY ISABEL	...	PERÚ	SI
2014102019	RODRIGUEZ	PEREZ	GIANFRANCO	...	PERÚ	NO
2014102020	ARANCIBIA	PEREZ	IVAN ISAAC	...	PERÚ	SI
2014102021	RODRIGUEZ	SOLIS	GENARO ALBERTO	...	PERÚ	SI
2014102022	GUTIERREZ	SERNA	EDITH	...	PERÚ	SI

Elaboración: los autores

- **Archivo Datos_Matricula.XLS:** Se tiene una base de 10 ingresantes, entre matriculados y no matriculados para verificar esta funcionalidad.

Tabla LXXXII – Datos de matricula

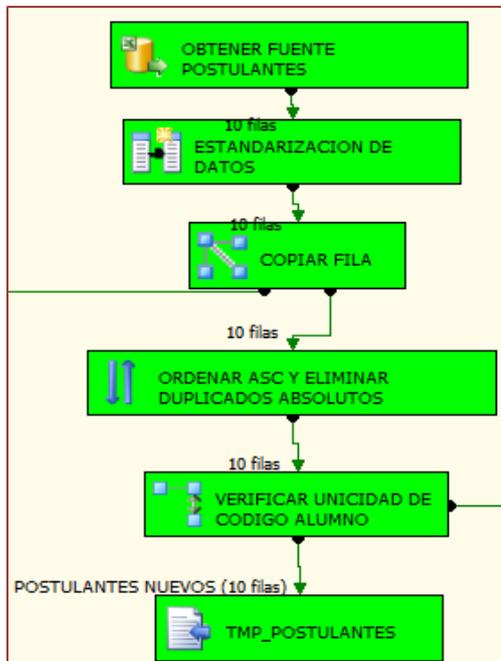
CODALUMNO	PATERNO	MATERNO	NOMBRES	MATRICULA	FECMATRICULA
2014102013	CCONISLLA	ARIAS	ANDRES WILLIAMS	SI	30/01/2014
2014102014	MANDUJANO	REYES	MIJAEL JOSE	SI	30/01/2014
2014102015	TAMARACCA	TAPIA	AXEL GIANFRANCO	SI	30/01/2014
2014102016	SUSANIBAR	ROQUE	MADELEYNE	SI	31/01/2014
2014102017	PRUDENCIO	PARRA	JUNIOR GABRIEL	NO	
2014102018	BENAVIDES	GARRO	MERY ISABEL	SI	31/01/2014
2014102019	RODRIGUEZ	PEREZ	GIANFRANCO	NO	
2014102020	ARANCIBIA	PEREZ	IVAN ISAAC	SI	30/01/2014
2014102021	RODRIGUEZ	SOLIS	GENARO ALBERTO	SI	31/01/2014
2014102022	GUTIERREZ	SERNA	EDITH	SI	30/01/2014

Elaboración: los autores

b) Proceso ETL

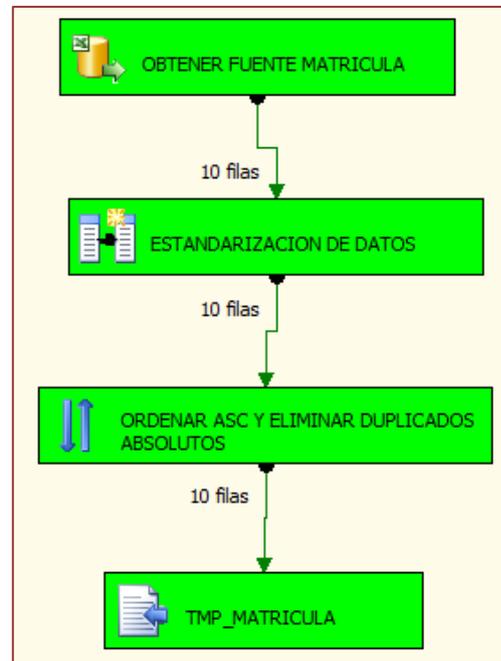
A continuación, los procesos ETL involucrados en el caso de prueba.

**Ilustración LXXXII - Extracción
fuente postulantes**



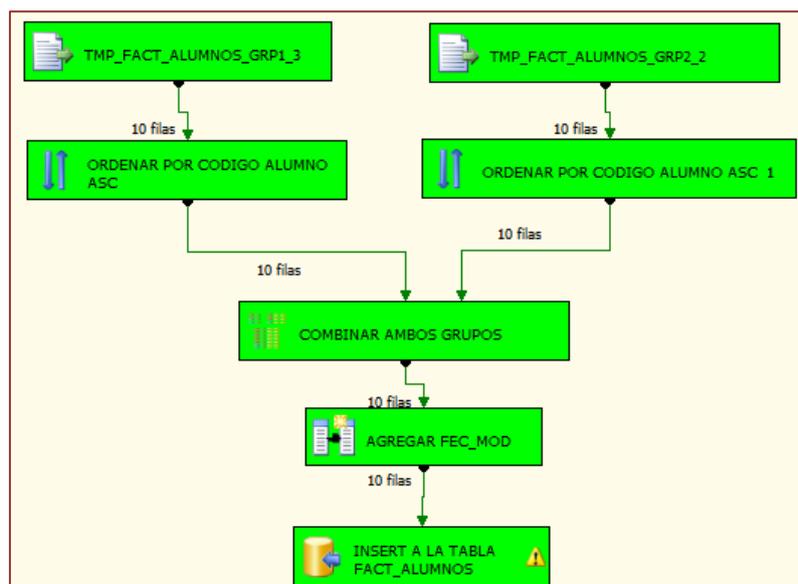
Elaboración: los autores

**Ilustración LXXXIII - Extracción
fuente matrícula**



Elaboración: los autores

Ilustración LXXXIV - Carga de la tabla FACT_ALUMNO

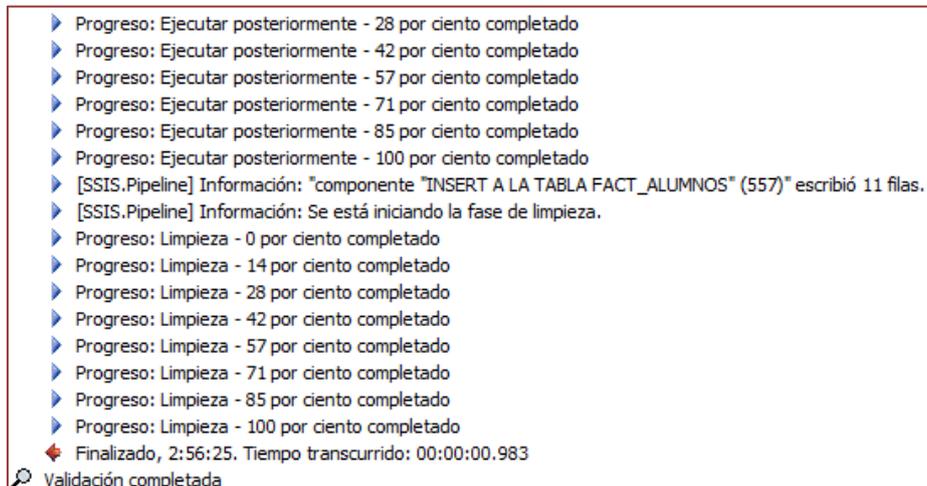


Elaboración: los autores

c) Log de ejecución

Éste es el log de ejecución del flujo de datos CW_FACT_ALUMNOS, donde se aprecia la cantidad de registros cargados a la FACT_ALUMNOS.

Ilustración LXXXV - Log de ejecución de carga



Elaboración: los autores

d) Resultado final

Una vez terminada la carga, podemos observar que la tabla FACT_ALUMNO fue cargada correctamente con 10 registros.

Ilustración LXXXVI - Tabla FACT_ALUMNO

```
select * from dbo.FACT_ALUMNO
```

	ID_ALUMNO	NOMBRE	FEC_NACIMIENTO	EDAD	TELEFONO_1	TELEFONO_2	PUNTAJE	MERITO
1	2014102013	CCONISLLA ARIAS ANDRES WILLIAMS	1997-08-26 00:00:00	16	958043966	934273999	215	386
2	2014102014	MANDUJANO REYES MIJUEL JOSE	1997-09-27 00:00:00	16	919674237	912177651	241	94
3	2014102015	TAMARACCA TAPIA AXEL GIANFRANCO	1997-05-29 00:00:00	17	935590967	975906757	241	92
4	2014102016	SUSANIBAR ROQUE MADELEYNE	1995-10-08 00:00:00	18	980797479	945196912	209	289
5	2014102017	PRUDENCIO PARRA JUNIOR GABRIEL	1997-10-20 00:00:00	16	910550269	948765025	217	350
6	2014102018	BENAVIDES GARRO MERY ISABEL	1997-08-28 00:00:00	16	972777952	914957850	236	168
7	2014102020	ARANCIBIA PEREZ IVAN ISAAC	1997-06-20 00:00:00	16	996567971	956892527	247	54
8	2014102021	RODRIGUEZ SOLIS GENARO ALBERTO	1997-11-14 00:00:00	16	960473843	977179379	189	422
9	2014102022	GUTIERREZ SERNA EDITH	1997-10-19 00:00:00	16	984959397	940659143	185	445
10	2014102023	GUTIERREZ RIOJA AYRTON RAPHAEL	1997-06-17 00:00:00	16	945776804	913302050	238	150

Elaboración: los autores

5.1.1.2 Funcionalidad #2: Limpieza de duplicados absolutos

a) Datos de entrada

- Archivo Datos Postulante.XLS: Se tiene una base de 5 postulantes duplicados para verificar esta funcionalidad.

Tabla LXXXIII – Datos postulantes duplicados

CODALUMNO	PATERNO	MATERNO	NOMBRES	...	PAÍS	INGRESO
2014102013	CCONISLLA	ARIAS	ANDRES WILLIAMS	...	PERÚ	SI
2014102013	CCONISLLA	ARIAS	ANDRES WILLIAMS	...	PERÚ	SI
2014102014	MANDUJANO	REYES	MIJAEL JOSE	...	PERÚ	SI
2014102014	MANDUJANO	REYES	MIJAEL JOSE	...	PERÚ	SI
2014102015	TAMARACCA	TAPIA	AXEL GIANFRANCO	...	PERÚ	SI
2014102015	TAMARACCA	TAPIA	AXEL GIANFRANCO	...	PERÚ	SI
2014102016	SUSANIBAR	ROQUE	MADELEYNE	...	PERÚ	SI
2014102016	SUSANIBAR	ROQUE	MADELEYNE	...	PERÚ	SI
2014102017	PRUDENCIO	PARRA	JUNIOR GABRIEL	...	PERÚ	SI
2014102017	PRUDENCIO	PARRA	JUNIOR GABRIEL	...	PERÚ	SI

Elaboración: los autores

- **Archivo Datos_Matricula.XLS:** Se tiene una base de 5 ingresantes duplicados, entre matriculados y no matriculados para verificar esta funcionalidad.

Tabla LXXXIV – Datos matriculas duplicadas

CODALUMN O	PATERNO	MATERN O	NOMBRES	...	MATRICUL A	FECMATR ICULA
2014102013	CCONISLLA	ARIAS	ANDRES WILLIAMS	...	SI	30/01/2014
2014102013	CCONISLLA	ARIAS	ANDRES WILLIAMS	...	SI	30/01/2014
2014102014	MANDUJAN O	REYES	MIJAEL JOSE	...	SI	30/01/2014
2014102014	MANDUJAN O	REYES	MIJAEL JOSE	...	SI	30/01/2014
2014102015	TAMARACC A	TAPIA	AXEL GIANFRANCO	...	SI	30/01/2014
2014102015	TAMARACC A	TAPIA	AXEL GIANFRANCO	...	SI	30/01/2014
2014102016	SUSANIBA R	ROQUE	MADELEYNE	...	SI	31/01/2014
2014102016	SUSANIBA R	ROQUE	MADELEYNE	...	SI	31/01/2014
2014102017	PRUDENCI O	PARRA	JUNIOR GABRIEL	...	NO	
2014102017	PRUDENCI O	PARRA	JUNIOR GABRIEL	...	NO	

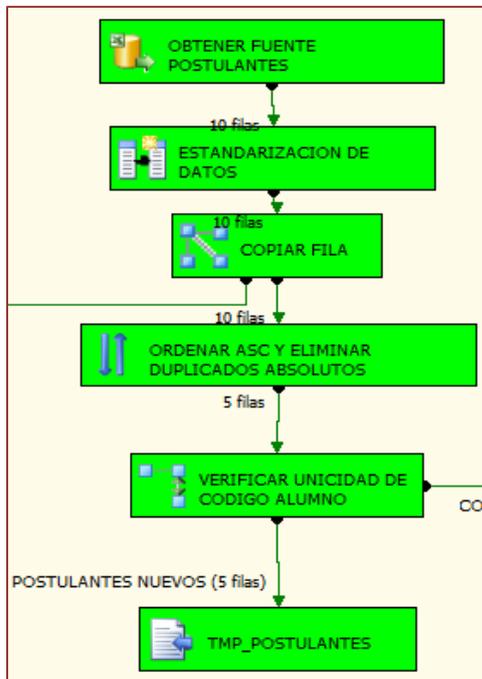
Elaboración: los autores

b) Proceso ETL

A continuación, los procesos ETL involucrados en el caso de prueba.

Ilustración LXXXVII - Extracción fuente

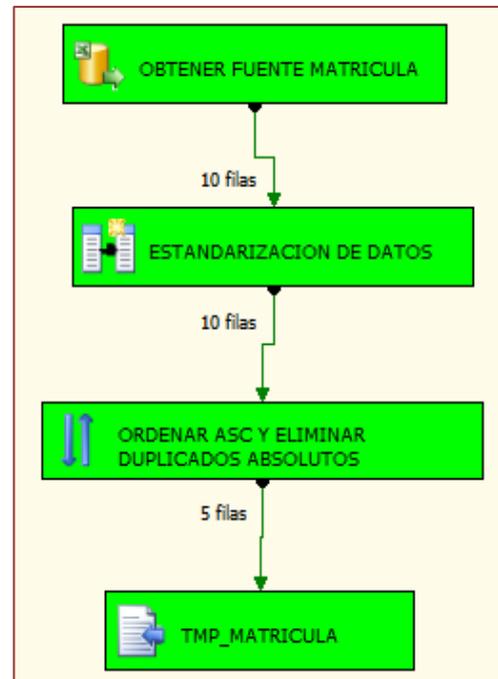
postulantes



Elaboración: los autores

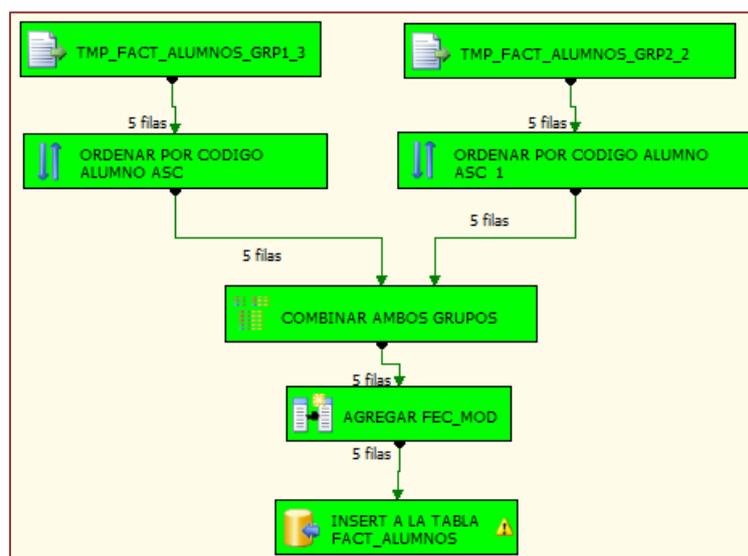
Ilustración LXXXVIII - Extracción fuente

matrícula



Elaboración: los autores

Ilustración LXXXIX - Carga de la tabla FACT_ALUMNO



Elaboración: los autores

c) Log de ejecución

Se presenta el log de ejecución del Flujo de datos EX_SRC_POSTULANTES, donde se aprecia la cantidad de registros cargados al archivo temporal TMP_POSTULANTE.

Ilustración XC - Log EX_SRC_POSTULANTES

```
▶ Progreso: Ejecutar posteriormente - 61 por ciento completado
▶ Progreso: Ejecutar posteriormente - 69 por ciento completado
▶ Progreso: Ejecutar posteriormente - 76 por ciento completado
▶ Progreso: Ejecutar posteriormente - 84 por ciento completado
▶ Progreso: Ejecutar posteriormente - 92 por ciento completado
▶ [TMP_SEV_DOC [1426]] Información: Finalizó el procesamiento del archivo "C:\Users\Diego\Desktop\TESIS 2014\ELT\TEMP\TMP_SEV_DOC.TXT".
▶ Progreso: Ejecutar posteriormente - 100 por ciento completado
▶ [SSIS.Pipeline] Información: "componente "TMP_POSTULANTES" (517)" escribió 5 filas.
```

Elaboración: los autores

Se presenta el log de ejecución del Flujo de datos EX_SRC_MATRICULA, donde se aprecia la cantidad de registros cargados al archivo temporal TMP_MATRICULA.

Ilustración XCI - Log EX_SRC_Matrícula

```
▶ Progreso: Ejecutar posteriormente - 0 por ciento completado
▶ Progreso: Ejecutar posteriormente - 25 por ciento completado
▶ Progreso: Ejecutar posteriormente - 50 por ciento completado
▶ Progreso: Ejecutar posteriormente - 75 por ciento completado
▶ [TMP_MATRICULA [517]] Información: Finalizó el procesamiento del archivo "C:\Users\Diego\Desktop\TESIS 2014\ELT\TEMP\MATRICULA.TXT".
▶ Progreso: Ejecutar posteriormente - 100 por ciento completado
▶ [SSIS.Pipeline] Información: "componente "TMP_MATRICULA" (517)" escribió 5 filas.
```

Elaboración: los autores

Se presenta el log de ejecución del Flujo de datos de CW_FACT_ALUMNO, donde se aprecia la cantidad de registros cargados al a la tabla FACT_ALUMNO.

Ilustración XCII - Log FACT_ALUMNO

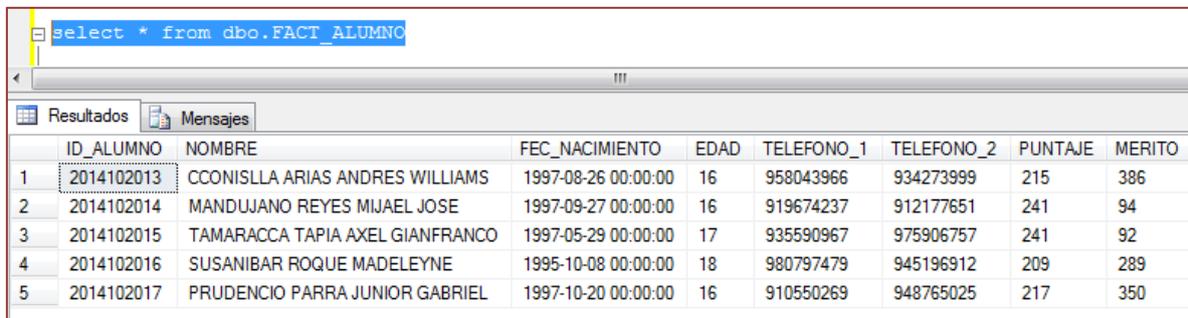
```
▶ Progreso: Ejecutar posteriormente - 28 por ciento completado
▶ Progreso: Ejecutar posteriormente - 42 por ciento completado
▶ Progreso: Ejecutar posteriormente - 57 por ciento completado
▶ Progreso: Ejecutar posteriormente - 71 por ciento completado
▶ Progreso: Ejecutar posteriormente - 85 por ciento completado
▶ Progreso: Ejecutar posteriormente - 100 por ciento completado
▶ [SSIS.Pipeline] Información: "componente "INSERT A LA TABLA FACT_ALUMNOS" (557)" escribió 5 filas.
```

Elaboración: los autores

d) Resultado final

Una vez terminada la carga, podemos observar que la tabla FACT_ALUMNO fue cargada correctamente con 5 registros.

Ilustración XCIII - Tabla FACT_ALUMNO



The screenshot shows a SQL query window with the following query: `select * from dbo.FACT_ALUMNO`. Below the query, there are tabs for 'Resultados' and 'Mensajes'. The 'Resultados' tab is active, displaying a table with 5 rows and 10 columns: ID_ALUMNO, NOMBRE, FEC_NACIMIENTO, EDAD, TELEFONO_1, TELEFONO_2, PUNTAJE, and MERITO. The data is as follows:

	ID_ALUMNO	NOMBRE	FEC_NACIMIENTO	EDAD	TELEFONO_1	TELEFONO_2	PUNTAJE	MERITO
1	2014102013	CCONISLLA ARIAS ANDRES WILLIAMS	1997-08-26 00:00:00	16	958043966	934273999	215	386
2	2014102014	MANDUJANO REYES MIJAEEL JOSE	1997-09-27 00:00:00	16	919674237	912177651	241	94
3	2014102015	TAMARACCA TAPIA AXEL GIANFRANCO	1997-05-29 00:00:00	17	935590967	975906757	241	92
4	2014102016	SUSANIBAR ROQUE MADELEYNE	1995-10-08 00:00:00	18	980797479	945196912	209	289
5	2014102017	PRUDENCIO PARRA JUNIOR GABRIEL	1997-10-20 00:00:00	16	910550269	948765025	217	350

Elaboración: los autores

5.1.1.3 Funcionalidad #3: Severidad tipo 2 - Completitud

a) Datos de entrada

- **Archivo Datos Postulante.XLS:** Se tiene una base de 10 postulantes para verificar esta funcionalidad. Para efecto de la prueba se colocaron dos documentos en nulo.

Tabla LXXXV – Datos nulos de postulantes

CODALUMNO	PATERNO	MATERNO	NOMBRES	...	NUMERODOC	INGRESO
2014102143	MOSQUEIRA	TAPIA	HANS	...	77477622	SI
2014102144	VELASQUEZ	PEREZ	SILVANA MARIA	...	75362249	SI
2014102145	FERNANDEZ	BURGA	BLANCA MARISOL	...	77440108	NO
2014102146	SALAMANCA	ARIAS	VICTOR ALONSO	...	NULL	SI
2014102147	HALANOCCA	MORON	MIGUEL HERBERT	...	77269064	SI
2014102148	SEMINARIO	RAMOS	ITALO MARTIN	...	77245655	SI
2014102149	MONTALBAN	CALLE	HILVER JODAN	...	77414759	SI
2014102150	CARRETERO	SILVA	CLAUDIA XIMENA	...	77102876	NO
2014102151	CARRETERO	SILVA	SILVANA PAOLA	...	NULL	NO
2014102152	SANTILLAN	ROSSI	KIMBERLY ANAYKA GENOVEVA	...	78317887	NO

Elaboración: los autores

- **Archivo Datos_Matricula.XLS:** Se tiene una base de 6 ingresantes, entre matriculados y no matriculados para verificar esta funcionalidad.

Tabla LXXXVI – Datos de ingresantes

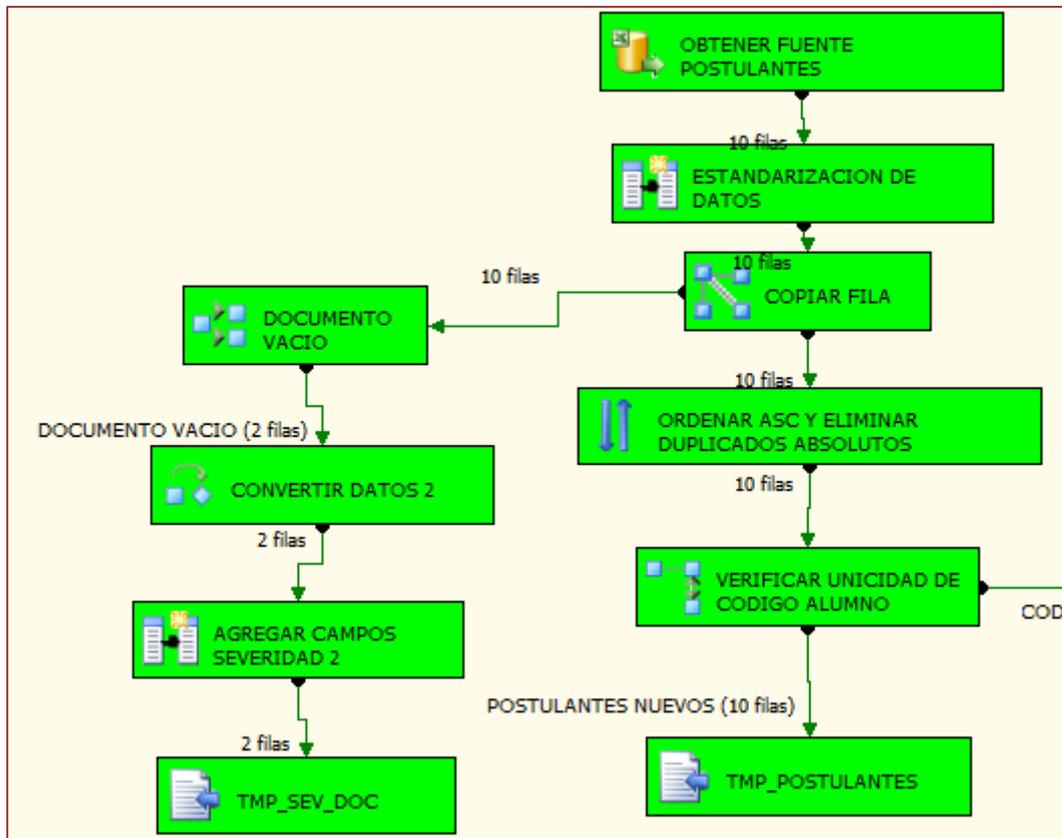
CODALUMNO	PATERNO	MATERNO	NOMBRES	...	MATRICULA	FECMATRICULA
2014102143	MOSQUEIRA	TAPIA	HANS	...	SI	30/01/2014
2014102144	VELASQUEZ	PEREZ	SILVANA MARIA	...	SI	30/01/2014
2014102146	SALAMANCA	ARIAS	VICTOR ALONSO	...	SI	31/01/2014
2014102147	HALANOCCA	MORON	MIGUEL HERBERT	...	SI	30/01/2014
2014102148	SEMINARIO	RAMOS	ITALO MARTIN	...	SI	31/01/2014
2014102149	MONTALBAN	CALLE	HILVER JODAN	...	SI	30/01/2014

Elaboración: los autores

b) Proceso ETL

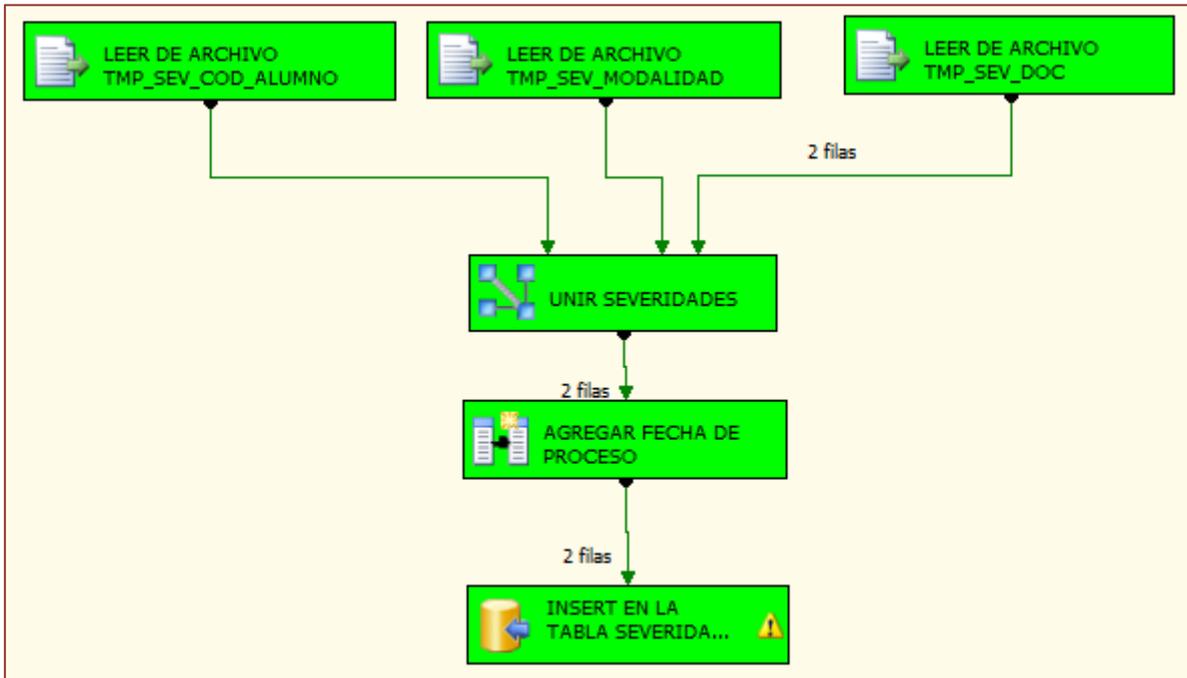
A continuación, los procesos ETL involucrados en el caso de prueba.

Ilustración XCIV - Extracción fuente postulantes



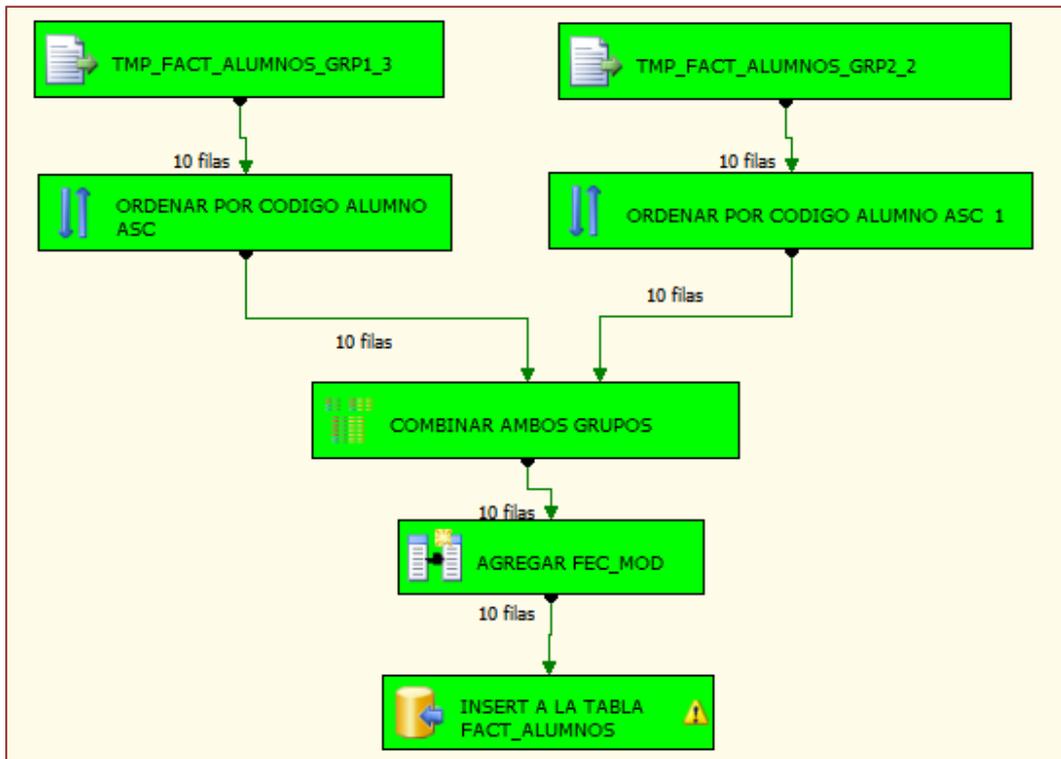
Elaboración: los autores

Ilustración XCV - Carga de severidades



Elaboración: los autores

Ilustración XCVI - Carga tabla FACT_ALUMNO



Elaboración: los autores

c) Logs de ejecución

Se presenta el log de ejecución del Flujo de datos EX_SRC_POSTULANTES, donde se aprecia la cantidad de registros cargados al archivo temporal TMP_POSTULANTE y TMP_SEV_DOC.

Ilustración XCVII - Log de EX_SRC_POSTULANTES

```
▶ Progreso: Ejecutar posteriormente - 61 por ciento completado
▶ Progreso: Ejecutar posteriormente - 69 por ciento completado
▶ Progreso: Ejecutar posteriormente - 76 por ciento completado
▶ Progreso: Ejecutar posteriormente - 84 por ciento completado
▶ Progreso: Ejecutar posteriormente - 92 por ciento completado
▶ [TMP_SEV_DOC [1426]] Información: Finalizó el procesamiento del archivo "C:\Users\Diego\Desktop\TESIS 2014\ELT\TEMP\TMP_SEV_DOC.TXT".
▶ Progreso: Ejecutar posteriormente - 100 por ciento completado
▶ [SSIS.Pipeline] Información: "componente "TMP_POSTULANTES" (517)" escribió 10 filas.
▶ [SSIS.Pipeline] Información: "componente "TMP_SEV_COD_ALUMN" (1088)" escribió 0 filas.
▶ [SSIS.Pipeline] Información: "componente "TMP_SEV_DOC" (1426)" escribió 2 filas.
```

Elaboración: los autores

Se presenta el log de ejecución del Flujo de datos CW_SEVERIDADES, donde se aprecia la cantidad de registros cargados a la tabla SEVERIDADES.

Ilustración XCVIII - Log de CW_SEVERIDADES

```
▶ Progreso: Ejecutar posteriormente - 0 por ciento completado
▶ Progreso: Ejecutar posteriormente - 16 por ciento completado
▶ Progreso: Ejecutar posteriormente - 33 por ciento completado
▶ Progreso: Ejecutar posteriormente - 50 por ciento completado
▶ [LEER DE ARCHIVO TMP_SEV_MODALIDAD [161]] Información: Finalizó el procesamiento del archivo "C:\Users\Diego\Desktop\TESIS 2014\ELT\TEMP\TMP_SEV_MODALIDAD.TXT".
▶ Progreso: Ejecutar posteriormente - 66 por ciento completado
▶ [LEER DE ARCHIVO TMP_SEV_DOC [194]] Información: Finalizó el procesamiento del archivo "C:\Users\Diego\Desktop\TESIS 2014\ELT\TEMP\TMP_SEV_DOC.TXT".
▶ Progreso: Ejecutar posteriormente - 83 por ciento completado
▶ [LEER DE ARCHIVO TMP_SEV_COD_ALUMNO [235]] Información: Finalizó el procesamiento del archivo "C:\Users\Diego\Desktop\TESIS 2014\ELT\TEMP\TMP_SEV_COD_ALUMNO.TXT".
▶ Progreso: Ejecutar posteriormente - 100 por ciento completado
▶ [SSIS.Pipeline] Información: "componente "INSERT EN LA TABLA SEVERIDADES" (60)" escribió 2 filas.
```

Elaboración: los autores

Se presenta el log de ejecución del Flujo de datos de CW_FACT_ALUMNO, donde se aprecia la cantidad de registros cargados a la tabla FACT_ALUMNO.

Ilustración XCIX - Log de carga FACT_ALUMNO

```
▶ Progreso: Ejecutar posteriormente - 28 por ciento completado
▶ Progreso: Ejecutar posteriormente - 42 por ciento completado
▶ Progreso: Ejecutar posteriormente - 57 por ciento completado
▶ Progreso: Ejecutar posteriormente - 71 por ciento completado
▶ Progreso: Ejecutar posteriormente - 85 por ciento completado
▶ Progreso: Ejecutar posteriormente - 100 por ciento completado
▶ [SSIS.Pipeline] Información: "componente "INSERT A LA TABLA FACT_ALUMNOS" (557)" escribió 10 filas.
```

Elaboración: los autores

d) Resultado final

Se puede observar que se registraron dos severidades tipo 2 en la tabla final de severidades.

Ilustración C - Tabla severidades

```
select * from SEVERIDADES
```

	ID_SEVERIDAD	FUENTE	DESC_SEVERIDAD	FEC_MOD	CONTENIDO_CAMPOS	ID_TIPO_SEVERIDAD
1	3236	DATOS_POSTULANTE.TXT	DOCUMENTO VACIO	2014-06-13 04:02:00	2014102146 VICTOR ALONSO...	2
2	3237	DATOS_POSTULANTE.TXT	DOCUMENTO VACIO	2014-06-13 04:02:00	2014102151 SILVANA PAOLA ...	2

Elaboración: los autores

Además, para los dos alumnos cuyos documentos estaban vacíos, se les colocó un punto.

Ilustración CI - Tabla FACT_ALUMNO

```
select ID_ALUMNO, NOMBRE, TIPO_DOCUMENTO, NUM_DOCUMENTO from dbo.FACT_ALUMNO
```

	ID_ALUMNO	NOMBRE	TIPO_DOCUMENTO	NUM_DOCUMENTO
1	2014102143	MOSQUEIRA TAPIA HANS	DNI	77477622
2	2014102144	VELASQUEZ PEREZ SILVANA MARIA	DNI	75362249
3	2014102145	FERNANDEZ BURGA BLANCA MARISOL	DNI	77440108
4	2014102146	SALAMANCA ARIAS VICTOR ALONSO	DNI	.
5	2014102147	HALANOCCA MORON MIGUEL HERBERT	DNI	77269064
6	2014102148	SEMINARIO RAMOS ITALO MARTIN	DNI	77245655
7	2014102149	MONTALBAN CALLE HILVER JODAN	DNI	77414759
8	2014102150	CARRETERO SILVA CLAUDIA XIMENA	DNI	77102876
9	2014102151	CARRETERO SILVA SILVANA PAOLA	DNI	.
10	2014102152	SANTILLAN ROSSI KIMBERLY ANAYKA GENOVEVA	DNI	78317887

Elaboración: los autores

5.1.1.4 Funcionalidad #4: Severidad tipo 1 – Identidad referencial

a) Datos de entrada

- **Archivo Datos Postulante.XLS:** Se tiene una base de 10 postulantes para verificar esta funcionalidad. Para efecto de la prueba se colocaron dos modalidades en nulo.

Tabla LXXXVII – Datos de postulantes

CODALUMNO	PATERO	MATERO	NOMBRES	...	MODALIDAD
2014102143	MOSQUEIRA	TAPIA	HANS	...	CONCURSO ORDINARIO DE ADMISIÓN
2014102144	VELASQUEZ	PEREZ	SILVANA MARIA	...	NULL
2014102145	FERNANDEZ	BURGA	BLANCA MARISOL	...	CONCURSO ORDINARIO DE ADMISIÓN
2014102146	SALAMANCA	ARIAS	VICTOR ALONSO	...	EXONERADO POR PRIMERA ALTERNATIVA
2014102147	HALANOCCA	MORON	MIGUEL HERBERT	...	CONCURSO ORDINARIO DE ADMISIÓN
2014102148	SEMINARIO	RAMOS	ITALO MARTIN	...	EXONERADO POR PRIMERA ALTERNATIVA
2014102149	MONTALBAN	CALLE	HILVER JODAN	...	NULL
2014102150	CARRETERO	SILVA	CLAUDIA XIMENA	...	EXONERADO POR PRIMERA ALTERNATIVA
2014102151	CARRETERO	SILVA	SILVANA PAOLA	...	EXONERADO POR PRIMERA ALTERNATIVA
2014102152	SANTILLAN	ROSSI	KIMBERLY ANAYKA GENOVEVA	...	CONCURSO ORDINARIO DE ADMISIÓN

Elaboración: los autores

- **Archivo Datos_Matricula.XLS:** Se tiene una base de 6 ingresantes, entre matriculados y no matriculados para verificar esta funcionalidad.

Tabla LXXXVIII – Datos de ingresantes

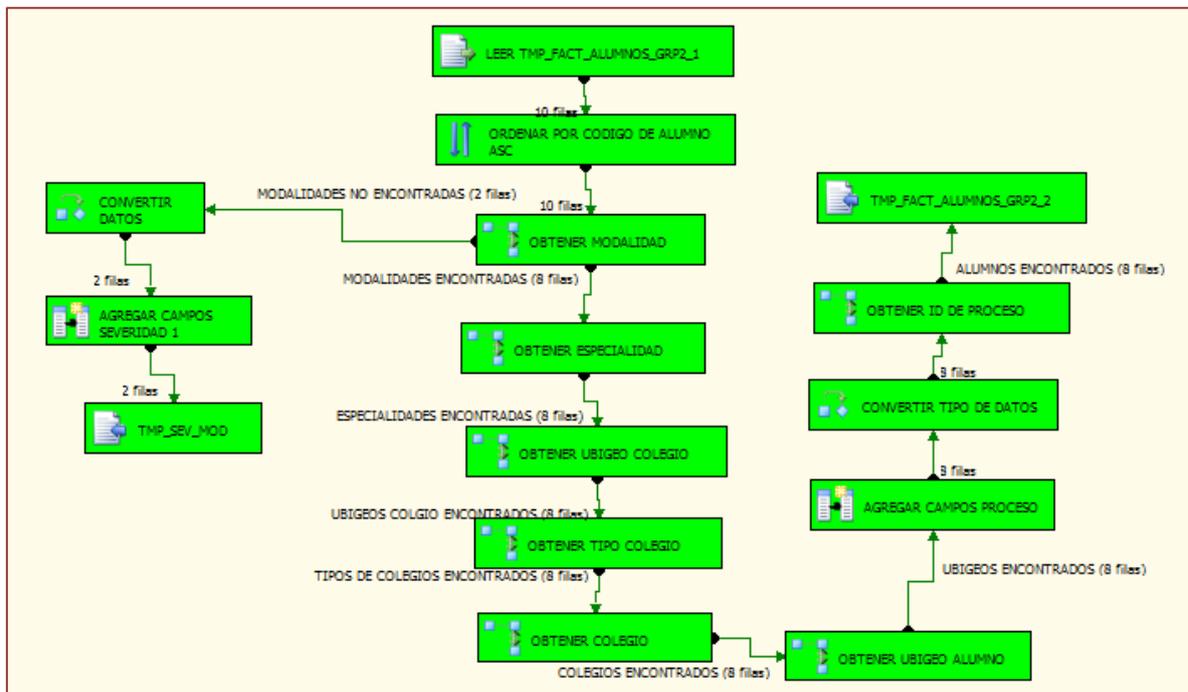
CODALUMNO	PATERO	MATERNO	NOMBRES	MATRICULA	FECMATRICULA
2014102143	MOSQUEIRA	TAPIA	HANS	SI	30/01/2014
2014102144	VELASQUEZ	PEREZ	SILVANA MARIA	SI	30/01/2014
2014102146	SALAMANCA	ARIAS	VICTOR ALONSO	SI	31/01/2014
2014102147	HALANOCCA	MORON	MIGUEL HERBERT	SI	30/01/2014
2014102148	SEMINARIO	RAMOS	ITALO MARTIN	SI	31/01/2014
2014102149	MONTALBAN	CALLE	HILVER JODAN	SI	30/01/2014

Elaboración: los autores

b) Proceso ETL

A continuación, los procesos ETL involucrados en el caso de prueba.

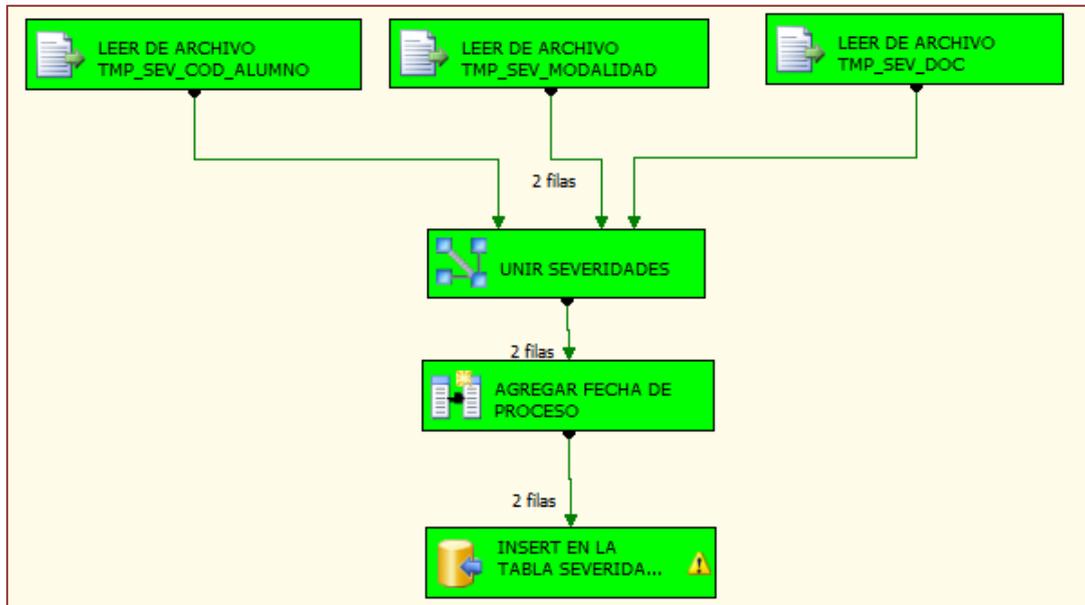
Ilustración CII – RN_FACT_ALUMNO_GRP2_2



Elaboración: los autores

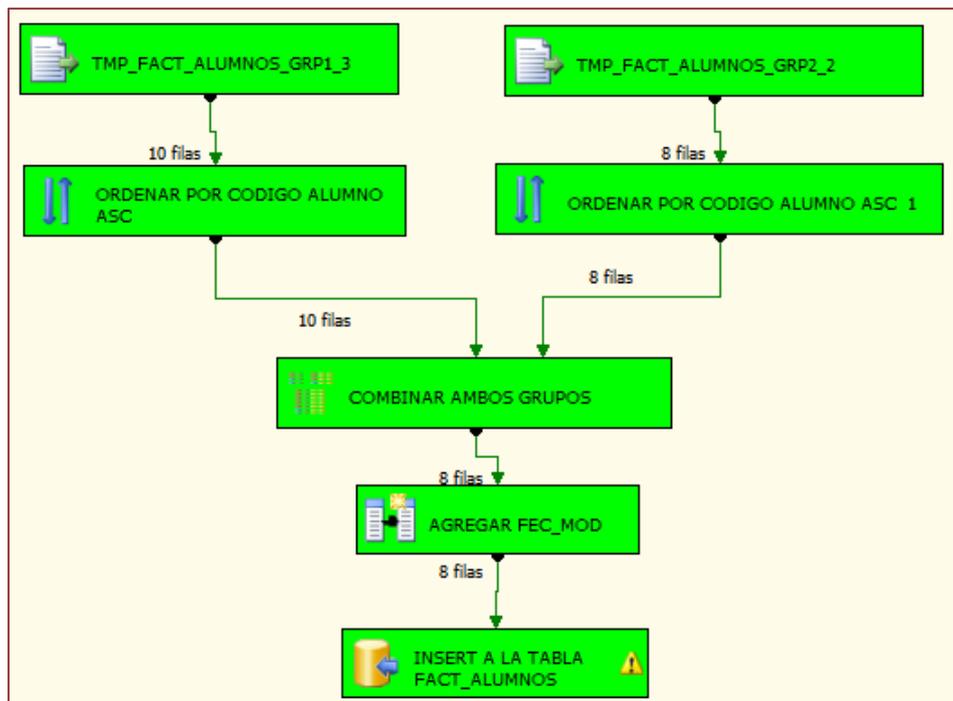
Al lado izquierdo se observa el flujo cuando no se encuentra una modalidad específica.

Ilustración CIII - Carga de severidades



Elaboración: los autores

Ilustración CIV - Carga tabla FACT_ALUMNO



Elaboración: los autores

c) Logs de ejecución

Se presenta el log de ejecución del Flujo de datos RN_FACT_ALUMNO_GRP2_2, donde se aprecia la cantidad de registros cargados al archivo temporal TMP_POSTULANTE y TMP_SEV_MOD.

Ilustración CV - Log de RN_FACT_ALUMNO_GRP2_2

```
▶ [TMP_FACT_ALUMNOS_GRP2_2 [601]] Información: Finalizó el procesamiento del archivo "C:\Users\Diego\Desktop\TESIS 2014\ELT\TEMP\TMP_FACT_ALUMNOS_GRP2_2.TXT".
▶ Progreso: Ejecutar posteriormente - 80 por ciento completado
▶ Progreso: Ejecutar posteriormente - 86 por ciento completado
▶ [TMP_SEV_MOD [791]] Información: Finalizó el procesamiento del archivo "C:\Users\Diego\Desktop\TESIS 2014\ELT\TEMP\TMP_SEV_MODALIDAD.TXT".
▶ Progreso: Ejecutar posteriormente - 93 por ciento completado
▶ Progreso: Ejecutar posteriormente - 100 por ciento completado
▶ [SSIS.Pipeline] Información: "componente "TMP_FACT_ALUMNOS_GRP2_2" (601)" escribió 8 filas.
```

Elaboración: los autores

Se presenta el log de ejecución del Flujo de datos CW_SEVERIDADES, donde se aprecia la cantidad de registros cargados a la tabla SEVERIDADES.

Ilustración CVI - Log de CW_SEVERIDADES

```
▶ [LEER DE ARCHIVO TMP_SEV_MODALIDAD [161]] Información: Finalizó el procesamiento del archivo "C:\Users\Diego\Desktop\TESIS 2014\ELT\TEMP\TMP_SEV_MODALIDAD.TXT".
▶ Progreso: Ejecutar posteriormente - 66 por ciento completado
▶ [LEER DE ARCHIVO TMP_SEV_DOC [194]] Información: Finalizó el procesamiento del archivo "C:\Users\Diego\Desktop\TESIS 2014\ELT\TEMP\TMP_SEV_DOC.TXT".
▶ Progreso: Ejecutar posteriormente - 83 por ciento completado
▶ [LEER DE ARCHIVO TMP_SEV_COD_ALUMNO [235]] Información: Finalizó el procesamiento del archivo "C:\Users\Diego\Desktop\TESIS 2014\ELT\TEMP\TMP_SEV_COD_ALUMNO.TXT".
▶ Progreso: Ejecutar posteriormente - 100 por ciento completado
▶ [SSIS.Pipeline] Información: "componente "INSERT EN LA TABLA SEVERIDADES" (60)" escribió 2 filas.
```

Elaboración: los autores

Se presenta el log de ejecución del Flujo de datos de CW_FACT_ALUMNO, donde se aprecia la cantidad de registros cargados a la tabla FACT_ALUMNO.

Ilustración CVII - Log de carga FACT_ALUMNO

```
▶ [TMP_FACT_ALUMNOS_GRP1_3 [1]] Información: Finalizó el procesamiento del archivo "C:\Users\Diego\Desktop\TESIS 2014\ELT\TEMP\TMP_FACT_ALUMNOS_GRP1_3.TXT".
▶ Progreso: Ejecutar posteriormente - 14 por ciento completado
▶ [TMP_FACT_ALUMNOS_GRP2_2 [9]] Información: Finalizó el procesamiento del archivo "C:\Users\Diego\Desktop\TESIS 2014\ELT\TEMP\TMP_FACT_ALUMNOS_GRP2_2.TXT".
▶ Progreso: Ejecutar posteriormente - 28 por ciento completado
▶ Progreso: Ejecutar posteriormente - 42 por ciento completado
▶ Progreso: Ejecutar posteriormente - 57 por ciento completado
▶ Progreso: Ejecutar posteriormente - 71 por ciento completado
▶ Progreso: Ejecutar posteriormente - 85 por ciento completado
▶ Progreso: Ejecutar posteriormente - 100 por ciento completado
▶ [SSIS.Pipeline] Información: "componente "INSERT A LA TABLA FACT_ALUMNOS" (557)" escribió 8 filas.
```

Elaboración: los autores

d) Resultado final

Se puede observar que se registraron dos severidades tipo 1 en la tabla final de severidades.

Ilustración CVIII - Tabla severidades

```
select * from SEVERIDADES
```

ID_SEVERIDAD	FUENTE	DESC_SEVERIDAD	FEC_MOD	CONTENIDO_CAMPOS	ID_TIPO_SEVERIDAD
1	DATOS_POSTULANTE.TXT	MODALIDAD NO EXISTE	2014-06-13 05:04:00	2014102144 .	1
2	DATOS_POSTULANTE.TXT	MODALIDAD NO EXISTE	2014-06-13 05:04:00	2014102149 .	1

Elaboración: los autores

Además, los dos alumnos cuyas sus modalidades estaban vacías, no se cargaron a la tabla FACT_ALUMNO.

Ilustración CIX - Tabla FACT_ALUMNO

```
select ALU.ID_ALUMNO, ALU.NOMBRE, ALU.ID_MODALIDAD, MOD.MODALIDAD  
FROM dbo.FACT_ALUMNO ALU INNER JOIN DIM_MODALIDAD MOD  
ON ALU.ID_MODALIDAD = MOD.ID_MODALIDAD
```

ID_ALUMNO	NOMBRE	ID_MODALIDAD	MODALIDAD
2014102146	SALAMANCA ARIAS VICTOR ALONSO	7	EXONERADO POR PRIMERA ALTERNATIVA
2014102148	SEMINARIO RAMOS ITALO MARTIN	7	EXONERADO POR PRIMERA ALTERNATIVA
2014102150	CARRETERO SILVA CLAUDIA XIMENA	7	EXONERADO POR PRIMERA ALTERNATIVA
2014102151	CARRETERO SILVA SILVANA PAOLA	7	EXONERADO POR PRIMERA ALTERNATIVA
2014102143	MOSQUEIRA TAPIA HANS	9	CONCURSO ORDINARIO DE ADMISIÓN
2014102145	FERNANDEZ BURGA BLANCA MARISOL	9	CONCURSO ORDINARIO DE ADMISIÓN
2014102147	HALANOCCA MORON MIGUEL HERBERT	9	CONCURSO ORDINARIO DE ADMISIÓN
2014102152	SANTILLAN ROSSI KIMBERLY ANAYKA GENOVEVA	9	CONCURSO ORDINARIO DE ADMISIÓN

Elaboración: los autores

5.1.1.5 Funcionalidad #5: Limpieza de temporales

a) Datos de entrada

No son necesarios datos de entrada. No aplica.

b) Proceso ETL

A continuación, el proceso ETL involucrados en el caso de prueba.

Ilustración CX - Limpieza archivos temporales

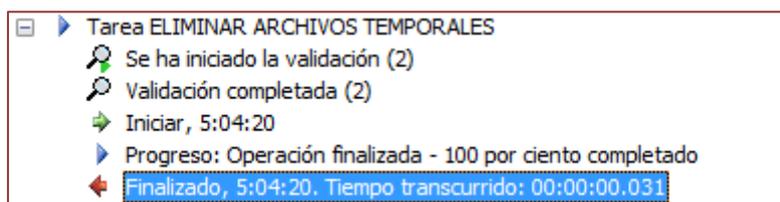


Elaboración: los autores

c) Logs de ejecución

Se presenta el log de ejecución del proceso de eliminación de archivos temporales.

Ilustración CXI - Eliminación archivos temporales

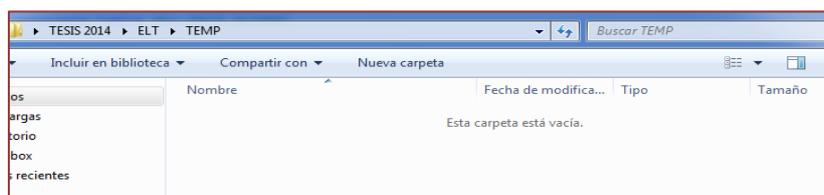


Elaboración: los autores

a) Resultado final

Como se puede observar, los archivos fueron eliminados correctamente.

Ilustración CXII - Carpeta de archivos temporales



Elaboración: los autores

5.2 Caso de prueba #2 - Consistencia de datos

Tabla LXXXIX - Caso prueba #2

CASO DE PRUEBA #2 – Consistencia de datos				
OBJETO A ANALIZAR		Servicios OLAP		
OBJETIVO		Verificar que el cubo Postulantes muestre la misma información de la tabla DIM_CARRERA con DIM_FACULTAD		
PRE-REQUISITOS		<ul style="list-style-type: none"> • Disponibilidad de la base de datos BD_DATAMART_FIA • Disponibilidad del cubo Postulantes 		
DATOS DE ENTRADA		Realizar la misma consulta		
FUNCIONALIDADES A PROBAR				
#	FUNCIÓN A PROBAR	ACCIÓN	RESULTADO ESPERADO	RESULTADO OBTENIDO
1	Valores en la BD_DATAMART_FIA frente a los valores del cubo	Realizar la consulta en la base de datos del Datamart y en el Cubo	El cubo de Postulantes muestre la misma cantidad de Carreras con su respectiva Facultad que DIM_CARRERA con DIM_FACULTAD	Muestra la misma información

Elaboración: los autores

5.2.1 Funcionalidad #1: Consistencia de datos

a) Consulta a realizar

Mostrar las carreras con sus respectivas facultades

b) Datos de entrada

```
Select C.CARRERA , F.FACULTAD from dbo.DIM_CARRERA C inner join  
dbo.DIM_FACULTAD F on C.ID_FACULTAD = F.ID_FACULTAD where  
C.ID_FACULTAD = 5
```

c) Resultado del datamart

Ilustración CXIII - Resultado datamart consulta



	CARRERA	FACULTAD
1	ARQUITECTURA	INGENIERÍA Y ARQUITECTURA
2	CIENCIAS AERONÁUTICAS	INGENIERÍA Y ARQUITECTURA
3	INGENIERÍA CIVIL	INGENIERÍA Y ARQUITECTURA
4	INGENIERÍA DE COMPUTACIÓN Y SISTEMAS	INGENIERÍA Y ARQUITECTURA
5	INGENIERÍA ELECTRÓNICA	INGENIERÍA Y ARQUITECTURA
6	INGENIERÍA EN INDUSTRIAS ALIMENTARIAS	INGENIERÍA Y ARQUITECTURA
7	INGENIERÍA INDUSTRIAL	INGENIERÍA Y ARQUITECTURA

Elaboración: los autores

d) Resultado del cubo

Ilustración CXIV - Resultado del cubo

Facultad	Carrera	Numero Postulantes
INGENIERÍA Y ARQUITECTURA	ARQUITECTURA	181
	CIENCIAS AERONÁUTICAS	159
	INGENIERÍA CIVIL	192
	INGENIERÍA DE COMPUTACIÓN Y SISTEMAS	226
	INGENIERÍA ELECTRÓNICA	151
	INGENIERÍA EN INDUSTRIAS ALIMENTARIAS	177
	INGENIERÍA INDUSTRIAL	238
	Total	1.324
Total general	1.324	

Elaboración: los autores

5.3 Caso de prueba #3 – Consistencia de Datos_Reportes

Tabla XC - Caso prueba #3

CASO DE PRUEBA #3 – Consistencia de Datos_Reportes				
OBJETO A ANALIZAR		Reportes Excel		
OBJETIVO		Verificar que el Reporte Excel muestre la misma información que el cubo de información		
PRE-REQUISITOS		<ul style="list-style-type: none"> • Disponibilidad de la base de datos BD_DATAMART_FIA • Disponibilidad del cubo Postulantes • Disponibilidad de Reporte diseñado en Excel 		
DATOS DE ENTRADA		Realizar la misma consulta		
FUNCIONALIDADES A PROBAR				
#	FUNCIÓN A PROBAR	ACCIÓN	RESULTADO ESPERADO	RESULTADO OBTENIDO
1	Valores en el reporte Excel frente a los valores del cubo	Realizar la consulta en el Reporte Excel y en el Cubo	El reporte Excel muestre los mismos datos que la consulta en el cubo	Muestra la misma información

Elaboración: los autores

5.3.1 Funcionalidad #1: Consistencia de Datos_Reportes

a) Consulta a realizar

Mostrar cantidad de alumnos según estado de su matricula

b) Resultado del datamart

Ilustración CXV – Resultado cubo

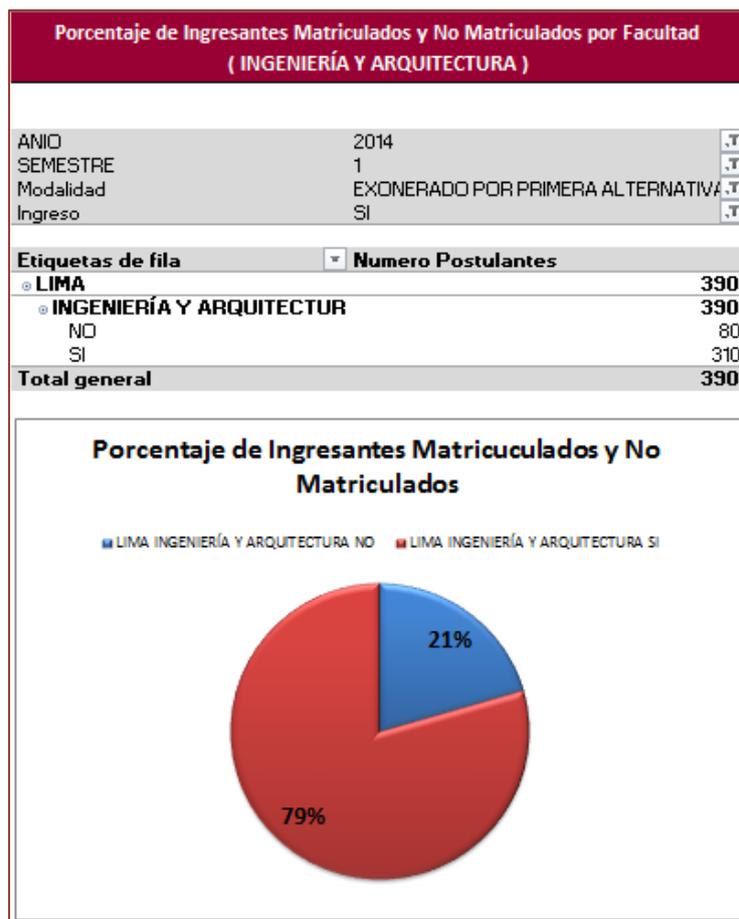
Dimensión	Jerarquía	Operador	Expresión de filtro
Proceso de Admisión	ANIO	Igual	{ 2014 }
Proceso de Admisión	SEMESTRE	Igual	{ 1 }
Modalidad de Ingreso	Modalidad	Igual	{ EXONERADO POR PRIMERA ALTERNATIVA }
Ingreso	Ingreso	Igual	{ SI }

Coloque campos de filtro aquí			
Sede	Facultad		
LIMA	INGENIERÍA Y ARQUITECTURA	Total	Total general
Matricula	Numero Postulantes	Numero Postulantes	Numero Postulantes
NO	80	80	80
SI	310	310	310
Total general	390	390	390

Elaboración: los autores

c) Resultado del cubo

Ilustración CXVI – Resultado Excel



Elaboración: los autores

5.4 Caso de prueba #4 – Indicador No_Matriculados

Tabla XCI - Caso prueba #4

CASO DE PRUEBA #4 – Indicador No_Matriculados				
OBJETO A ANALIZAR		Indicador Excel		
OBJETIVO		Verificar que el Indicador que se encuentra dentro del Reporte Excel muestre la misma información que el cubo de información		
PRE-REQUISITOS		<ul style="list-style-type: none"> • Disponibilidad de la base de datos BD_DATAMART_FIA • Disponibilidad del cubo Postulantes • Disponibilidad de Indicador en el Reporte diseñado en Excel 		
DATOS DE ENTRADA		Realizar la misma consulta		
FUNCIONALIDADES A PROBAR				
#	FUNCIÓN A PROBAR	ACCIÓN	RESULTADO ESPERADO	RESULTADO OBTENIDO
1	Valores en el reporte Excel frente a los valores del cubo	Realizar la consulta en el Reporte Excel y en el Cubo	El reporte Excel muestre los mismos datos que la consulta en el cubo	Muestra la misma información

Elaboración: los autores

5.4.1 Funcionalidad #1: Consistencia de Datos_Reportes

a) Consulta a realizar

Mostrar el indicador principal de alumnos no matriculados

b) Resultado del datamart

Ilustración CXVII – Resultado Cubo_Indicador

Mostrar estructura	Valor	Objetivo	Estado
KPI			
Alumnos_NO_Matriculados	120	100	

Elaboración: los autores

c) Resultado del cubo

Ilustración CXVIII – Resultado Excel_Indicador

ALUMNOS NO MATRICULADOS			
Alumnos_NO_Matriculados	Objetivo Alumnos_NO_Matriculad	Estado Alumnos_NO_Matriculados	Alumnos_NO_Matriculados
120	100		
ALUMNOS NO MATRICULADOS	Arquitectura_NO_Matriculados	Estado Arquitectura_NO_Matriculados	
	23		
	Aeronauticas_NO_Matriculados	Estado Aeronauticas_NO_Matriculados	
	11		
	Civil_NO_Matriculados	Estado Civil_NO_Matriculados	
	17		
	Sistemas_NO_Matriculados	Estado Sistemas_NO_Matriculados	
	16		
Electronica_NO_Matriculados	Estado Electronica_NO_Matriculados		
15			
Alimentarias_NO_Matriculados	Estado Alimentarias_NO_Matriculados		
24			
Industrial_NO_Matriculados	Estado Industrial_NO_Matriculados		
14			

Elaboración: los autores

CAPÍTULO VI

DISCUSIÓN Y APLICACIONES

6.1 Discusión y aplicación

A continuación, se muestra un contraste entre los objetivos específicos que se plantearon en la primera etapa de la tesis y las actividades que se realizaron a lo largo del desarrollo. Se describe también, de qué manera se aplicaron dichas actividades para lograr los objetivos planteados.

Tabla XCII - Discusión y aplicación

Objetivos específicos	Punto específico de aplicación	Descripción de la aplicación
Realizar análisis de negocio para la oficina de admisión y para la FIA - USMP, obteniendo los requerimientos e indicadores claves	4.2. Def. de los requerimientos del negocio	El análisis del negocio consistió en un conjunto de entrevistas que se realizaron a los diferentes roles tanto de la Oficina de Admisión como de la FIA - USMP. Las entrevistas fueron realizadas a roles pertenecientes a los tres niveles de la organización: nivel estratégico, táctico y operativo con el fin de poder cubrir con la totalidad de necesidades.
	4.2.1. Preparación y desarrollo de entrevistas	
	4.2.2. Definición de req. funcionales	
	4.2.3. Definición de req. no funcionales	
Identificar y analizar las diversas fuentes de datos que servirán como input para el Datamart de la Oficina de Registros Académicos.	4.2.4. Análisis de fuentes de datos	Se identificaron y analizaron las diferentes fuentes de datos para poblar los modelos.
	4.2.4.1. Fuentes Maestras	
	4.2.4.2. Otras fuentes	

Elaborar un modelo de base de datos multidimensional que permita la explotación de la información de ingresantes.	4.3. Modelamiento dimensional	Se definió el modelo de estructura de datos a utilizar en el modelo y se analizaron e identificaron las dimensiones, medidas y la tabla de hechos para posteriormente realizar el diseño del modelo lógico de base de datos de la tesis.
	4.3.1. Def. del modelo de la estructura de datos	
	4.3.2. Establecimiento del nivel de granularidad	
	4.3.3. Elección y detalle de las dimensiones	
	4.3.4. Identif. de medidas y tabla de hechos	
	4.3.5. Desarrollo del modelo lógico dimensional	Se brinda el esquema del almacenamiento que utilizamos en el desarrollo del datamart, También se muestra las tablas diseñadas con su respectivo tipo de datos, las cuales dan soporte al diseño físico. Luego se desarrolla el diseño físico del datamart
	4.4. Diseño físico	
	4.4.1. Def. del esquema de almacenamiento	
	4.4.2. Diseño de tablas y columnas	
	4.4.3. Diseño de Índices y Funciones	
4.4.4. Desarrollar Diseño Físico	Se diseñaron los procesos ETL para la extracción de los datos desde las distintas fuentes, aplicando reglas de limpieza de datos, para posteriormente cargar la información en el datamart.	
4.5. Diseño y desarrollo de los proc. de ETL		
4.5.1. Elaborar el mapeo de datos		
4.5.2. Diseño y desarrollo de los proc. de ETL		
Desarrollar los procesos de extracción, transformación y carga de los datos al modelo de base de datos multidimensional	4.5.3. Diseño de la secuencia de ejecución	La explotación de la información se realizó a través de un cubo de información (POSTULANTES) y a través de reportes, lo cuales pueden ser reportes frecuentes o personalizados
	4.8. Diseño y desarrollo de aplicaciones BI	
	4.8.1. Diseño y desarrollo de reportes	
Explotar la información a través de cubos OLAP y reportes.	4.8.2. Diseño y desarrollo de cubos OLAP	

Elaboración: los autores

Como se pudo observar, todos los objetivos específicos han sido cubiertos por alguna actividad en la presente tesis.

CONCLUSIONES

1. La metodología Kimball logró guiar exitosamente toda la construcción del datamart. Su definición de etapas claras permitió planificar y documentar el análisis, diseño e implementación de la solución de manera clara y sin ambigüedades. Si bien Kimball está orientado al desarrollo de todas las etapas de la metodología, se ha demostrado en la presente tesis que, empleando las etapas más críticas de la metodología, se pudo desarrollar la construcción de una solución basada en Inteligencia de Negocios.
2. Las necesidades de información comprendidas entre la Oficina de Admisión USMP y la FIA –USMP fueron identificadas satisfactoriamente debido a que se tuvieron entrevistas con los actores de negocio para recabar la información. Esto contribuyó a identificar requerimientos claros y precisos que fueron documentados y utilizados para la construcción del modelo multidimensional.
3. La correcta identificación y análisis de las fuentes de datos fueron fundamentales para el desarrollo del modelo dentro de los tiempos establecidos en la planificación.
4. El modelo multidimensional de la solución logró abarcar las necesidades de información identificadas y fue representado utilizando diagramas de fácil comprensión que permitieron una correcta validación de este.
5. Los procesos de extracción, transformación y carga de los datos, lograron poblar un datamart confiable, coherente y consistente que cumpla con los requisitos de información de los usuarios de la Oficina de Registros Académicos. Luego se verificó su funcionamiento a través de casos de prueba.
6. El cubo de información y los reportes elaborados permitieron mostrar la importancia de la explotación de la información y de los indicadores que generan una ventaja competitiva a la hora de tomar decisiones.

7. El datamart, permite a los usuarios contar con la información en el momento que la necesiten y para lo que la necesiten, sin dependencia del área de sistemas.

RECOMENDACIONES

En este punto se presentan algunas recomendaciones sobre el trabajo realizado en esta tesis, asimismo, se presentan algunas posibilidades de trabajos futuros que pueden ampliar las funcionalidades de este modelo.

1. Revisar los resultados de varios reportes. Los datos deben ser congruentes entre ellos e idénticos para los casos en que muestran lo mismo, dado los filtros.
2. Es muy importante desarrollar una buena fase de análisis para evitar que a lo largo del desarrollo del modelo surjan problemas que ameriten una reestructuración de los procesos, mapeos o de los reportes mismos. Algunos inconvenientes no saltan a la vista hasta que se tiene el reporte terminado, puesto que aparecen incongruencias en los datos del informe o se identifica que los datos no eran agregables. En estos casos, se debe regresar a los procesos anteriores para resolver el problema.
3. Se recomienda la utilización de herramientas que no sean libres y que cuenten con soporte y documentación. Esto permitirá levantar observaciones y errores con menor dificultad comparado a herramientas de uso libre que no cuenten con respaldo.
4. Este modelo ha sido diseñado y desarrollado para poder ser implementado o adaptado en diferentes facultades, hasta en diferentes sedes. De esta manera el alcance no solo estará limitado por la necesidad de información de la FIA –USMP sino que podría abarcar facultades como Ciencias de la Comunicación, Turismo y de Psicología, Ciencias Contables Económicas y Financieras, u otras facultades, O también la Filial Norte.

FUENTES DE INFORMACIÓN

Bibliográficas:

Cano, J. L. (2007). *Business Intelligence: Competir con Información* . Madrid: Fundación Cultural Banesto.

Kimball Ralph, C. J. (2004). *The Data Warehouse ETL Toolkit*. Indianapolis, United States of America: Wiley Publishing.

Electrónicas:

Araníbar, J. C. (Mayo de 2003). Recuperado el 19 de Marzo de 2014, de sitio web de Scientific Electronic Library Online: <http://www.scielo.org.bo/pdf/rcc/n12/a10.pdf>

Arroyo, Y. (Marzo de 2009). *Biblioteca: Univerdad Veracruzana*. Recuperado el 29 de Marzo de 2014, de sitio web de Univerdad Veracruzana: <http://cdigital.uv.mx/bitstream/123456789/29396/1/Arroyo%20Mtz.pdf>

Gartner Group. (1996). *Gartner Group Report*. Recuperado el 20 de Marzo de 2014, de <http://www.Gartner.com/>

Inmon, B. (2005). *Library: Inmon Corporate Information Factory*. Recuperado el 27 de Marzo de 2014, de Inmon Corporate Information Factory: <http://www.inmoncif.com/library/glossary/#D>

Instituto Tecnológico de Buenos Aires. (2009). *Tesis Magister: Instituto Tecnológico de Buenos Aires*. Recuperado el 30 de Marzo de 2014, de Sitio web de Instituto Tecnológico de Buenos Aires: <http://www2.itba.edu.ar/archivos/secciones/nader-tesisdemagister.pdf>

- Intel Corporation. (2009). *Curso: Business Intelligence*. Recuperado el 31 de Marzo de 2014, de http://www.tacticasoftware.com/CRM-CURSO/Business_Intelligence.pdf
- Microsoft Corporation. (2014). *Información general de arquitectura lógica (Analysis Services - Datos multidimensionales): Microsoft Corporation*. Recuperado el Abril de 2014, de sitio web de Microsoft Corporation: [http://technet.microsoft.com/es-es/library/ms174587\(v=sql.105\).aspx](http://technet.microsoft.com/es-es/library/ms174587(v=sql.105).aspx)
- Microsoft Corporation. (2014). *Libros en Pantalla de SQL Server: Microsoft Corporation*. Recuperado el 25 de Abril de 2014, de sitio web de Microsoft Corporation : [http://msdn.microsoft.com/es-es/library/ms130214\(v=sql.105\).aspx](http://msdn.microsoft.com/es-es/library/ms130214(v=sql.105).aspx)
- Pentaho Open Source Business Intelligence. (4 de Febrero de 2012). *Pentaho BI Suite: Pentaho Open Source Business Intelligence*. Recuperado el 24 de abril de 2014, de sitio web de Pentaho Open Source Business Intelligence: <https://sites.google.com/site/pentahobisuite/home/caracteristicas/pentaho-bi-platform>
- Politecnico Gran Colombiano. (2012). *Politecnico Gran Colombiano*. Recuperado el 28 de Marzo de 2014, de sitio web de Politecnico Gran Colombiano: <http://www.poli.edu.co/comunica/egresados/TallerdeFundamentosdeBIPGCD1.pdf>
- Universidad de Cordova. (2010). *Gestión: Universidad de Cordova*. Recuperado el 1 de Abril de 2014, de sitio web de Universidad de Cordova: http://www.gestion.uco.es/gestion/datawarehouse/doc/m_usuario/07000_olap.pdf
- Universidad de Drexel. (2011). *Universidad de Drexel*. Recuperado el Marzo de 2014, de http://www.cis.drexel.edu/faculty/song/courses/info%20607/tutorial_OLAP/MOLAP_sub.htm
- Universidad San Martín de Porres. (2014). *Inicio: Oficina de admisión USMP*. Recuperado el 1 de Mayo de 2014, de sitio web de Oficina de Admisión USMP: <http://admission.usmp.edu.pe/>

ANEXOS

INDICE DE ANEXOS

	Página
Anexo 1: Cronograma de la tesis	200
Anexo 2: Cronograma de actividades	207
Anexo 3: Asignación de tareas por rol	210
Anexo 4: Preparación y desarrollo de entrevistas	212
Anexo 5: Dominio de datos	214
Anexo 6: Funciones de base de datos	217
Anexo 7: Script de creación de base de datos – DB_DATAMART_FIA	220
Anexo 8: Mapeo fuente COLEGIOS.XLS	229
Anexo 9: Mapeo fuente UBIGEOS.XLS	231
Anexo 10: Mapeo fuente MODALIDAD.XLS	232
Anexo 11: Mapeo fuente ESPECIALIDAD.XLS	234
Anexo 12: Mapeo fuente DATOS_MATRICULA.XLS	237
Anexo 13: Mapeo fuente DATOS_POSTULANTE.XLS	238
Anexo 14: Mapeo fuente DATOS_POSTULANTE.XLS - ANÁLISIS	241
Anexo 15: Diseño de flujo de datos	245
Anexo 16: Definición de severidades	269
Anexo 17: KPI's – Indicadores clave de rendimiento	276
Anexo 18: Reportes Pre-Definidos	292

ANEXO 1
CRONOGRAMA DE LA TESIS

a. Cronograma de la tesis

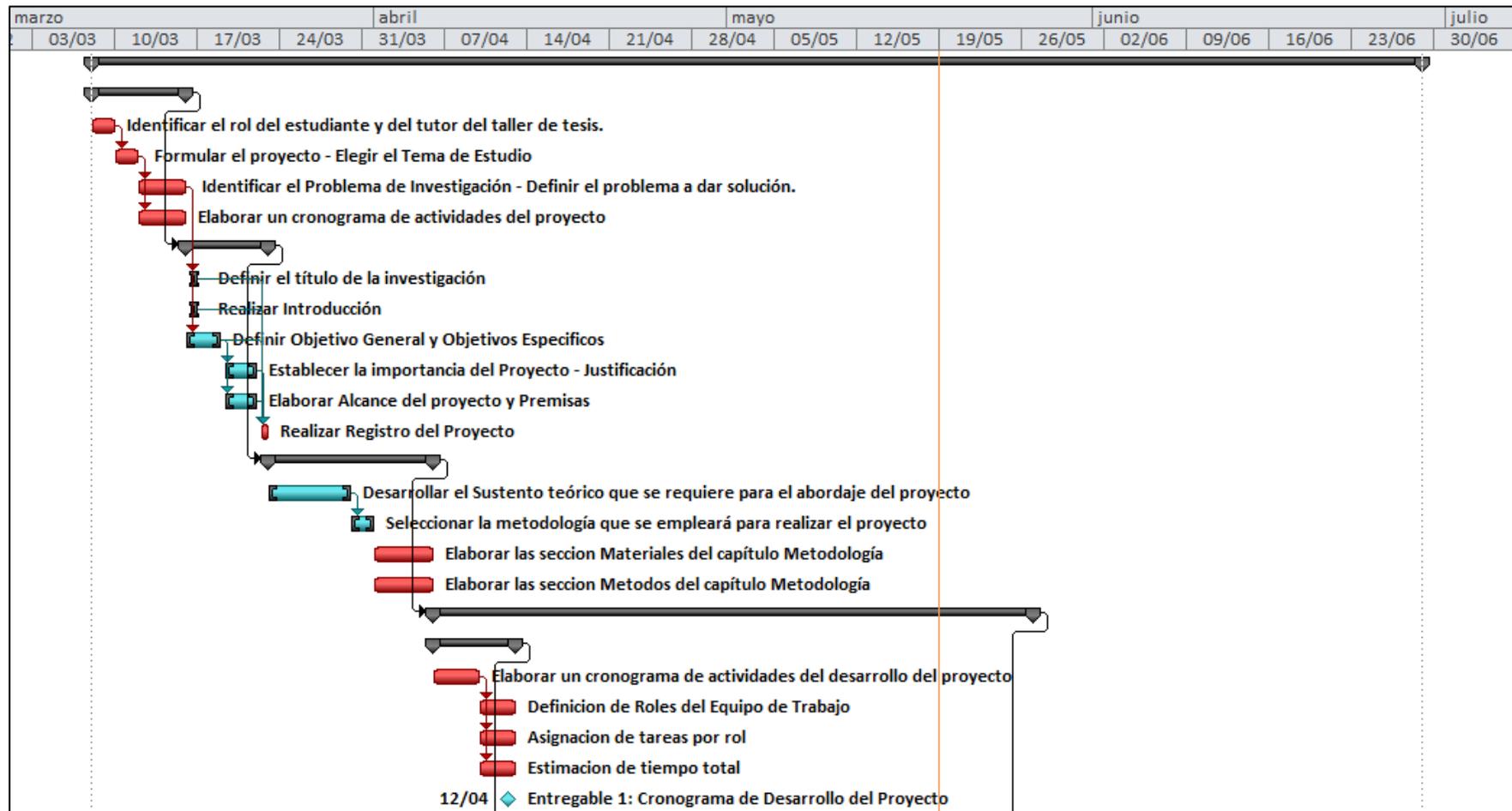
N °	Tareas	Duración	Comienzo	Fin	Predecesoras
1	DATAMART DE REGISTROS ACADÉMICOS - FIA	113 días	sáb 08/03/14	sáb 28/06/14	
2	EL PROYECTO DE TESIS	8 días	sáb 08/03/14	sáb 15/03/14	
3	Identificar el rol del estudiante y del tutor del taller de tesis.	2 días	sáb 08/03/14	dom 09/03/14	
4	Formular la tesis- Elegir el tema de estudio	2 días	lun 10/03/14	mar 11/03/14	3
5	Identificar el problema de investigación - Definir el problema a dar solución.	4 días	mié 12/03/14	sáb 15/03/14	4
6	Elaborar un cronograma de actividades de la tesis	4 días	mié 12/03/14	sáb 15/03/14	4
7	ELABORACIÓN DEL PROYECTO DE TESIS	7 días	dom 16/03/14	sáb 22/03/14	2
8	Definir el título de la investigación	1 día	dom 16/03/14	dom 16/03/14	5
9	Realizar Introducción	1 día	dom 16/03/14	dom 16/03/14	
10	Definir objetivo general y objetivos específicos	3 días	dom 16/03/14	mar 18/03/14	5
11	Establecer la importancia de la tesis - justificación	3 días	mié 19/03/14	vie 21/03/14	10
12	Elaborar alcance de la tesis y premisas	3 días	mié 19/03/14	vie 21/03/14	10
13	Realizar registro de la tesis	1 día	sáb 22/03/14	sáb 22/03/14	8,9,10,11,12
14	MARCO TEÓRICO Y METODOLOGIA DE LA TESIS	14 días	dom 23/03/14	sáb 05/04/14	7
15	Desarrollar el sustento teórico que se requiere para el abordaje de la tesis	7 días	dom 23/03/14	sáb 29/03/14	
16	Seleccionar la metodología que se empleará para realizar la tesis	2 días	dom 30/03/14	lun 31/03/14	15

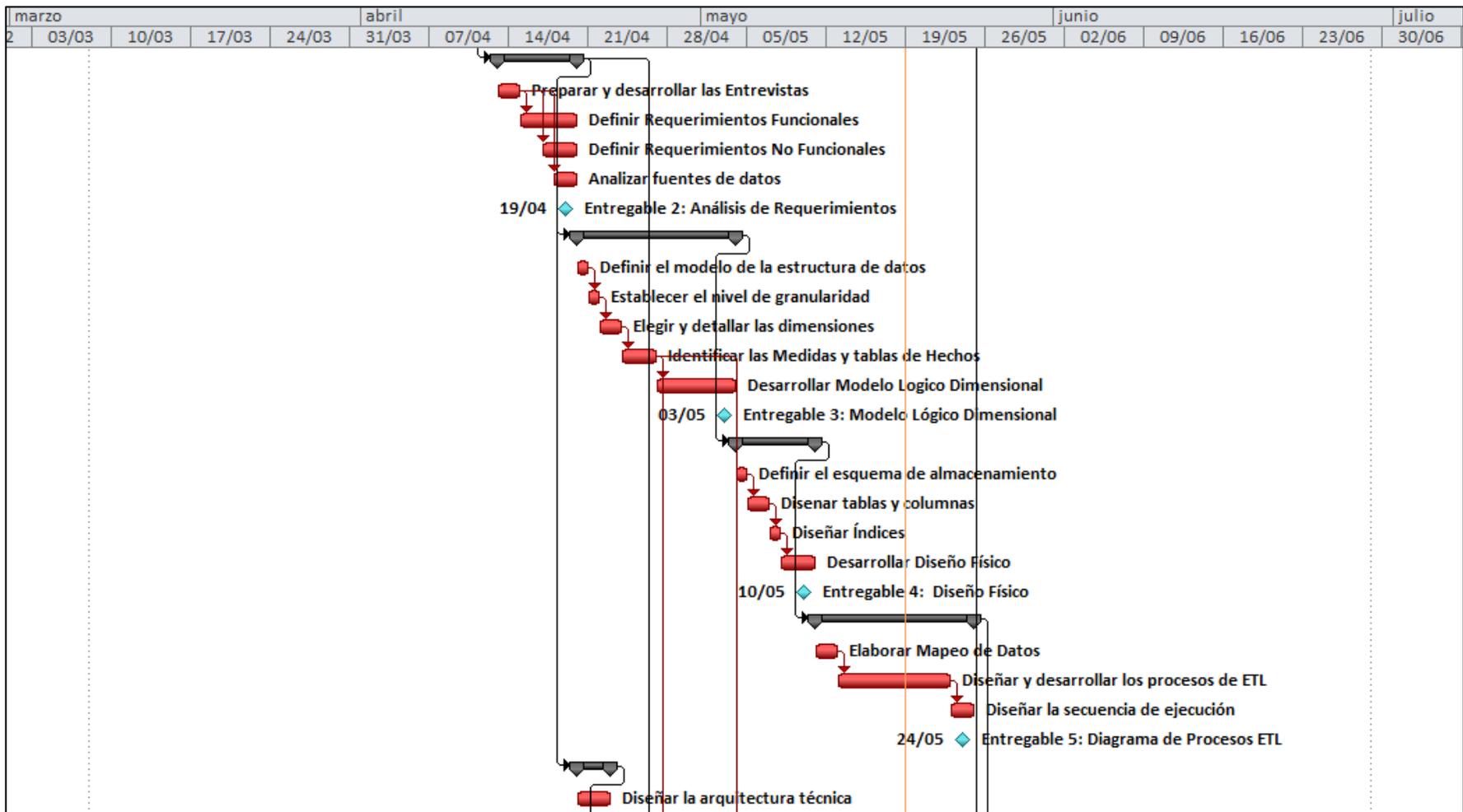
17	Elaborar la sección materiales del capítulo metodología	5 días	mar 01/04/14	sáb 05/04/14	
18	Elaborar la sección métodos del capítulo metodología	5 días	mar 01/04/14	sáb 05/04/14	
19	DESARROLLO DE LA TESIS	51 días	dom 06/04/14	lun 26/05/14	14
20	Planificación de la tesis	7 días	dom 06/04/14	sáb 12/04/14	
21	Elaborar un cronograma de actividades del desarrollo de la tesis	4 días	dom 06/04/14	mié 09/04/14	
22	Definición de roles del equipo de trabajo	3 días	jue 10/04/14	sáb 12/04/14	21
23	Asignación de tareas por rol	3 días	jue 10/04/14	sáb 12/04/14	21
24	Estimación de tiempo total	3 días	jue 10/04/14	sáb 12/04/14	21
25	Entregable 1: Cronograma de desarrollo de la tesis	0 días	sáb 12/04/14	sáb 12/04/14	
26	Definición de los requerimientos del negocio	7 días	dom 13/04/14	sáb 19/04/14	20
27	Preparar y desarrollar las entrevistas	2 días	dom 13/04/14	lun 14/04/14	
28	Definir requerimientos funcionales	5 días	mar 15/04/14	sáb 19/04/14	27
29	Definir requerimientos no funcionales	3 días	jue 17/04/14	sáb 19/04/14	27
30	Analizar fuentes de datos	2 días	vie 18/04/14	sáb 19/04/14	27
31	Entregable 2: Análisis de requerimientos	0 días	sáb 19/04/14	sáb 19/04/14	
32	Modelamiento dimensional	14 días	dom 20/04/14	sáb 03/05/14	26
33	Definir el modelo de la estructura de datos	1 día	dom 20/04/14	dom 20/04/14	
34	Establecer el nivel de granularidad	1 día	lun 21/04/14	lun 21/04/14	33
35	Elegir y detallar las dimensiones	2 días	mar 22/04/14	mié 23/04/14	34
36	Identificar las medidas y tablas de hechos	3 días	jue 24/04/14	sáb 26/04/14	35
37	Desarrollar modelo lógico dimensional	7 días	dom 27/04/14	sáb 03/05/14	36
38	Entregable 3: Modelo lógico dimensional	0 días	sáb 03/05/14	sáb 03/05/14	
39	Diseño físico	7 días	dom 04/05/14	sáb 10/05/14	32

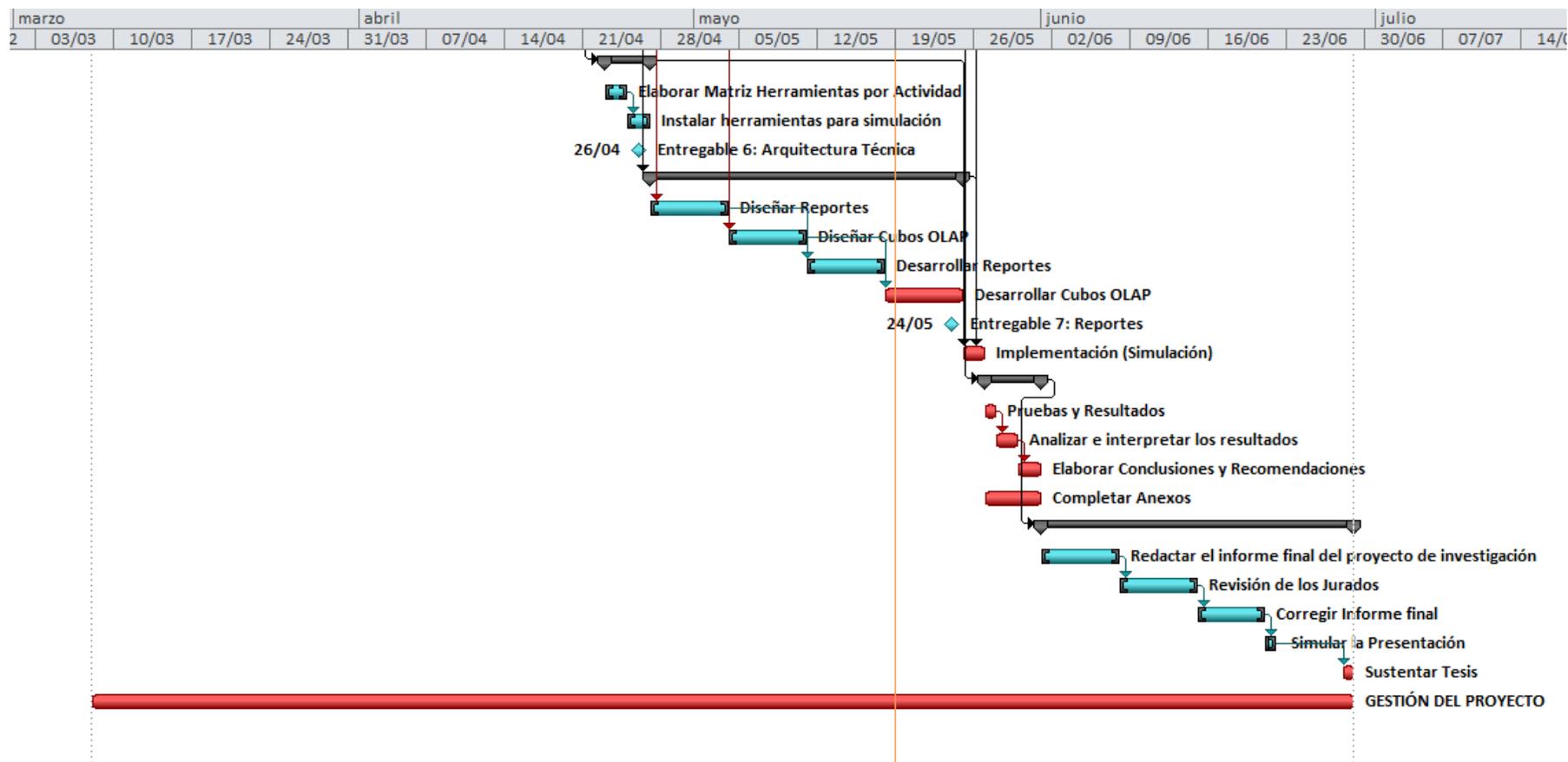
40	Definir el esquema de almacenamiento	1 día	dom 04/05/14	dom 04/05/14	
41	Diseñar tablas y columnas	2 días	lun 05/05/14	mar 06/05/14	40
42	Diseñar Índices	1 día	mié 07/05/14	mié 07/05/14	41
43	Desarrollar diseño físico	3 días	jue 08/05/14	sáb 10/05/14	42
44	Entregable 4: diseño físico	0 días	sáb 10/05/14	sáb 10/05/14	
45	Diseño y desarrollo de los procesos de ETL	14 días	dom 11/05/14	sáb 24/05/14	39
46	Elaborar mapeo de datos	2 días	dom 11/05/14	lun 12/05/14	
47	Diseñar y desarrollar los procesos de ETL	10 días	mar 13/05/14	jue 22/05/14	46
48	Diseñar la secuencia de ejecución	2 días	vie 23/05/14	sáb 24/05/14	47
49	Entregable 5: Diagrama de procesos ETL	0 días	sáb 24/05/14	sáb 24/05/14	
50	Diseño de la arquitectura técnica	3 días	dom 20/04/14	mar 22/04/14	26
51	Diseñar la arquitectura técnica	3 días	dom 20/04/14	mar 22/04/14	
52	Selección de los productos e instalación	4 días	mié 23/04/14	sáb 26/04/14	50
53	Elaborar matriz herramientas por actividad	2 días	mié 23/04/14	jue 24/04/14	
54	Instalar herramientas para simulación	2 días	vie 25/04/14	sáb 26/04/14	53
55	Entregable 6: Arquitectura técnica	0 días	sáb 26/04/14	sáb 26/04/14	
56	Diseño y desarrollo de aplicaciones BI	28 días	dom 27/04/14	sáb 24/05/14	26
57	Diseñar reportes	7 días	dom 27/04/14	sáb 03/05/14	36
58	Diseñar cubos OLAP	7 días	dom 04/05/14	sáb 10/05/14	36
59	Desarrollar reportes	7 días	dom 11/05/14	sáb 17/05/14	57
60	Desarrollar cubos OLAP	7 días	dom 18/05/14	sáb 24/05/14	58
61	Entregable 7: Reportes	0 días	sáb 24/05/14	sáb 24/05/14	
62	Implementación (Simulación)	2 días	dom 25/05/14	lun 26/05/14	52,45,56

63	PRUEBAS Y RESULTADOS	5 días	mar 27/05/14	sáb 31/05/14	19
64	Pruebas y resultados	1 día	mar 27/05/14	mar 27/05/14	
65	Analizar e interpretar los resultados	2 días	mié 28/05/14	jue 29/05/14	64
66	Elaborar conclusiones y recomendaciones	2 días	vie 30/05/14	sáb 31/05/14	65
67	Completar anexos	5 días	mar 27/05/14	sáb 31/05/14	
68	ELABORACIÓN DEL INFORME FINAL Y SUSTENTACIÓN DE LA TESIS	28 días	dom 01/06/14	sáb 28/06/14	63
69	Redactar el informe final del proyecto de investigación	7 días	dom 01/06/14	sáb 07/06/14	
70	Revisión de los jurados	7 días	dom 08/06/14	sáb 14/06/14	69
71	Corregir informe final	6 días	dom 15/06/14	vie 20/06/14	70
72	Simular la presentación	1 día	sáb 21/06/14	sáb 21/06/14	71
73	Sustentar tesis	1 día	sáb 28/06/14	sáb 28/06/14	72
74	GESTIÓN DEL PROYECTO	113 días	sáb 08/03/14	sáb 28/06/14	

b. Diagrama Gantt







ANEXO 2
CRONOGRAMA DE ACTIVIDADES

N °	Tareas	Duración	Comienzo	Fin	Predecesoras
1	DATAMART DE REGISTROS ACADÉMICOS - FIA	113 días	sáb 08/03/14	sáb 28/06/14	
19	DESARROLLO DE LA TESIS	51 días	dom 06/04/14	lun 26/05/14	14
20	Planificación de la tesis	7 días	dom 06/04/14	sáb 12/04/14	
21	Elaborar un cronograma de actividades del desarrollo de la tesis	4 días	dom 06/04/14	mié 09/04/14	
22	Definición de roles del equipo de trabajo	3 días	jue 10/04/14	sáb 12/04/14	21
23	Asignación de tareas por rol	3 días	jue 10/04/14	sáb 12/04/14	21
24	Estimación de tiempo total	3 días	jue 10/04/14	sáb 12/04/14	21
25	Entregable 1: Cronograma de desarrollo del proyecto	0 días	sáb 12/04/14	sáb 12/04/14	
26	Definición de los requerimientos del negocio	7 días	dom 13/04/14	sáb 19/04/14	20
27	Preparar y desarrollar las entrevistas	2 días	dom 13/04/14	lun 14/04/14	
28	Definir requerimientos funcionales	5 días	mar 15/04/14	sáb 19/04/14	27
29	Definir requerimientos no funcionales	3 días	jue 17/04/14	sáb 19/04/14	27
30	Analizar fuentes de datos	2 días	vie 18/04/14	sáb 19/04/14	27
31	Entregable 2: Análisis de requerimientos	0 días	sáb 19/04/14	sáb 19/04/14	
32	Modelamiento dimensional	14 días	dom 20/04/14	sáb 03/05/14	26
33	Definir el modelo de la estructura de datos	1 día	dom 20/04/14	dom 20/04/14	
34	Establecer el nivel de granularidad	1 día	lun 21/04/14	lun 21/04/14	33

35	Elegir y detallar las dimensiones	2 días	mar 22/04/14	mié 23/04/14	34
36	Identificar las medidas y tablas de hechos	3 días	jue 24/04/14	sáb 26/04/14	35
37	Desarrollar modelo lógico dimensional	7 días	dom 27/04/14	sáb 03/05/14	36
38	Entregable 3: Modelo lógico dimensional	0 días	sáb 03/05/14	sáb 03/05/14	
39	Diseño físico	7 días	dom 04/05/14	sáb 10/05/14	32
40	Definir el esquema de almacenamiento	1 día	dom 04/05/14	dom 04/05/14	
41	Diseñar tablas y columnas	2 días	lun 05/05/14	mar 06/05/14	40
42	Diseñar Índices	1 día	mié 07/05/14	mié 07/05/14	41
43	Desarrollar diseño físico	3 días	jue 08/05/14	sáb 10/05/14	42
44	Entregable 4: Diseño Físico	0 días	sáb 10/05/14	sáb 10/05/14	
45	Diseño y desarrollo de los procesos de ETL	14 días	dom 11/05/14	sáb 24/05/14	39
46	Elaborar mapeo de datos	2 días	dom 11/05/14	lun 12/05/14	
47	Diseñar y desarrollar los procesos de ETL	10 días	mar 13/05/14	jue 22/05/14	46
48	Diseñar la secuencia de ejecución	2 días	vie 23/05/14	sáb 24/05/14	47
49	Entregable 5: Diagrama de procesos ETL	0 días	sáb 24/05/14	sáb 24/05/14	
50	Diseño de la arquitectura técnica	3 días	dom 20/04/14	mar 22/04/14	26
51	Diseñar la arquitectura técnica	3 días	dom 20/04/14	mar 22/04/14	
52	Selección de los productos e instalación	4 días	mié 23/04/14	sáb 26/04/14	50
53	Elaborar matriz herramientas por actividad	2 días	mié 23/04/14	jue 24/04/14	
54	Instalar herramientas para simulación	2 días	vie 25/04/14	sáb 26/04/14	53
55	Entregable 6: Arquitectura técnica	0 días	sáb 26/04/14	sáb 26/04/14	
56	Diseño y desarrollo de aplicaciones BI	28 días	dom 27/04/14	sáb 24/05/14	26
57	Diseñar reportes	7 días	dom 27/04/14	sáb 03/05/14	36

58	Diseñar cubos OLAP	7 días	dom 04/05/14	sáb 10/05/14	36
59	Desarrollar reportes	7 días	dom 11/05/14	sáb 17/05/14	57
60	Desarrollar cubos OLAP	7 días	dom 18/05/14	sáb 24/05/14	58
61	Entregable 7: reportes	0 días	sáb 24/05/14	sáb 24/05/14	
62	Implementación (Simulación)	2 días	dom 25/05/14	lun 26/05/14	52,45,56

ANEXO 3
ASIGNACIÓN DE TAREAS POR ROL

N °	Tareas	Recursos
1	DATAMART DE REGISTROS ACADÉMICOS - FIA	
19	DESARROLLO DE LA TESIS	
20	Planificación de la tesis	
26	Definición de los requerimientos del negocio	
27	Preparar y desarrollar las entrevistas	Analista de negocio
28	Definir requerimientos funcionales	Analista de negocio
29	Definir requerimientos no funcionales	Analista de negocio
30	Analizar fuentes de datos	Analista de negocio
31	Entregable 2: Análisis de requerimientos	
32	Modelamiento dimensional	
33	Definir el modelo de la estructura de datos	Arquitecto y modelador de datos
34	Establecer el nivel de granularidad	Arquitecto y modelador de datos
35	Elegir y detallar las dimensiones	Arquitecto y modelador de datos
36	Identificar las medidas y tablas de hechos	Arquitecto y modelador de datos
37	Desarrollar Modelo lógico dimensional	Arquitecto y modelador de datos
38	Entregable 3: Modelo lógico dimensional	
39	Diseño físico	
40	Administrador de base de datos	Administrador de base de datos
41	Diseñar tablas y columnas	Administrador de base de datos
42	Diseñar índices	Administrador de base de datos
43	Desarrollar Diseño físico	Administrador de base de datos
44	Entregable 4: Diseño físico	
45	Diseño y desarrollo de los procesos de ETL	
46	Elaborar mapeo de datos	Diseñador y desarrollador ETL, Analista de Calidad
47	Diseñar y desarrollar los procesos de ETL	Diseñador y desarrollador ETL, Analista de calidad
48	Diseñar la secuencia de ejecución	Diseñador y desarrollador ETL, Analista de calidad
49	Entregable 5: Diagrama de procesos ETL	
50	Diseño de la arquitectura técnica	Arquitecto técnico

51	Diseñar la arquitectura técnica	Arquitecto técnico
52	Selección de los productos e instalación	Arquitecto técnico
53	Elaborar matriz herramientas por actividad	Arquitecto técnico
54	Instalar herramientas para simulación	Arquitecto técnico
55	Entregable 6: Arquitectura técnica	
56	Diseño y desarrollo de aplicaciones BI	
57	Diseñar reportes	Diseñador y desarrollador BI
58	Diseñar cubos OLAP	Diseñador y desarrollador BI
59	Desarrollar reportes	Diseñador y desarrollador BI
60	Desarrollar cubos OLAP	Diseñador y desarrollador BI
61	Entregable 7: Reportes	
62	Implementación (Simulación)	Analista de calidad, Administrador de base de datos, Arquitecto técnico, Diseñador y desarrollador BI, Diseñador y desarrollador ETL, Arquitecto y modelador de datos, Analista de pruebas

ANEXO 4

PREPARACIÓN Y DESARROLLO DE ENTREVISTAS

1. ¿Qué análisis le gustaría llevar a cabo?
2. ¿Cuál cree que sería el impacto financiero de un proyecto de inteligencia de negocios dirigido a la Oficina de Registros Académicos de la facultad?
3. ¿Qué indicadores tiene usted para medir el desempeño de la Oficina de Registros Académicos?
4. ¿Cuál es su posición dentro de la organización?
5. ¿Usted estaría de acuerdo en apoyar un proyecto de inteligencia de negocios en la facultad? ¿Por qué motivo?
6. ¿Principales responsabilidades?
7. ¿Qué Objetivos tiene y como apoya a la organización?
8. ¿Cuáles son sus indicadores de éxito?
9. ¿Qué impide cumplir con los objetivos, cual es el impacto en la organización?
10. ¿Qué tipo de análisis de rutina se realiza en la actualidad?
11. ¿Qué análisis le gustaría llevar a cabo respecto de la información de admisión?
Preguntas que se quieran resolver Ejm.: ¿Cuáles son las carreras que son más escogidas por los postulantes?
12. ¿Qué informes (reportes) utiliza actualmente?
13. ¿Qué datos sobre los informes (reportes) son importante?
14. ¿Qué informes haría diferente?
15. ¿Cuánta información histórica requiere?
16. ¿Qué oportunidades existen para mejorar su departamento basado en la mejora del acceso a la información?
17. ¿Qué indicadores le gustaría medir?
18. Descripción de los distintos procesos de admisión. Algún flujo, ppt o cualquier documentación que respalde dichos procesos.
19. ¿Qué datos utilizan?
20. ¿Cómo actualmente obtiene los datos?
21. ¿Qué hace con la información cuando la tienen?
22. ¿Existen posibles mejoras a su método/proceso actual?
23. ¿Realiza análisis de reportes AD-HOC? ¿Qué hacen con el análisis?

24. ¿Qué informes (reportes) utiliza actualmente? ¿Alguna documentación al respecto?
25. ¿Cuánto tiempo demora la elaboración de los reportes?, ¿Cuáles son los recursos utilizados, quienes los realizan?
26. ¿En cuánto tiempo entregan a cada facultad sus respectivos reportes? ¿A través de que medio (físico, electrónico)?
27. ¿Cuánto tiempo aproximadamente demoran las consultas que se hacen a la base de datos para la elaboración de los reportes?
28. ¿Cuántos reportes se entregan? ¿Con qué frecuencia, diario, mensual, semanal, semestrales, anual?
29. Brindar los diferentes reportes elaborados para la FIA.
30. ¿Qué otros proyectos de inteligencia de negocios se han realizados en el área?
Describir
31. Debido al análisis que se desea realizar sobre la información de los postulantes e ingresantes a la FIA, se requiere una profundidad histórica de 2 años de información. (4 procesos de admisión como mínimo) Por favor coordinemos para la entrega de dicha información.
32. ¿El área del base de datos de la facultad, maneja información de admisión? ¿qué información?
33. En el caso no sea así, ¿Cómo se envía la información de Admisión a la Oficina de Registros Académicos?
34. ¿En qué base de datos se encuentra almacenada la información de postulantes e ingresantes?
35. ¿En qué base de datos se encuentra almacenada la información de matrícula?
36. ¿Cuál es el motor de base de datos que se utiliza actualmente en la facultad? ¿Y En la oficina de admisión?
37. ¿Qué otra información relacionada a admisión se tiene en la base de datos?
38. Brindar el diseño físico o estructura de la base de datos de admisión. (Tablas, campos, tipo de datos, relaciones)
39. ¿Cómo se carga la información de ingresantes a la base de datos, brindada por la Oficina de Admisión?
40. ¿Cuántos reportes se entregan? ¿Con qué frecuencia diario, mensual, semanal, semestrales, anual?

ANEXO 5

DOMINIO DE DATOS

El dominio de datos definido para las dimensiones que no cambian en el tiempo es el siguiente:

DIM_ESTADO_CIVIL	
VALORES	DESCRIPCIÓN
'SOLTERO'	Postulantes que sean solteros
'CASADO'	Postulantes que sean casados
'DIVORCIADO'	Postulantes que sean divorciados
'VIUDO'	Postulantes que sean viudos

DIM_LETRA_INICIAL_APE	
VALORES	DESCRIPCIÓN
Letras desde la A a la Z	Todas las letras del abecedario

DIM_GENERO	
VALORES	DESCRIPCIÓN
'MASCULINO'	Genero Masculino
'FEMENINO'	Genero Femenino

DIM_INGRESO	
VALORES	DESCRIPCIÓN
'SI'	El postulante ingresa a al universidad
'NO'	El postulante no ingresa a la universidad

DIM_MATRICULA	
VALORES	DESCRIPCIÓN
'SI'	El ingresante matriculado
'NO'	El ingresante no matriculado

DIM_TRABAJA	
VALORES	DESCRIPCIÓN
'SI'	El postulante trabaja

'NO'	El postulante no trabaja
------	--------------------------

DIM_TIPO_COLEGIO	
VALORES	DESCRIPCIÓN
'PUBLICO'	Hace referencia a colegios publicos
'PRIVADO'	Hace referencia a colegios privados

DIM_RANGO_EDAD	
VALORES	DESCRIPCIÓN
'Edad entre: [- - 19>'	Rango de edades entre 0 y 19 (abierto)
'Edad entre: [19 - 22>'	Rango de edades entre 19 y 22 (abierto)
'Edad entre: [22 - 25>'	Rango de edades entre 22 y 25 (abierto)
'Edad entre: [25 - 28>'	Rango de edades entre 25 y 28 (abierto)
'Edad entre: [28 - +>'	Rango de edades de 28 a más

DIM_RANGO_EGRESO	
VALORES	DESCRIPCIÓN
'Años de Egreso entre: [0 - 1>'	Rango de años de egreso entre 0 y 1 (abierto)
'Años de Egreso entre: [1 - 4>'	Rango de años de egreso entre 1 y 4 (abierto)
'Años de Egreso entre: [4 - +>'	Rango de años de egreso de 4 a más (abierto)

DIM_RANGO_PUNTAJE	
VALORES	DESCRIPCIÓN
'Puntaje entre: [0 - 130> MUY BAJO'	Rango de puntaje entre 0 y 130 (abierto) - MUY BAJO
'Puntaje entre: [130 - 170> BAJO'	Rango de puntaje entre 130 y 170 (abierto) - BAJO
'Puntaje entre: [170 - 250> NORMAL'	Rango de puntaje entre 170 y 250 (abierto) - NORMAL
'Puntaje entre: [250 - 270> ALTO'	Rango de puntaje entre 250 y 270 (abierto) - ALTO
'Puntaje entre: [270 - 300> MUY ALTO'	Rango de puntaje entre 270 y 999 (abierto) - MUY ALTO

DIM_RANGO_MERITO	
VALORES	DESCRIPCIÓN
'Rango de Merito entre: [1 - 50> MUY ALTO'	Rango de Merito entre 1 y 50 (abierto) – MUY ALTO
'Rango de Merito entre: [50 - 120> ALTO'	Rango de Merito entre 50 y 120 (abierto) – ALTO

'Rango de Merito entre: [120 - 200> NORMAL ALTO'	Rango de Merito entre 120 y 200 (abierto) – NORMAL ALTO
'Rango de Merito entre: [200 - 320> NORMAL'	Rango de Merito entre 200 y 320 (abierto) – NORMAL
'Rango de Merito entre: [320 - 460> BAJO'	Rango de Merito entre 320 y 460 (abierto) – BAJO
'Rango de Merito entre: [460 - +> MUY BAJO'	Rango de Merito entre 460 y 9999 (abierto) – MUY BAJO

ANEXO 6

FUNCIONES DE BASE DE DATOS

En el proceso de ETL, se vio necesario por motivos de performance y menor complejidad, utilizar un conjunto de funciones para obtener los identificadores únicos de cada dimensión relacionada a un rango específico.

Se definieron cuatro funciones de rangos que se utilizaron a lo largo del proceso ETL relacionadas a cuatro variables medibles: edad, año de egreso, mérito y puntaje. Para todos los casos, el parámetro input es el valor que se quiere situar dentro del rango y el output de la función es el identificador único del rango (ID).

Además, se utilizó una función para calcular la edad de una persona en base a su fecha de nacimiento y la fecha en que se ejecuta la carga del datamart.

- **Función FN_CALCULO_EDAD**

```
CREATE FUNCTION FN_CALCULO_EDAD (@Fecha datetime)
Returns int
AS
BEGIN
Declare @Edad smallint
Declare @EDAD_STRING NVARCHAR(8)
SET @EDAD_STRING = CONVERT(NVARCHAR(8), CONVERT(DATE, @Fecha, 103), 112)
set @Edad=(datediff(day, @EDAD_STRING, GETDATE())-datediff(yy, @EDAD_STRING,
GETDATE()))/4)/365
return(@Edad)
END
GO
```

- **Función FN_CALCULAR_RANGO_EGRESO**

```
CREATE FUNCTION FN_CALCULAR_RANGO_EGRESO (@VALOR SMALLINT)
RETURNS SMALLINT
AS
BEGIN
DECLARE @VALOR_1 SMALLINT, @ID SMALLINT, @ID_RANGO SMALLINT, @MIN_VAL
SMALLINT, @MAX_VAL SMALLINT
SET @VALOR_1 = CAST(YEAR(GETDATE()) AS SMALLINT) - @VALOR - 1
DECLARE RANGOS CURSOR FOR
    SELECT ID_RANGO_EGRESO, MIN_VAL, MAX_VAL FROM DBO.DIM_RANGO_EGRESO
OPEN RANGOS
    FETCH RANGOS INTO @ID_RANGO, @MIN_VAL, @MAX_VAL
    WHILE (@@FETCH_STATUS = 0 )
    BEGIN
        IF @VALOR_1 >= @MIN_VAL AND @VALOR_1 < @MAX_VAL
            SET @ID = @ID_RANGO
        FETCH RANGOS INTO @ID_RANGO, @MIN_VAL, @MAX_VAL
    END
END
```

```

        END
    CLOSE RANGOS
    DEALLOCATE RANGOS
    RETURN (@ID)
END
GO

```

- **Función FN_CALCULAR_RANGO_EDAD**

```

CREATE FUNCTION FN_CALCULAR_RANGO_EDAD (@VALOR SMALLINT)
RETURNS SMALLINT
AS
BEGIN
    DECLARE @ID SMALLINT, @ID_RANGO SMALLINT, @MIN_VAL SMALLINT, @MAX_VAL
    SMALLINT
    DECLARE RANGOS CURSOR FOR
        SELECT ID_RANGO_EDAD, MIN_VAL, MAX_VAL FROM DBO.DIM_RANGO_EDAD
    OPEN RANGOS
    FETCH RANGOS INTO @ID_RANGO, @MIN_VAL, @MAX_VAL
    WHILE (@@FETCH_STATUS = 0 )
    BEGIN
        IF @VALOR >= @MIN_VAL AND @VALOR < @MAX_VAL
            SET @ID = @ID_RANGO
        FETCH RANGOS INTO @ID_RANGO, @MIN_VAL, @MAX_VAL
    END
    CLOSE RANGOS
    DEALLOCATE RANGOS
    RETURN (@ID)
END

GO

```

- **Función FN_CALCULAR_RANGO_MERITO**

```

CREATE FUNCTION FN_CALCULAR_RANGO_MERITO (@VALOR SMALLINT)
RETURNS SMALLINT
AS
BEGIN
    DECLARE @ID SMALLINT, @ID_RANGO SMALLINT, @MIN_VAL SMALLINT, @MAX_VAL
    SMALLINT
    DECLARE RANGOS CURSOR FOR
        SELECT ID_RANGO_MERITO, MIN_VAL, MAX_VAL FROM DBO.DIM_RANGO_MERITO
    OPEN RANGOS
    FETCH RANGOS INTO @ID_RANGO, @MIN_VAL, @MAX_VAL
    WHILE (@@FETCH_STATUS = 0 )
    BEGIN
        IF @VALOR >= @MIN_VAL AND @VALOR < @MAX_VAL
            SET @ID = @ID_RANGO
        FETCH RANGOS INTO @ID_RANGO, @MIN_VAL, @MAX_VAL
    END
    CLOSE RANGOS
    DEALLOCATE RANGOS
    RETURN (@ID)
END

GO

```

- **Función FN_CALCULAR_RANGO_PUNTAJE**

```
CREATE FUNCTION FN_CALCULAR_RANGO_PUNTAJE (@VALOR SMALLINT)
RETURNS SMALLINT
AS
BEGIN
DECLARE @ID SMALLINT, @ID_RANGO SMALLINT, @MIN_VAL SMALLINT, @MAX_VAL
SMALLINT
DECLARE RANGOS CURSOR FOR
    SELECT ID_RANGO_PUNTAJE, MIN_VAL, MAX_VAL FROM DBO.DIM_RANGO_PUNTAJE
OPEN RANGOS
    FETCH RANGOS INTO @ID_RANGO, @MIN_VAL, @MAX_VAL
    WHILE (@@FETCH_STATUS = 0 )
    BEGIN
        IF @VALOR >= @MIN_VAL AND @VALOR < @MAX_VAL
            SET @ID = @ID_RANGO
        FETCH RANGOS INTO @ID_RANGO, @MIN_VAL, @MAX_VAL
    END
CLOSE RANGOS
DEALLOCATE RANGOS
RETURN (@ID)
END

GO
```

ANEXO 7

SCRIPT DE CREACIÓN DE BASE DE DATOS – DB_DATAMART_FIA

```
CREATE TABLE DIM_CARRERA
(
ID_CARRERA smallint IDENTITY ( 1,1 ),
CARRERA nvarchar(100) NOT NULL ,
CARRERA_ABREV nvarchar(50) NOT NULL ,
ID_FACULTAD smallint NOT NULL ,
FEC_MOD smalldatetime NOT NULL
)
go
```

```
ALTER TABLE DIM_CARRERA
ADD CONSTRAINT XPKDIM_CARRERA PRIMARY KEY
CLUSTERED (ID_CARRERA ASC)
Go
```

```
CREATE TABLE DIM_COLEGIO
(
ID_COLEGIO smallint IDENTITY ( 1,1 ),
COLEGIO nvarchar(100) NOT NULL ,
ID_TIPO_COLEGIO smallint NOT NULL ,
DIRECCION_COLEGIO nvarchar(120) NULL ,
ID_UBIGEO_COLEGIO smallint NOT NULL ,
FEC_MOD smalldatetime NOT NULL
)
Go
ALTER TABLE DIM_COLEGIO
```

```
ADD CONSTRAINT XPKDIM_COLEGIO PRIMARY KEY
CLUSTERED (ID_COLEGIO ASC)
```

Go

```
CREATE TABLE DIM_ESTADO_CIVIL
(
ID_ESTADO_CIVIL smallint IDENTITY ( 1,1 ),
ESTADO_CIVIL nvarchar(15) NOT NULL ,
FEC_MOD smalldatetime NOT NULL
)
go
```

```
ALTER TABLE DIM_ESTADO_CIVIL
ADD CONSTRAINT XPKDIM_ESTADO_CIVIL PRIMARY
KEY CLUSTERED (ID_ESTADO_CIVIL ASC)
go
```

```
CREATE TABLE DIM_FACULTAD
(
ID_FACULTAD smallint IDENTITY ( 1,1 ),
FACULTAD nvarchar(100) NOT NULL ,
FACULTAD_ABREV nvarchar(50) NOT NULL ,
ID_SEDE smallint NOT NULL ,
FEC_MOD smalldatetime NOT NULL
)
go
```

```
ALTER TABLE DIM_FACULTAD
ADD CONSTRAINT XPKDIM_FACULTAD PRIMARY KEY
CLUSTERED (ID_FACULTAD ASC)
```

```

go

CREATE TABLE DIM_GENERO

(

ID_GENERO smallint IDENTITY ( 1,1 ) ,

GENERO nvarchar(15) NOT NULL ,

FEC_MOD smalldatetime NOT NULL)

go

ALTER TABLE DIM_GENERO

ADD CONSTRAINT XPKDIM_GENERO PRIMARY KEY

CLUSTERED (ID_GENERO ASC)

go

CREATE TABLE DIM_INGRESO

(

ID_INGRESO smallint IDENTITY ( 1,1 ) ,

INGRESO char(2) NOT NULL ,

FEC_MOD smalldatetime NOT NULL

)

go

ALTER TABLE DIM_INGRESO

ADD CONSTRAINT XPKDIM_INGRESO PRIMARY KEY

CLUSTERED (ID_INGRESO ASC)

go

CREATE TABLE DIM_LETRA_INICIAL_APE

(

ID_LETRA_INICIAL_APE smallint IDENTITY ( 1,1 ) ,

LETRA_INICIAL_APE char(1) NOT NULL ,

FEC_MOD smalldatetime NOT NULL

)

go

ALTER TABLE DIM_LETRA_INICIAL_APE

```

```

ADD CONSTRAINT XPKDIM_LETRA_INICIAL_APE

PRIMARY KEY CLUSTERED (ID_LETRA_INICIAL_APE

ASC)

go

CREATE TABLE DIM_MATRICULA

(

ID_MATRICULA smallint IDENTITY ( 1,1 ) ,

MATRICULA char(2) NOT NULL ,

FEC_MOD smalldatetime NOT NULL

)

go

ALTER TABLE DIM_MATRICULA

ADD CONSTRAINT XPKDIM_MATRICULA PRIMARY KEY

CLUSTERED (ID_MATRICULA ASC)

go

CREATE TABLE DIM_MODALIDAD

(

ID_MODALIDAD smallint IDENTITY ( 1,1 ) ,

MODALIDAD nvarchar(50) NOT NULL ,

ID_TIPO_PRUEBA smallint NOT NULL ,

FEC_MOD smalldatetime NOT NULL

)

go

ALTER TABLE DIM_MODALIDAD

ADD CONSTRAINT XPKDIM_MODALIDAD PRIMARY KEY

CLUSTERED (ID_MODALIDAD ASC)

go

CREATE TABLE DIM_PROCESO

(

ID_PROCESO smallint IDENTITY ( 1,1 ) ,

ANIO smallint NOT NULL ,

```

```

SEMESTRE smallint NOT NULL ,

FEC_MOD smalldatetime NOT NULL

)

go

ALTER TABLE DIM_PROCESO

ADD CONSTRAINT XPKDIM_PROCESO PRIMARY KEY
CLUSTERED (ID_PROCESO ASC)

go

CREATE TABLE DIM_RANGO_EDAD

(

ID_RANGO_EDAD smallint IDENTITY ( 1,1 ) ,

RANGO_EDAD nvarchar(50) NOT NULL ,

MIN_VAL smallint NOT NULL ,

MAX_VAL smallint NOT NULL ,

FEC_MOD smalldatetime NOT NULL

)

go

ALTER TABLE DIM_RANGO_EDAD

ADD CONSTRAINT XPKDIM_RANGO_EDAD PRIMARY KEY
CLUSTERED (ID_RANGO_EDAD ASC)

go

CREATE TABLE DIM_RANGO_EGRESO

(

ID_RANGO_EGRESO smallint IDENTITY ( 1,1 ) ,

RANGO_EGRESO nvarchar(50) NOT NULL ,

MIN_VAL smallint NOT NULL ,

MAX_VAL smallint NOT NULL ,

FEC_MOD smalldatetime NOT NULL

)

go

```

```

ALTER TABLE DIM_RANGO_EGRESO

ADD CONSTRAINT XPKDIM_RANGO_EGRESO PRIMARY
KEY CLUSTERED (ID_RANGO_EGRESO ASC)

go

CREATE TABLE DIM_RANGO_MERITO

(

ID_RANGO_MERITO smallint IDENTITY ( 1,1 ) ,

RANGO_MERITO nvarchar(50) NOT NULL ,

MIN_VAL smallint NOT NULL ,

MAX_VAL smallint NOT NULL ,

FEC_MOD smalldatetime NOT NULL

)

go

ALTER TABLE DIM_RANGO_MERITO

ADD CONSTRAINT XPKDIM_RANGO_MERITO PRIMARY
KEY CLUSTERED (ID_RANGO_MERITO ASC)

go

CREATE TABLE DIM_RANGO_PUNTAJE

(

ID_RANGO_PUNTAJE smallint IDENTITY ( 1,1 ) ,

RANGO_PUNTAJE nvarchar(50) NOT NULL ,

MIN_VAL smallint NOT NULL ,

MAX_VAL smallint NOT NULL ,

FEC_MOD smalldatetime NOT NULL

)

go

ALTER TABLE DIM_RANGO_PUNTAJE

ADD CONSTRAINT XPKDIM_RANGO_PUNTAJE PRIMARY
KEY CLUSTERED (ID_RANGO_PUNTAJE ASC)

go

CREATE TABLE DIM_SEDE

```

```

(
ID_SEDE smallint IDENTITY ( 1,1 ) ,
SEDE nvarchar(30) NOT NULL ,
FEC_MOD smalldatetime NOT NULL
)
go
ALTER TABLE DIM_SEDE
ADD CONSTRAINT XPKDIM_SEDE PRIMARY KEY
CLUSTERED (ID_SEDE ASC)
go
CREATE TABLE DIM_TIPO_COLEGIO
(
ID_TIPO_COLEGIO smallint IDENTITY ( 1,1 ) ,
TIPO_COLEGIO nvarchar(20) NOT NULL ,
FEC_MOD smalldatetime NOT NULL
)
go
ALTER TABLE DIM_TIPO_COLEGIO
ADD CONSTRAINT XPKDIM_TIPO_COLEGIO PRIMARY
KEY CLUSTERED (ID_TIPO_COLEGIO ASC)
go
CREATE TABLE DIM_TIPO_PRUEBA
(
ID_TIPO_PRUEBA smallint IDENTITY ( 1,1 ) ,
TIPO_PRUEBA nvarchar(50) NOT NULL ,
DETALLE_PRUEBA nvarchar(100) NULL ,
FEC_MOD smalldatetime NOT NULL
)
go
ALTER TABLE DIM_TIPO_PRUEBA

```

```

ADD CONSTRAINT XPKDIM_TIPO_PRUEBA PRIMARY KEY
CLUSTERED (ID_TIPO_PRUEBA ASC)
go
CREATE TABLE DIM_TRABAJO
(
ID_TRABAJO smallint IDENTITY ( 1,1 ) ,
TRABAJO char(2) NOT NULL ,
FEC_MOD smalldatetime NOT NULL
)
go
ALTER TABLE DIM_TRABAJO
ADD CONSTRAINT XPKDIM_TRABAJO PRIMARY KEY
CLUSTERED (ID_TRABAJO ASC)
go
CREATE TABLE DIM_UBIGEO
(
ID_UBIGEO smallint IDENTITY ( 1,1 ) ,
UBIGEO nvarchar(50) NOT NULL ,
ID_UBIGEO_PADRE smallint NULL ,
FEC_MOD smalldatetime NOT NULL
)
go
ALTER TABLE DIM_UBIGEO
ADD CONSTRAINT XPKDIM_UBIGEO PRIMARY KEY
CLUSTERED (ID_UBIGEO ASC)
go
CREATE TABLE DIM_UBIGEO_COLEGIO
(
ID_UBIGEO_COLEGIO smallint NOT NULL ,
UBIGEO_COLEGIO nvarchar(50) NOT NULL ,
ID_UBI_COL_PADRE smallint NULL ,

```

```

FEC_MOD smalldatetime NOT NULL
)
go

ALTER TABLE DIM_UBIGEO_COLEGIO

ADD CONSTRAINT XPKDIM_UBIGEO_COLEGIO PRIMARY
KEY CLUSTERED (ID_UBIGEO_COLEGIO ASC)

go

CREATE TABLE FACT_ALUMNO

(

ID_ALUMNO nvarchar(10) NOT NULL ,

NOMBRE nvarchar(100) NOT NULL ,

FEC_NACIMIENTO smalldatetime NOT NULL ,

EDAD smallint NOT NULL ,

TELEFONO_1 nvarchar(15) NULL ,

TELEFONO_2 nvarchar(15) NULL ,

PUNTAJE smallint NOT NULL ,

MERITO smallint NOT NULL ,

DIRECCION nvarchar(150) NOT NULL ,

EMAIL nvarchar(60) NULL ,

TIPO_DOCUMENTO nvarchar(40) NOT NULL ,

NUM_DOCUMENTO nvarchar(15) NOT NULL ,

FEC_MATRICULA smalldatetime NULL ,

ID_UBIGEO smallint NOT NULL ,

ID_MODALIDAD smallint NOT NULL ,

ID_COLEGIO smallint NOT NULL ,

ID_CARRERA smallint NOT NULL ,

ID_GENERO smallint NOT NULL ,

ID_PROCESO smallint NOT NULL ,

ID_RANGO_EDAD smallint NOT NULL ,

ID_RANGO_MERITO smallint NOT NULL ,

```

```

ID_RANGO_PUNTAJE smallint NOT NULL ,

ID_LETRA_INICIAL_APE smallint NOT NULL ,

ID_RANGO_EGRESO smallint NOT NULL ,

ID_ESTADO_CIVIL smallint NOT NULL ,

ID_TRABAJA smallint NOT NULL ,

ID_MATRICULA smallint NOT NULL ,

ID_INGRESO smallint NOT NULL ,

FEC_MOD smalldatetime NOT NULL

)

go

ALTER TABLE FACT_ALUMNO

ADD CONSTRAINT XPKFACT_ALUMNO PRIMARY KEY
CLUSTERED (ID_ALUMNO ASC)

go

ALTER TABLE DIM_CARRERA

ADD CONSTRAINT R_22 FOREIGN KEY (ID_FACULTAD)
REFERENCES DIM_FACULTAD(ID_FACULTAD)

ON DELETE NO ACTION

ON UPDATE NO ACTION

go

ALTER TABLE DIM_COLEGIO

ADD CONSTRAINT R_23 FOREIGN KEY
(ID_TIPO_COLEGIO) REFERENCES
DIM_TIPO_COLEGIO(ID_TIPO_COLEGIO)

ON DELETE NO ACTION

ON UPDATE NO ACTION

go

ALTER TABLE DIM_COLEGIO

ADD CONSTRAINT R_33 FOREIGN KEY
(ID_UBIGEO_COLEGIO) REFERENCES
DIM_UBIGEO_COLEGIO(ID_UBIGEO_COLEGIO)

```

```

ON DELETE NO ACTION

ON UPDATE NO ACTION

go

ALTER TABLE DIM_FACULTAD

ADD CONSTRAINT R_21 FOREIGN KEY (ID_SEDE)
REFERENCES DIM_SEDE(ID_SEDE)

ON DELETE NO ACTION

ON UPDATE NO ACTION

go

ALTER TABLE DIM_MODALIDAD

ADD CONSTRAINT R_24 FOREIGN KEY
(ID_TIPO_PRUEBA) REFERENCES
DIM_TIPO_PRUEBA(ID_TIPO_PRUEBA)

ON DELETE NO ACTION

ON UPDATE NO ACTION

go

ALTER TABLE DIM_UBIGEO

ADD CONSTRAINT R_20 FOREIGN KEY
(ID_UBIGEO_PADRE) REFERENCES
DIM_UBIGEO(ID_UBIGEO)

ON DELETE NO ACTION

ON UPDATE NO ACTION

go

ALTER TABLE DIM_UBIGEO_COLEGIO

ADD CONSTRAINT R_32 FOREIGN KEY
(ID_UBI_COL_PADRE) REFERENCES
DIM_UBIGEO_COLEGIO(ID_UBIGEO_COLEGIO)

ON DELETE NO ACTION

ON UPDATE NO ACTION

go

ALTER TABLE FACT_ALUMNO

ADD CONSTRAINT R_4 FOREIGN KEY (ID_UBIGEO)
REFERENCES DIM_UBIGEO(ID_UBIGEO)

```

```

ON DELETE NO ACTION

ON UPDATE NO ACTION

go

ALTER TABLE FACT_ALUMNO

ADD CONSTRAINT R_5 FOREIGN KEY (ID_MODALIDAD)
REFERENCES DIM_MODALIDAD(ID_MODALIDAD)

ON DELETE NO ACTION

ON UPDATE NO ACTION

go

ALTER TABLE FACT_ALUMNO

ADD CONSTRAINT R_6 FOREIGN KEY (ID_COLEGIO)
REFERENCES DIM_COLEGIO(ID_COLEGIO)

ON DELETE NO ACTION

ON UPDATE NO ACTION

go

ALTER TABLE FACT_ALUMNO

ADD CONSTRAINT R_7 FOREIGN KEY (ID_CARRERA)
REFERENCES DIM_CARRERA(ID_CARRERA)

ON DELETE NO ACTION

ON UPDATE NO ACTION

go

ALTER TABLE FACT_ALUMNO

ADD CONSTRAINT R_8 FOREIGN KEY (ID_GENERO)
REFERENCES DIM_GENERO(ID_GENERO)

ON DELETE NO ACTION

ON UPDATE NO ACTION

go

ALTER TABLE FACT_ALUMNO

ADD CONSTRAINT R_9 FOREIGN KEY (ID_PROCESO)
REFERENCES DIM_PROCESO(ID_PROCESO)

ON DELETE NO ACTION

ON UPDATE NO ACTION

```

```

go

ALTER TABLE FACT_ALUMNO

ADD CONSTRAINT R_10 FOREIGN KEY
(ID_RANGO_EDAD) REFERENCES
DIM_RANGO_EDAD(ID_RANGO_EDAD)

ON DELETE NO ACTION

ON UPDATE NO ACTION

go

ALTER TABLE FACT_ALUMNO

ADD CONSTRAINT R_11 FOREIGN KEY
(ID_RANGO_MERITO) REFERENCES
DIM_RANGO_MERITO(ID_RANGO_MERITO)

ON DELETE NO ACTION

ON UPDATE NO ACTION

go

ALTER TABLE FACT_ALUMNO

ADD CONSTRAINT R_12 FOREIGN KEY
(ID_RANGO_PUNTAJE) REFERENCES
DIM_RANGO_PUNTAJE(ID_RANGO_PUNTAJE)

ON DELETE NO ACTION

ON UPDATE NO ACTION

go

ALTER TABLE FACT_ALUMNO

ADD CONSTRAINT R_13 FOREIGN KEY
(ID_LETRA_INICIAL_APE) REFERENCES
DIM_LETRA_INICIAL_APE(ID_LETRA_INICIAL_APE)

ON DELETE NO ACTION

ON UPDATE NO ACTION

go

ALTER TABLE FACT_ALUMNO

ADD CONSTRAINT R_14 FOREIGN KEY
(ID_RANGO_EGRESO) REFERENCES
DIM_RANGO_EGRESO(ID_RANGO_EGRESO)

ON DELETE NO ACTION

```

```

ON UPDATE NO ACTION

go

ALTER TABLE FACT_ALUMNO

ADD CONSTRAINT R_15 FOREIGN KEY
(ID_ESTADO_CIVIL) REFERENCES
DIM_ESTADO_CIVIL(ID_ESTADO_CIVIL)

ON DELETE NO ACTION

ON UPDATE NO ACTION

go

ALTER TABLE FACT_ALUMNO

ADD CONSTRAINT R_16 FOREIGN KEY (ID_TRABAJA)
REFERENCES DIM_TRABAJA(ID_TRABAJA)

ON DELETE NO ACTION

ON UPDATE NO ACTION

go

ALTER TABLE FACT_ALUMNO

ADD CONSTRAINT R_29 FOREIGN KEY (ID_MATRICULA)
REFERENCES DIM_MATRICULA(ID_MATRICULA)

ON DELETE NO ACTION

ON UPDATE NO ACTION

go

ALTER TABLE FACT_ALUMNO

ADD CONSTRAINT R_30 FOREIGN KEY (ID_INGRESO)
REFERENCES DIM_INGRESO(ID_INGRESO)

ON DELETE NO ACTION

ON UPDATE NO ACTION

go

CREATE FUNCTION FN_CALCULO_EDAD (@Fecha
datetime)

Returns int

AS

BEGIN

```

```

Declare @Edad smallint

Declare @EDAD_STRING NVARCHAR(8)

SET          @EDAD_STRING          =
CONVERT(NVARCHAR(8),CONVERT(DATE,@Fecha,103),
112)

set @Edad=(datediff(day, @EDAD_STRING,GETDATE())-
datediff(yy, @EDAD_STRING, GETDATE())/4)/365

return(@Edad)

END

go

--OTRAS FUNCIONES

CREATE FUNCTION FN_CALCULAR_RANGO_EGRESO
(@VALOR SMALLINT)

RETURNS SMALLINT

AS

BEGIN

DECLARE @VALOR_1 SMALLINT, @ID SMALLINT,
@ID_RANGO SMALLINT, @MIN_VAL SMALLINT,
@MAX_VAL SMALLINT

SET @VALOR_1 = CAST(YEAR(GETDATE()) AS SMALLINT)
- @VALOR - 1

DECLARE RANGOS CURSOR FOR

SELECT ID_RANGO_EGRESO, MIN_VAL, MAX_VAL FROM
DBO.DIM_RANGO_EGRESO

OPEN RANGOS

FETCH RANGOS INTO @ID_RANGO, @MIN_VAL,
@MAX_VAL

WHILE (@@FETCH_STATUS = 0 )

BEGIN

IF @VALOR_1 >= @MIN_VAL AND @VALOR_1 <
@MAX_VAL

SET @ID = @ID_RANGO

FETCH RANGOS INTO @ID_RANGO, @MIN_VAL,
@MAX_VAL

```

```

END

CLOSE RANGOS

DEALLOCATE RANGOS

RETURN (@ID)

END

go

CREATE FUNCTION FN_CALCULAR_RANGO_EDAD
(@VALOR SMALLINT)

RETURNS SMALLINT

AS

BEGIN

DECLARE @ID SMALLINT, @ID_RANGO SMALLINT,
@MIN_VAL SMALLINT, @MAX_VAL SMALLINT

DECLARE RANGOS CURSOR FOR

SELECT ID_RANGO_EDAD, MIN_VAL, MAX_VAL FROM
DBO.DIM_RANGO_EDAD

OPEN RANGOS

FETCH RANGOS INTO @ID_RANGO, @MIN_VAL,
@MAX_VAL

WHILE (@@FETCH_STATUS = 0 )

BEGIN

IF @VALOR >= @MIN_VAL AND @VALOR < @MAX_VAL

SET @ID = @ID_RANGO

FETCH RANGOS INTO @ID_RANGO, @MIN_VAL,
@MAX_VAL

END

CLOSE RANGOS

DEALLOCATE RANGOS

RETURN (@ID)

END

go

```

```

CREATE FUNCTION FN_CALCULAR_RANGO_MERITO
(@VALOR SMALLINT)

RETURNS SMALLINT

AS

BEGIN

DECLARE @ID SMALLINT, @ID_RANGO SMALLINT,
@MIN_VAL SMALLINT, @MAX_VAL SMALLINT

DECLARE RANGOS CURSOR FOR

SELECT ID_RANGO_MERITO, MIN_VAL, MAX_VAL FROM
DBO.DIM_RANGO_MERITO

OPEN RANGOS

FETCH RANGOS INTO @ID_RANGO, @MIN_VAL,
@MAX_VAL

WHILE (@@FETCH_STATUS = 0 )

BEGIN

IF @VALOR >= @MIN_VAL AND @VALOR < @MAX_VAL

SET @ID = @ID_RANGO

FETCH RANGOS INTO @ID_RANGO, @MIN_VAL,
@MAX_VAL

END

CLOSE RANGOS

DEALLOCATE RANGOS

RETURN (@ID)

END

go

CREATE FUNCTION FN_CALCULAR_RANGO_PUNTAJE
(@VALOR SMALLINT)

RETURNS SMALLINT

AS

BEGIN

DECLARE @ID SMALLINT, @ID_RANGO SMALLINT,
@MIN_VAL SMALLINT, @MAX_VAL SMALLINT

```

```

DECLARE RANGOS CURSOR FOR

SELECT ID_RANGO_PUNTAJE, MIN_VAL, MAX_VAL
FROM DBO.DIM_RANGO_PUNTAJE

OPEN RANGOS

FETCH RANGOS INTO @ID_RANGO, @MIN_VAL,
@MAX_VAL

WHILE (@@FETCH_STATUS = 0 )

BEGIN

IF @VALOR >= @MIN_VAL AND @VALOR < @MAX_VAL

SET @ID = @ID_RANGO

FETCH RANGOS INTO @ID_RANGO, @MIN_VAL,
@MAX_VAL

END

CLOSE RANGOS

DEALLOCATE RANGOS

RETURN (@ID)

END

go

```

ANEXO 8
MAPEO DE FUENTE COLEGIOS.XLS

NOMBRE DE FUENTE	COLEGIOS.XLS
DESCRIPCION	Contiene la información de todos los colegios registrados a la fecha por la universidad.
PERIODICIDAD DE CARGA	Semestral
PAQUETE DE CARGA	ETL_DIMENSIONES_SCD
PRECEDENCIAS	UBIGEOS_R.XLS
DESTINOS	DIM_COLEGIO, DIM_TIPO_COLEGIO

a. Mapeo DIM_COLEGIO

DESTINO						ORIGEN			LIMPIEZA DE DATOS	REGLAS ETL	SEVERIDADES
CAMPO	DESCRIPCIÓN	TIPO DATO	TAMAÑO	INDICE	NULL	CAMPO	DESCRIPCIÓN	TIPO DATO			
ID_COLEGIO	Identificador Único de Colegio	SMALLINT		PK	NO					Se autogenera un código	
COLEGIO	Nombre del Centro Educativo	NVARCHAR	100		NO	NOMBRE	Nombre del Centro Educativo	Texto	Si es Null o vacío, llenar con punto. Sino Trim y Uppercase	Lineal	Si es null o vacío, severidad tipo 2. Si se carga información, severidad tipo 3 de cuenta de registros insertados.
ID_TIPO_COLEGIO	Identificador de Tipo de Colegio	SMALLINT		FK	NO	GESTION	Tipo de Colegio	Texto	Si es Null o vacío, pase lineal. Sino Trim y Uppercase	Se obtiene el ID de la tabla DIM_TIPO_COLEGIO buscando por GESTION	
ID_UBIGEO_COLEGIO	Identificador de ubigeo del colegio	SMALLINT		FK	NO	DEPARTAMENTO	Departamento de Ubigeo	Texto	Si es Null o vacío, pase lineal. Sino Trim y Uppercase	Se obtiene el ID de la tabla DIM_UBIGEO_COLEGIO buscando por DEPARTAMENTO, PROVINCIA y DISTRITO.	Si no se encuentra el ID, severidad tipo 1.
						PROVINCIA	Provincia de Ubigeo del Colegio	Texto	Si es Null o vacío, pase lineal. Sino Trim y Uppercase		
						DISTRITO	Distrito de Ubigeo del colegio	Texto	Si es Null o vacío, pase lineal. Sino Trim y Uppercase		
DIRECCION_COLEGIO	Dirección del Colegio	NVARCHAR	120		SI	DIRECCION	Dirección del Colegio	Texto	Si es Null o vacío, llenar con "REFERENCIAL". Sino Trim y Uppercase	Lineal	
FEC_MOD	Ultima Fecha de modificación de la tabla	SMALLDATETIME			NO					Se obtiene la fecha del sistema	

b. Mapeo DIM_TIPO_COLEGIO

DESTINO						ORIGEN			LIMPIEZA DE DATOS	REGLAS ETL	SEVERIDADES
CAMPO	DESCRIPCIÓN	TIPO DATO	TAMAÑO	INDICE	NULL	CAMPO	DESCRIPCIÓN	TIPO DATO			
ID_TIPO_COLEGIO	Indetificador Unico de Tipo Colegio	SMALLINT		PK	NO					Se autogenera un código	
TIPO_COLEGIO	Tipo de Colegio de procedencia	NVARCHAR	20		NO	GESTION	Tipo de Colegio	Texto	Si es Null o vacio, llenar con punto. Sino Trim y Uppercase	Lineal	Si es null o vacio, severidad tipo 2. Si se carga informacion, severidad tipo 3 de cuenta de registros insertados.
FEC_MOD	Ultima Fecha de modificación de la tabla	SMALLDATETIME			NO					Se obtiene la fecha del sistema	

**ANEXO 9
MAPEO DE FUENTE UBIGEOS.XLS**

NOMBRE DE FUENTE	UBIGEOS_R.XLS
DESCRIPCION	Contiene la información de todos departamentos, provincias y distritos del Perú. Además, se presenta información de continentes, países y ciudades en el caso de que sean ubigeos externos al Perú
PERIODICIDAD DE CARGA	Semestral
PAQUETE DE CARGA	ETL_DIMENSIONES_SCD
PRECEDENCIAS	Ninguna
DESTINOS	DIM_UBIGEO, DIM_UBIGEO_COLEGIO

a. Mapeo DIM_UBIGEO

DESTINO						ORIGEN			LIMPIEZA DE DATOS	REGLAS ETL	SEVERIDADES
CAMPO	DESCRIPCIÓN	TIPO DATO	TAMAÑO	INDICE	NULL	CAMPO	DESCRIPCIÓN	TIPO DATO			
ID_UBIGEO	Identificador único de ubicación geográfica	SMALLINT		PK	NO					Se autogenera un código	
UBIGEO	Ubicación geográfica	NVARCHAR	50		NO	NODEP	Nombre del Departamento	Texto	Si es Null o vacío, pase lineal. Sino Trim y Uppercase	Lineal	Si se cumple la condición de la Tabla 1 , entonces severidad tipo 1. Si se carga información, severidad tipo 3 de cuenta de registros insertados.
						NOPRV	Nombre de la Provincia	Texto			
						NODIS	Nombre del Distrito	Texto			
ID_UBIGEO_PADRE	Identificador único de la ubicación geográfica Padre	SMALLINT		FK	SI					Se le asigna un código que hace referencia a la misma Dimension (Referencia Circular)	
FEC_MOD	Ultima Fecha de modificación de la tabla	SMALLDATETIME			NO					Se obtiene la fecha del sistema	

NOTA:	No se detalla el mapeo de la dimensión DIM_UBIGEO_COLEGIO debido a que dicha dimensión es una copia exacta de la dimensión DIM_UBIGEO. Esto se debe a que el motor OLAP no permite crear referencias cerradas y es necesario utilizar el ubigeo tanto para la FACT_ALUMNOS como para la DIM_COLEGIO
--------------	---

TABLA 1		
NODEP	NOPRV	NODIS
Null o Vacio	Null o Vacio	Null o Vacio
Null o Vacio	Valor OK	Valor OK
Null o Vacio	Null o Vacio	Valor OK
Null o Vacio	Valor OK	Null o Vacio
Valor OK	Null o Vacio	Valor OK

**ANEXO 10
MAPEO DE FUENTE MODALIDAD.XLS**

NOMBRE DE FUENTE	MODALIDAD.XLS
DESCRIPCION	Contiene la información de todas las modalidades de ingreso a la Universidad, con su respectivo tipo de prueba.
PERIODICIDAD DE CARGA	Semestral
PAQUETE DE CARGA	ETL_DIMENSIONES_SCD
PRECEDENCIAS	Ninguna
DESTINOS	DIM_MODALIDAD, DIM_TIPO_PRUEBA

a. Mapeo DIM_MODALIDAD

DESTINO						ORIGEN			LIMPIEZA DE DATOS	REGLAS ETL	SEVERIDADES
CAMPO	DESCRIPCIÓN	TIPO DATO	TAMAÑO	INDICE	NULL	CAMPO	DESCRIPCIÓN	TIPO DATO			
ID_MODALIDAD	Identificador Único de Modalidad	SMALLINT		PK	NO					Se autogenera un código	
MODALIDAD	Nombre de la Modalidad	NVARCHAR	50		NO	MODALIDAD	Tipo de Modalidad	Texto	Si es Null o vacío, llenar con punto. Sino Trim y Uppercase	Lineal	Si es null o vacío, severidad tipo 2. Si se carga información, severidad tipo 3 de cuenta de registros insertados
ID_TIPO_PRUEBA	Identificador Único del tipo de prueba	SMALLINT		FK	NO	TIPO PRUEBA	Tipo de Examen de la modalidad	Texto	Si es Null o vacío, pase lineal. Sino Trim y Uppercase	Se obtiene el ID de la tabla DIM_TIPO_PRUEBA buscando por TIPO_PRUEBA	Si no se encuentra el ID, severidad tipo 1.
FEC_MOD	Ultima Fecha de modificación de la tabla	SMALLDATETIME			NO					Se obtiene la fecha del sistema	

b. Mapeo DIM_TIPO_PRUEBA

DESTINO						ORIGEN			LIMPIEZA DE DATOS	REGLAS ETL	SEVERIDADES
CAMPO	DESCRIPCIÓN	TIPO DATO	TAMAÑO	INDICE	NULL	CAMPO	DESCRIPCIÓN	TIPO DATO			
ID_TIPO_PRUEBA	Identificador Único del tipo de prueba	SMALLINT		PK	NO					Se autogenera un código	
TIPO_PRUEBA	Tipo de Examen de ingreso	NVARCHAR	50		NO	TIPO PRUEBA	Tipo de examen de la modalidad	Texto	Si es Null o vacío, llenar con punto. Sino Trim y Uppercase	Lineal	Si es null o vacío, severidad tipo 2. Si se carga información, severidad tipo 3 de cuenta de registros insertados
DETALLE_PRUEBA	Detalle del tipo de examen de ingreso	NVARCHAR	100		SI	DETALLE	Detalle del examen	Texto	Si es Null o vacío, llenar con punto. Sino Trim y Uppercase	Lineal	
FEC_MOD	Ultima Fecha de modificación de la tabla	SMALLDATETIME			NO					Se obtiene la fecha del sistema	

ANEXO 11
MAPEO DE FUENTE ESPECIALIDAD.XLS

NOMBRE DE FUENTE	ESPECIALIDAD.XLS
DESCRIPCION	Presenta información de las diversas carreras que pertenecen a una cierta facultad y estas a su vez a una sede en particular.
PERIODICIDAD DE CARGA	Semestral
PAQUETE DE CARGA	ETL_DIMENSIONES_SCD
PRECEDENCIAS	Ninguna
DESTINOS	DIM_CARRERA, DIM_FACULTAD, DIM_SEDE

a. Mapeo DIM_CARRERA

DESTINO						ORIGEN			LIMPIEZA DE DATOS	REGLAS ETL	SEVERIDADES
CAMPO	DESCRIPCIÓN	TIPO DATO	TAMAÑO	INDICE	NULL	CAMPO	DESCRIPCIÓN	TIPO DATO			
ID_CARRERA	Identificador Único de la carrera profesional	SMALLINT		PK	NO					Se autogenera un código	
CARRERA	Nombre de la carrera profesional	NVARCHAR	100		NO	CARRERA	Nombre de la carrera profesional	Texto	Si es Null o vacío, llenar con punto. Sino Trim y Uppercase	Lineal	Si es null o vacío, severidad tipo 2. Si se carga información, severidad tipo 3 de cuenta de registros insertados.
CARRERA_ABREV	Abreviatura de la Carrera	NVARCHAR	50		NO	CARRERA	Nombre de la carrera profesional	Texto	Si es Null o vacío, llenar con punto. Sino Trim y Uppercase	Se trunca el campo CARRERA a máximo 50 caracteres	
ID_FACULTAD	Identificador Único de la facultad	SMALLINT		FK	NO	FACULTAD	Nombre de la facultad	Texto	Si es Null o vacío, pase lineal. Sino Trim y Uppercase	Se obtiene el ID de la tabla DIM_FACULTAD buscando por FACULTAD	Si no se encuentra el ID, severidad tipo 1.
FEC_MOD	Ultima Fecha de modificación de la tabla	SMALLDATETIME			NO					Se obtiene la fecha del sistema	

b. Mapeo DIM_FACULTAD

DESTINO						ORIGEN			LIMPIEZA DE DATOS	REGLAS ETL	SEVERIDADES
CAMPO	DESCRIPCIÓN	TIPO DATO	TAMAÑO	INDICE	NULL	CAMPO	DESCRIPCIÓN	TIPO DATO			
ID_FACULTAD	Identificador Único de la facultad	SMALLINT		PK	NO					Se autogenera un código	
FACULTAD	Nombre de la facultad	NVARCHAR	100		NO	FACULTAD	Nombre de la facultad	Texto	Si es Null o vacío, llenar con punto. Sino Trim y Uppercase	Lineal	Si es null o vacío, severidad tipo 2 , Si se carga información, severidad tipo 3 de cuenta de registros insertados.
FACULTAD_ABREV	Abreviatura de la facultad	NVARCHAR	50		NO	FACULTAD	Nombre de la facultad	Texto	Si es Null o vacío, llenar con punto. Sino Trim y Uppercase	Se trunca el campo FACULTAD a máximo 50 caracteres	
ID_SEDE	Identificador Único de la sede	SMALLINT		FK	NO	SEDE	Nombre de la Sede Principal o Filial	Texto	Si es Null o vacío, pase lineal. Sino Trim y Uppercase	Se obtiene el ID de la tabla DIM_SEDE buscando por SEDE	Si no se encuentra el ID, severidad tipo 1.
FEC_MOD	Ultima Fecha de modificación de la tabla	SMALLDATETIME			NO					Se obtiene la fecha del sistema	

c. Mapeo DIM_SEDE

DESTINO						ORIGEN			LIMPIEZA DE DATOS	REGLAS ETL	SEVERIDADES
CAMPO	DESCRIPCIÓN	TIPO DATO	TAMAÑO	INDICE	NULL	CAMPO	DESCRIPCIÓN	TIPO DATO			
ID_SEDE	Identificador Único de la sede	SMALLINT		PK	NO					Se autogenera un código	
SEDE	Nombre de la Sede de la Universidad	NVARCHAR	30		NO	SEDE	Nombre de la Sede Principal o Filial	Texto	Si es Null o vacío, llenar con punto. Sino Trim y Uppercase	Lineal	Si es null o vacío, severidad tipo 2. Si se carga información, severidad tipo 3 de cuenta de registros insertados.
FEC_MOD	Ultima Fecha de modificación de la tabla	SMALLDATETIME			NO					Se obtiene la fecha del sistema	

ANEXO 12
MAPEO DE FUENTE DATOS_MATRICULA.XLS

NOMBRE DE FUENTE	DATOS_MATRICULA.XLS
DESCRIPCION	Contiene información de todos los ingresantes que se han matriculado o no en un determinado periodo académico
PERIODICIDAD DE CARGA	Semestral
PAQUETE DE CARGA	ETL_FACT_ALUMNO
PRECEDENCIAS	COLEGIOS.XLS, UBIGEOS_R.XLS, MODALIDAD.XLS, ESPECIALIDAD.XLS
DESTINOS	FACT_ALUMNO

a. Mapeo FACT_ALUMNO

DESTINO						ORIGEN			LIMPIEZA DE DATOS	REGLAS ETL	SEVERIDADES
CAMPO	DESCRIPCIÓN	TIPO DATO	TAMAÑO	INDICE	NULL	CAMPO	DESCRIPCIÓN	TIPO DATO			
FEC_MATRICULA	Fecha de Matricula	SMALLDATETIME			SI	FECMATRICULA	Fecha de Matricula	Fecha		Lineal	
ID_MATRICULA	Identificador único del estado de la matricula	SMALLINT		FK	NO	MATRICULA	Situación actual de Matricula	Texto	Si es Null o vacío, pase lineal. Sino Trim y Uppercase	Se obtiene el ID de la tabla DIM_MATRICULA buscando por MATRICULA	Si no se encuentra el ID, severidad tipo 1.

ANEXO 13
MAPEO DE FUENTE DATOS_POSTULANTE.XLS

NOMBRE DE FUENTE	DATOS_POSTULANTE.XLS
DESCRIPCION	Presenta los datos de todos los postulantes a las diversas carreras de la Facultad de Ingeniería y Arquitectura.
PERIODICIDAD DE CARGA	Semestral
PAQUETE DE CARGA	ETL_FACT_ALUMNO
PRECEDENCIAS	COLEGIOS.XLS, UBIGEOS_R.XLS, MODALIDAD.XLS, ESPECIALIDAD.XLS
DESTINOS	FACT_ALUMNO, DIM_PROCESO

a. Mapeo FACT_ALUMNO

DESTINO						ORIGEN			LIMPIEZA DE DATOS	REGLAS ETL	SEVERIDADES
CAMPO	DESCRIPCIÓN	TIPO DATO	TAMAÑO	INDICE	NULL	CAMPO	DESCRIPCIÓN	TIPO DATO			
ID_ALUMNO	Identificador único del alumno	NVARCHAR	10	PK	NO	CODALUMNO	Código del postulante	Texto		Se autogenera un código	
NOMBRE	Nombres y del Apellidos Alumno	NVARCHAR	100		NO	PATERNO	Apellido Paterno	Texto	Si es Null, llenar con vacío. Sino Trim y Uppercase	Concatenar campos PATERNO + MATERNO + NOMBRES.	Si alguno de los campos es null o vacío, severidad tipo 2.
						MATERNO	Apellido materno	Texto			
						NOMBRES	Nombres	Texto			
FEC_NACIMIENTO	Fecha de Nacimiento del Alumno	SMALLDATETIME			NO	FNACIMIEN	Fecha de Nacimiento	Fecha		Lineal	Si el campo null o vacío, severidad tipo 1.
EDAD	Campo calculado de la edad del alumno	SMALLINT			NO	FNACIMIEN	Fecha de Nacimiento	Fecha			
TELEFONO_1	Teléfono contacto 1 de	NVARCHAR	15		SI	TELEFONO1	1er Teléfono ingresante o contacto	Texto		Lineal	
TELEFONO_2	Teléfono contacto 2 de	NVARCHAR	15		SI	TELEFONO2	2do Teléfono ingresante o contacto	Texto			

DESTINO						ORIGEN			LIMPIEZA DE DATOS	REGLAS ETL	SEVERIDADES	
CAMPO	DESCRIPCIÓN	TIPO DATO	TAMAÑO	INDICE	NULL	CAMPO	DESCRIPCIÓN	TIPO DATO				
PUNTAJE	Puntaje del Ingreso	SMALLINT			NO	PUNTAJE	Puntaje de ingreso	Numérico	Si es Null o vacío, llenar con punto. Sino Trim y Uppercase	Lineal	Si alguno de los campos es null o vacío, severidad tipo 1.	
MERITO	Orden de mérito de ingreso	SMALLINT			NO	MERITO	Orden de mérito de ingreso	Numérico				
DIRECCION	Dirección de residencia	NVARCHAR	150		NO	DIRECCION	Dirección del postulante	Texto			Si el campo es null o vacío, severidad tipo 2.	
EMAIL	Email de contacto del alumno	NVARCHAR	60		SI	EMAIL	Correo electrónico	Texto				
TIPO_DOCUMENTO	Tipo de Documento	NVARCHAR	40		NO	TIPDOCUM	Tipo documento de identificación	Texto				
NUM_DOCUMENTO	Número del documento	NVARCHAR	15		NO	NUMERODOC	Nro. Documento de identificación	Texto	Si alguno de los campos es null o vacío, severidad tipo 2.			
FEC_MATRICULA	Fecha de Matricula	SMALLDATETIME			SI	Se carga desde la fuente DATOS MATRICULA.XLS						
ID_UBIGEO	Identificador único de la ubicación geográfica	SMALLINT		FK	NO	UBIRESIDENCIA	Dirección de residencia del ingresante	Texto	Si es Null o vacío, pase lineal. Sino Trim y Uppercase	Se obtiene el ID buscando por UBIRESIDENCIA y comparándolo con la concatenación DEPARTAMENTO - PROVINCIA - DISTRITO de la tabla DIM_UBIGEO	Si no se encuentra el ID, severidad tipo 1.	
ID_MODALIDAD	Identificador único de la modalidad	SMALLINT		FK	NO	MODALIDAD	Modalidad de admisión	Texto				Se obtiene el ID de la tabla DIM_MODALIDAD buscando por MODALIDAD
ID_COLEGIO	Identificador único del colegio	SMALLINT		FK	NO	NOMCOLEGI	Nombre del colegio de procedencia	Texto				Se obtiene el ID de la tabla DIM_TIPO_COLEGIO buscando por TIPCOLEGI. Se obtiene el ID buscando por UBICOLEGI y comparándolo con la concatenación DEPARTAMENTO - PROVINCIA - DISTRITO de la tabla DIM_UBIGEO_COLEGIO. Se obtiene el ID de la tabla DIM_COLEGIO buscando por NOMCOLEGI, ID_TIPO_COLEGIO y ID_UBIGEO_COLEGIO
						UBICOLEGI	Ubicación geográfica del Colegio	Texto				
						TIPCOLEGI	Tipo de Colegio	Texto				

DESTINO						ORIGEN			LIMPIEZA DE DATOS	REGLAS ETL	SEVERIDADES
CAMPO	DESCRIPCIÓN	TIPO DATO	TAMAÑO	INDICE	NULL	CAMPO	DESCRIPCIÓN	TIPO DATO			
ID_CARRERA	Identificador único de la carrera	SMALLINT		FK	NO	ESPECIALIDAD	Nombre de la carrera de ingreso	Texto	Si es Null o vacío, pase lineal. Sino Trim y Uppercase	Se obtiene el ID de la tabla DIM_CARRERA buscando por ESPECIALIDAD	Si no se encuentra el ID, severidad tipo 1.
ID_GENERO	Identificador único del género	SMALLINT		FK	NO	SEXO	Sexo del ingresante	Texto		Se obtiene el ID de la tabla DIM_GENERO buscando por SEXO	
ID_PROCESO	Identificador único del proceso de admisión	SMALLINT		FK	NO	CODALUMNO	Código del postulante	Texto		Se obtiene el ID de la tabla DIM_PROCESO buscando por los 5 primeros dígitos del campo CODALUMNO (Año y semestre)	
ID_RANGO_EDAD	Identificador único del rango de edad	SMALLINT		FK	NO					Se obtiene el ID de la tabla DIM_RANGO_EDAD buscando por el campo EDAD en el RANGO_EDAD	
ID_RANGO_MERITO	Identificador único del rango de mérito	SMALLINT		FK	NO	MERITO	Orden de mérito de ingreso	Numérico		Se obtiene el ID de la tabla DIM_RANGO_MERITO buscando por el campo MERITO en el RANGO_MERITO	
ID_RANGO_PUNTAJE	Identificador único del rango del puntaje de ingreso	SMALLINT		FK	NO	PUNTAJE	Puntaje de ingreso	Numérico		Se obtiene el ID de la tabla DIM_RANGO_PUNTAJE buscando por el campo PUNTAJE en el RANGO_PUNTAJE	
ID_LETRA_INICIAL_APE	Identificador único de la letra inicial del apellido	SMALLINT		FK	NO	PATERNO	Apellido Paterno	Texto	Si es Null, llenar con vacío. Sino Trim y Uppercase	Se obtiene el ID de la tabla DIM_LETRA_INICIAL_APE buscando por la primera letra de campo PATERNO	
ID_RANGO_EGRESO	Identificador único del rango de egreso del colegio	SMALLINT		FK	NO	EGRESOCOL	Año de Egreso del Colegio	Numérico		Se obtiene el ID de la tabla DIM_RANGO_EGRESO buscando por el campo EGRESOCOL en el RANGO_EGRESO	
ID_ESTADO_CIVIL	Identificador único del estado civil	SMALLINT		FK	NO	ESTADOCIVI	Estado civil del ingresante	Texto	Si es Null o vacío, pase lineal. Sino Trim y Uppercase	Se obtiene el ID de la tabla DIM_ESTADO_CIVIL buscando por el campo ESTADOCIVI	
ID_TRABAJA	Identificador único de la situación laboral actual	SMALLINT		FK	NO	TRABAJA	Situación laboral del ingresante	Texto		Se obtiene el ID de la tabla DIM_TRABAJA buscando por el campo TRABAJA	
ID_MATRICULA	Identificador único del estado de la matrícula	SMALLINT		FK	NO	Se carga desde la fuente DATOS MATRICULA.XLS					
ID_INGRESO	Identificador único del estado de ingreso del postulante	SMALLINT		FK	NO	INGRESO	Situación actual de la postulación	Texto	Si es Null o vacío, pase lineal. Sino Trim y Uppercase	Se obtiene el ID de la tabla DIM_INGRESO buscando por el campo INGRESO	Si no se encuentra el ID, severidad tipo 1.
FEC_MOD	Ultima Fecha de modificación de la tabla	SMALLDATETIME			NO					Se obtiene la fecha del sistema	

ANEXO 14

MAPEO DE FUENTE DATOS_POSTULANTE.XLS – ANÁLISIS

NOMBRE DE FUENTE	DATOS_POSTULANTE.XLS - ANÁLISIS
DESCRIPCION	Información sobre otras dimensiones que es poco probable que cambien en el tiempo.
PERIODICIDAD DE CARGA	Una sola vez
PAQUETE DE CARGA	ETL_DIMENSIONES_NCD
PRECEDENCIAS	Ninguna
DESTINOS	DIM_ESTADO_CIVIL, DIM_GENERO, DIM_TRABAJA, DIM_INGRESO, DIM_MATRICULA, DIM_LETRA_INICIAL_APE, DIM_RANGO_MERITO, DIM_RANGO_EDAD, DIM_RANGO_PUNTAJE, DIM_RANGO_EGRESO

a. Mapeo DIM_ESTADO_CIVIL

DESTINO						ORIGEN			LIMPIEZA DE DATOS	REGLAS ETL	SEVERIDADES
CAMPO	DESCRIPCIÓN	TIPO DATO	TAMAÑO	INDICE	NULL	CAMPO	DESCRIPCIÓN	TIPO DATO			
ID_ESTADO_CIVIL	Identificador único del estado civil	SMALLINT		PK	NO				Se autogenera un código		
ESTADO_CIVIL	Descripción del Estado Civil	NVARCHAR	15		NO				Se ingresan los valores predeterminados 'SOLTERO', 'CASADO', 'VIUDO' Y 'DIVORCIADO'		
FEC_MOD	Ultima Fecha de modificación de la tabla	SMALLDATETIME			NO				Se obtiene la fecha del sistema		

b. Mapeo DIM_GENERO

DESTINO						ORIGEN			LIMPIEZA DE DATOS	REGLAS ETL	SEVERIDADES
CAMPO	DESCRIPCIÓN	TIPO DATO	TAMAÑO	INDICE	NULL	CAMPO	DESCRIPCIÓN	TIPO DATO			
ID_GENERO	Identificador único del género	SMALLINT		PK	NO				Se autogenera un código		
GENERO	Descripción del género	NVARCHAR	15		NO				Se ingresan los valores predeterminados 'MASCULINO' y 'FEMENINO'		
FEC_MOD	Ultima Fecha de modificación de la tabla	SMALLDATETIME			NO				Se obtiene la fecha del sistema		

c. Mapeo DIM_TRABAJA

DESTINO						ORIGEN			LIMPIEZA DE DATOS	REGLAS ETL	SEVERIDADES
CAMPO	DESCRIPCIÓN	TIPO DATO	TAMAÑO	INDICE	NULL	CAMPO	DESCRIPCIÓN	TIPO DATO			
ID_TRABAJA	Identificador único de la situación laboral actual	SMALLINT		PK	NO					Se autogenera un código	
TRABAJA	Descripción del género	CHAR	2		NO					Se ingresan los valores predeterminados 'SI y 'NO'	
FEC_MOD	Ultima Fecha de modificación de la tabla	SMALLDATETIME			NO					Se obtiene la fecha del sistema	

d. Mapeo DIM_INGRESO

DESTINO						ORIGEN			LIMPIEZA DE DATOS	REGLAS ETL	SEVERIDADES
CAMPO	DESCRIPCIÓN	TIPO DATO	TAMAÑO	INDICE	NULL	CAMPO	DESCRIPCIÓN	TIPO DATO			
ID_INGRESO	Identificador único de la identifica la situación laboral actual	SMALLINT		PK	NO					Se autogenera un código	
INGRESO	Estado Ingreso del Postulante	CHAR	2		NO					Se ingresan los valores predeterminados 'SI y 'NO'	
FEC_MOD	Ultima Fecha de modificación de la tabla	SMALLDATETIME			NO					Se obtiene la fecha del sistema	

e. Mapeo DIM_MATRICULA

DESTINO						ORIGEN			LIMPIEZA DE DATOS	REGLAS ETL	SEVERIDADES
CAMPO	DESCRIPCIÓN	TIPO DATO	TAMAÑO	INDICE	NULL	CAMPO	DESCRIPCIÓN	TIPO DATO			
ID_MATRICULA	Identificador único del estado de la matrícula	SMALLINT		PK	NO					Se autogenera un código	
MATRICULA	Estado de la matrícula del ingresante	CHAR	2		NO					Se ingresan los valores predeterminados 'SI y 'NO'	
FEC_MOD	Ultima Fecha de modificación de la tabla	SMALLDATETIME			NO					Se obtiene la fecha del sistema	

f. Mapeo DIM_LETRA_INICIAL_APE

DESTINO						ORIGEN			LIMPIEZA DE DATOS	REGLAS ETL	SEVERIDADES
CAMPO	DESCRIPCIÓN	TIPO DATO	TAMAÑO	INDICE	NULL	CAMPO	DESCRIPCIÓN	TIPO DATO			
COD_LETRA_INICIAL_APE	Identificador único de la letra del abecedario	SMALLINT		PK	NO					Se autogenera un código	
LETRA_INICIAL_APE	Letra del Abecedario	CHAR	1		NO					Se ingresan todas las letras del abecedario.	
FEC_MOD	Ultima Fecha de modificación de la tabla	SMALLDATETIME			NO					Se obtiene la fecha del sistema	

g. Mapeo DIM_RANGO_MERITO

DESTINO						ORIGEN			LIMPIEZA DE DATOS	REGLAS ETL	SEVERIDADES
CAMPO	DESCRIPCIÓN	TIPO DATO	TAMAÑO	INDICE	NULL	CAMPO	DESCRIPCIÓN	TIPO DATO			
ID_RANGO_MERITO	Identificador único del rango de merito	SMALLINT		PK	NO					Se autogenera un código	
RANGO_MERITO	Descripción del rango de merito	NVARCHAR	50		NO					Se ingresan la descripción de rangos predeterminados por el usuario.	
MIN_VAL	Mínimo valor del rango	SMALLINT			NO					Se ingresa el valor mínimo del rango	
MAX_VAL	Máximo valor del rango	SMALLINT			NO					Se ingresa el valor máximo del rango	
FEC_MOD	Ultima Fecha de modificación de la tabla	SMALLDATETIME			NO					Se obtiene la fecha del sistema	

h. Mapeo DIM_RANGO_EDAD

DESTINO						ORIGEN			LIMPIEZA DE DATOS	REGLAS ETL	SEVERIDADES
CAMPO	DESCRIPCIÓN	TIPO DATO	TAMAÑO	INDICE	NULL	CAMPO	DESCRIPCIÓN	TIPO DATO			
ID_RANGO_EDAD	Identificador único Del rango de edad	SMALLINT		PK	NO					Se autogenera un código	
RANGO_EDAD	Descripción del rango de la edad	NVARCHAR	50		NO					Se ingresan la descripción de rangos predeterminados por el usuario.	
MIN_VAL	Mínimo valor del rango	SMALLINT			NO					Se ingresa el valor mínimo del rango	
MAX_VAL	Máximo valor del rango	SMALLINT			NO					Se ingresa el valor máximo del rango	
FEC_MOD	Ultima Fecha de modificación de la tabla	SMALLDATETIME			NO					Se obtiene la fecha del sistema	

i. Mapeo DIM_RANGO_PUNTAJE

DESTINO						ORIGEN			LIMPIEZA DE DATOS	REGLAS ETL	SEVERIDADES
CAMPO	DESCRIPCIÓN	TIPO DATO	TAMAÑO	INDICE	NULL	CAMPO	DESCRIPCIÓN	TIPO DATO			
ID_RANGO_PUNTAJE	Identificador único del rango de puntaje	SMALLINT		PK	NO					Se autogenera un código	
RANGO_PUNTAJE	Descripción del rango del puntaje	NVARCHAR	50		NO					Se ingresan la descripción de rangos predeterminados por el usuario.	
MIN_VAL	Mínimo valor del rango	SMALLINT			NO					Se ingresa el valor mínimo del rango	
MAX_VAL	Máximo valor del rango	SMALLINT			NO					Se ingresa el valor máximo del rango	
FEC_MOD	Ultima Fecha de modificación de la tabla	SMALLDATETIME			NO					Se obtiene la fecha del sistema	

j. Mapeo DIM_RANGO_EGRESO

DESTINO						ORIGEN			LIMPIEZA DE DATOS	REGLAS ETL	SEVERIDADES
CAMPO	DESCRIPCIÓN	TIPO DATO	TAMAÑO	INDICE	NULL	CAMPO	DESCRIPCIÓN	TIPO DATO			
ID_RANGO_EGRESO	Identificador único del rango de egreso	SMALLINT		PK	NO					Se autogenera un código	
RANGO_EGRESO	Descripción del rango del egreso	NVARCHAR	50		NO					Se ingresan la descripción de rangos predeterminados por el usuario.	
MIN_VAL	Mínimo valor del rango	SMALLINT			NO					Se ingresa el valor mínimo del rango	
MAX_VAL	Máximo valor del rango	SMALLINT			NO					Se ingresa el valor máximo del rango	
FEC_MOD	Ultima Fecha de modificación de la tabla	SMALLDATETIME			NO					Se obtiene la fecha del sistema	

ANEXO 15
DISEÑO DE FLUJO DE DATOS

ÍNDICE

	Página
1. Paquete ETL_DIMENSIONES_NCD	248
2. Paquete ETL_DIMENSIONES_SCD	248
2.1 ETL_DIM_ESPECIALIDAD	249
2.2 ELT_DIM_UBIGEO	251
2.3 ETL_DIM_UBIGEO_COLEGIO	254
2.4 ETL_DIM_COLEGIO	255
2.5 ETL_DIM_MODALIDAD	256
2.6 CARGA DE SEVERIDADES – Dimensiones SCD	258
3. Paquete ETL_FACT_ALUMNO	259
3.1 EX_FUENTES_FACT_ALUMNO	260
3.2 RN_CW_DIM_PROCESO	261
3.3 RN_FACT_ALUMNO_GRP1	263
3.4 RN_FACT_ALUMNO_GRP2	265
3.5 CW_FACT_ALUMNO	266
3.6 CARGA DE SEVERIDADES – FACT_ALUMNOS	267
3.7 LIMPIEZA DE TEMPORALES	268

ÍNDICE DE ILUSTRACIONES

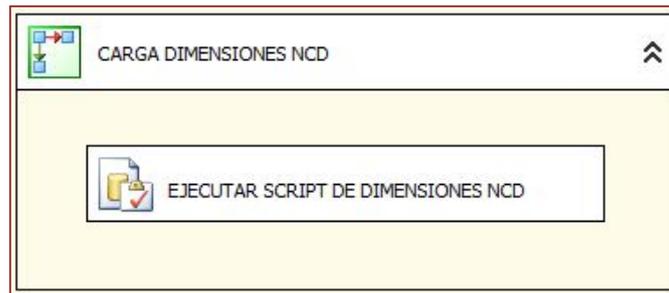
	Página
Ilustración I - ETL_DIMENSIONES_NCD	248
Ilustración II - ELT_DIMENSIONES_SCD	248
Ilustración III - ETL_DIM_ESPECIALIDAD	249
Ilustración IV - EX_SRC_POSTULANTES	249
Ilustración V - EX_DIM_SEDE	250
Ilustración VI - RN_CW_DIM_FACULTAD	250
Ilustración VII - RN_CW_DIM_CARRERA	251
Ilustración VIII - ELT_DIM_UBIGEO	251
Ilustración IX - EX_SRC_UBIGEO	252
Ilustración X - RN_CW_DIM_UBIGEO_DPTO	252
Ilustración XI - RN_CW_DIM_UBIGEO_PROV	253
Ilustración XII - RN_CW_DIM_UBIGEO_DISTRITO	253
Ilustración XIII - ETL_DIM_UBIGEO_COLEGIO	254
Ilustración XIV - ETL_DIM_UBIGEO_COLEGIO	254
Ilustración XV - ETL_DIM_COLEGIO	255
Ilustración XVI - EX_SRC_COLEGIOS	255
Ilustración XVII - RN_CW_DIM_COLEGIO	256
Ilustración XVIII - ETL_DIM_MODALIDAD	256
Ilustración XIX - EX_SRC_MODALIDAD	257
Ilustración XX - RN_CW_DIM_TIPO_PRUEBA	257
Ilustración XXI - RN_CW_DIM_MODALIDAD	258
Ilustración XXII - CARGA DE SEVERIDADES	258
Ilustración XXIII - CW_SEVERIDADES	259
Ilustración XXIV - ETL_FACT_ALUMNO	259
Ilustración XXV - EX_FUENTES_FACT_ALUMNO	260
Ilustración XXVI - EX_SRC_POSTULANTES	260
Ilustración XXVII - EX_SRC_MATRICULA	261
Ilustración XXVIII - RN_CW_DIM_PROCESO	261
Ilustración XXIX - RN_CW_DIM_PROCESO	262
Ilustración XXX - RN_FACT_ALUMNO_GRP1	263
Ilustración XXXI - RN_FACT_ALUMNOS_GRP1_1	263

Ilustración XXXII - RN_FACT_ALUMNOS_GRP1_2	264
Ilustración XXXIII - RN_FACT_ALUMNOS_GRP1_3	264
Ilustración XXXIV - RN_FACT_ALUMNO_GRP2	265
Ilustración XXXV - RN_FACT_ALUMNOS_GRP2_1	265
Ilustración XXXVI - RN_FACT_ALUMNOS_GRP2_2	266
Ilustración XXXVII - CW_FACT_ALUMNO	266
Ilustración XXXVIII - CW_FACT_ALUMNO	267
Ilustración XXXIX - CARGA DE SEVERIDADES	267
Ilustración XL - CW_SEVERIDADES	268
Ilustración XLI - LIMPIEZA DE TEMPORALES	268

1. Paquete ETL_DIMENSIONES_NCD

El paquete ETL_DIMENSIONES_NCD no presenta un flujo de datos dentro de la herramienta SSIS debido a que la información es cargada por medio de un script que se ejecuta desde una tarea SQL.

Ilustración I - ETL_DIMENSIONES_NCD

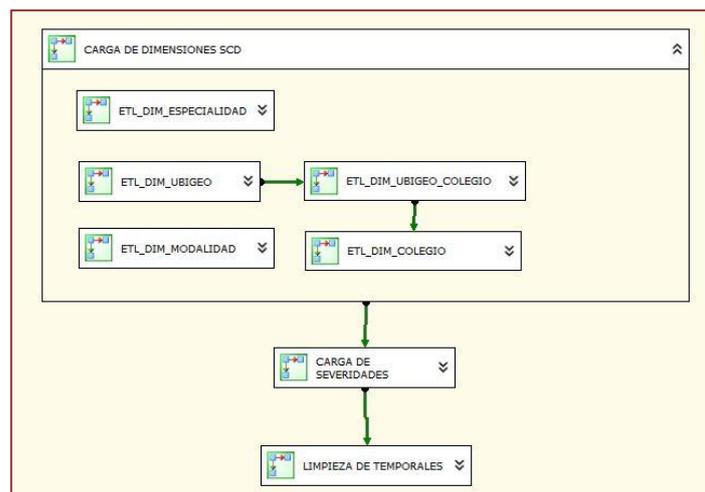


Elaboración: los autores

2. Paquete ETL_DIMENSIONES_SCD

El paquete ETL_DIMENSIONES_SCD presenta diferentes contenedores de secuencias, cada uno con un grupo de tareas de flujo de datos.

Ilustración II - ETL_DIMENSIONES_SCD

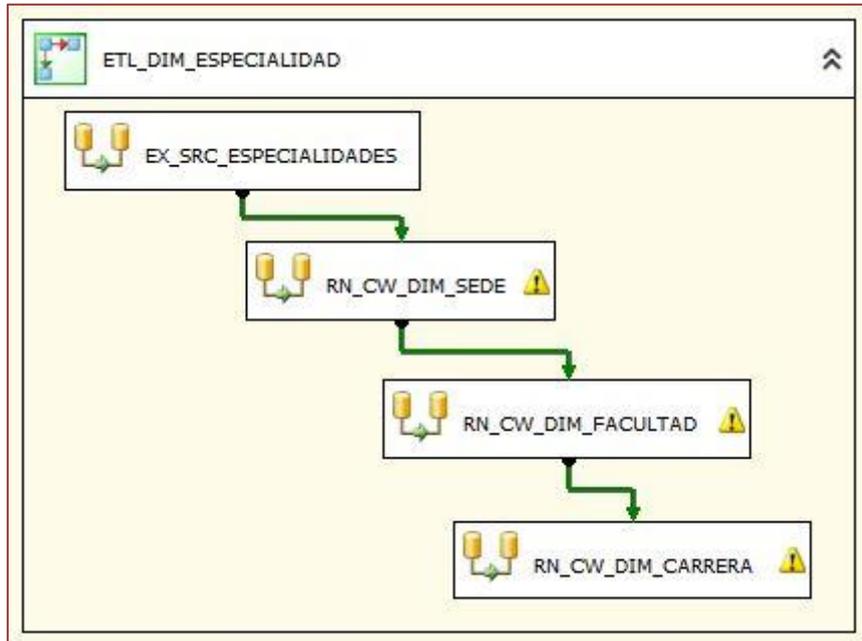


Elaboración: los autores

2.1 ETL_DIM_ESPECIALIDAD

Este contenedor presenta cuatro flujos de datos diferentes. Uno de extracción y los otros tres de reglas de negocio y carga final.

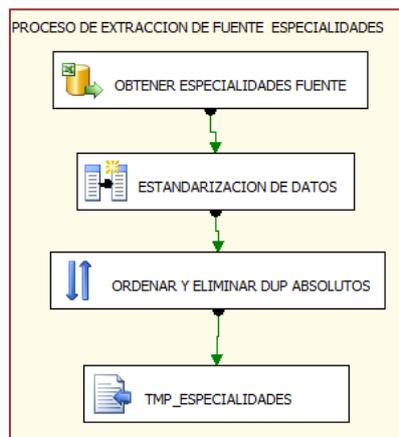
Ilustración III - ETL_DIM_ESPECIALIDAD



Elaboración: los autores

- Flujo de datos EX_SRC_ESPECIALIDADES

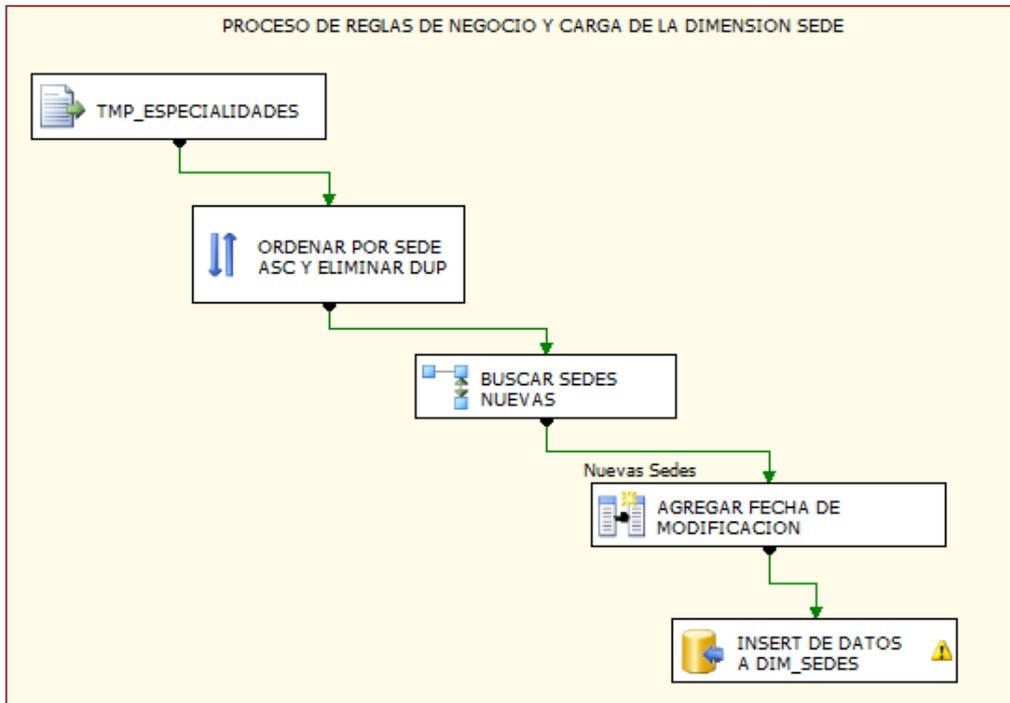
Ilustración IV - EX_SRC_POSTULANTES



Elaboración: los autores

- Flujo de datos RN_CW_DIM_SEDE

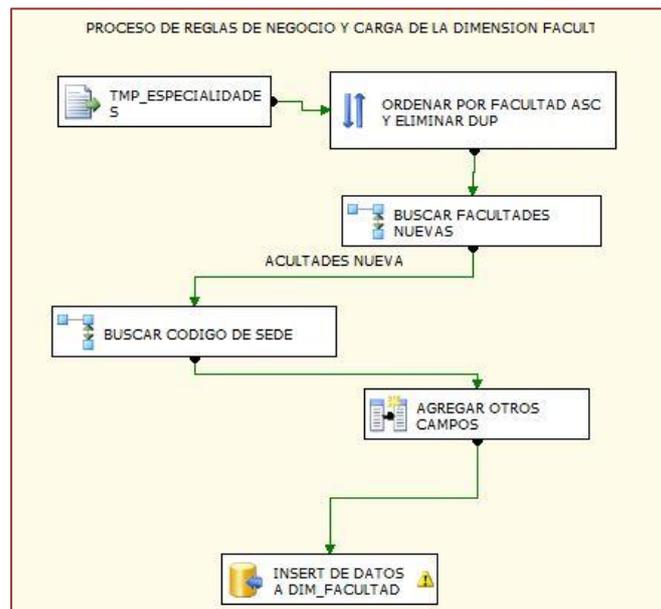
Ilustración V - EX_DIM_SEDE



Elaboración: los autores

- Flujo de datos RN_CW_DIM_FACULTAD

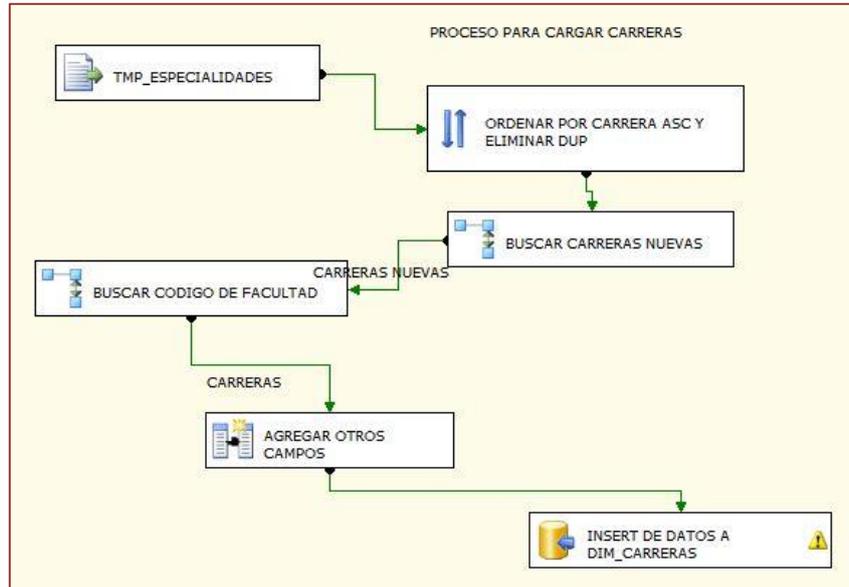
Ilustración VI - RN_CW_DIM_FACULTAD



Elaboración: los autores

- Flujo de datos RN_CW_DIM_CARRERA

Ilustración VII - RN_CW_DIM_CARRERA

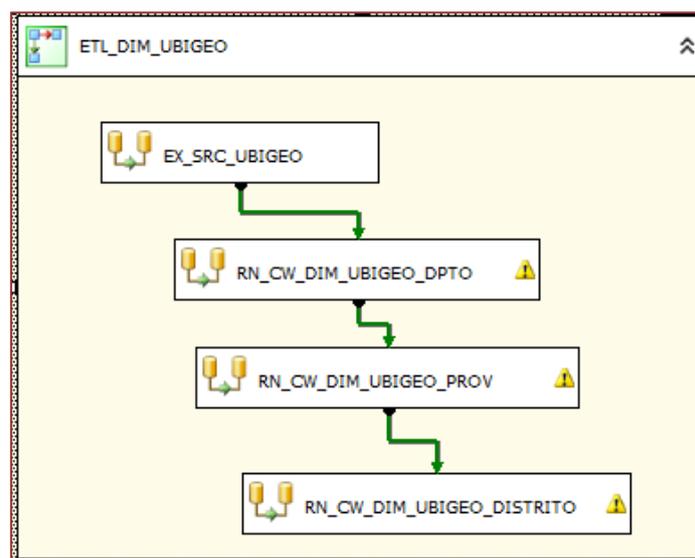


Elaboración: los autores

2.2 ELT_DIM_UBIGEO

Este contenedor presenta cuatro flujos de datos diferentes. Uno de extracción y los otros tres de reglas de negocio y carga final.

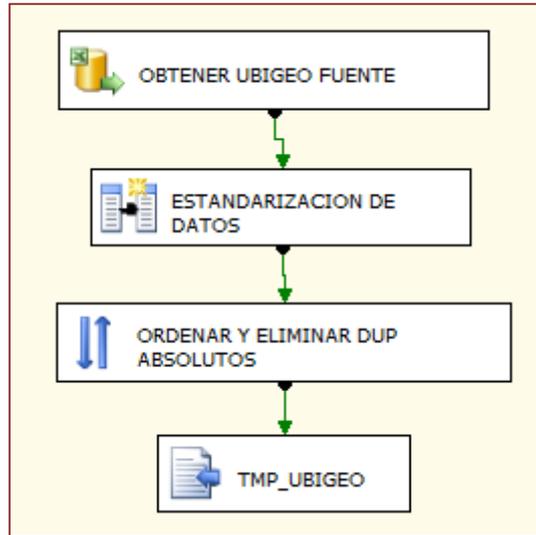
Ilustración VIII - ELT_DIM_UBIGEO



Elaboración: los autores

- Flujo de Datos EX_SRC_UBIGEO

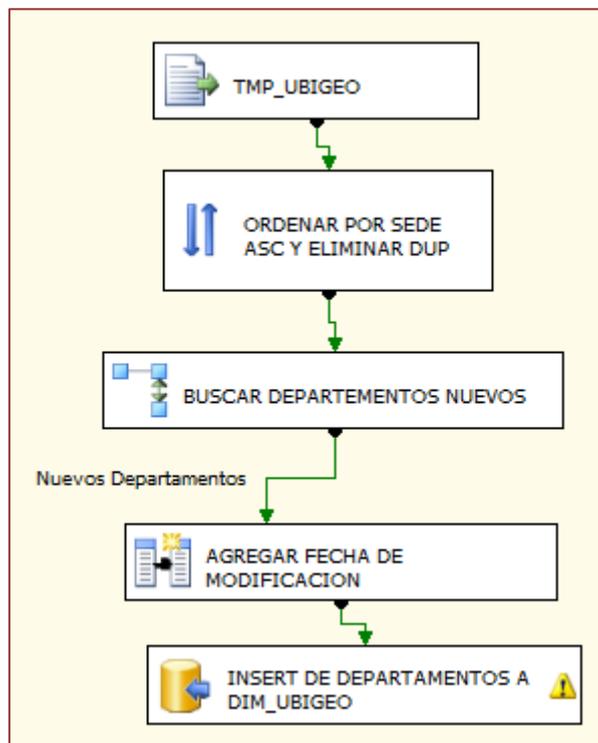
Ilustración IX - EX_SRC_UBIGEO



Elaboración: los autores

- Flujo de Datos RN_CW_DIM_UBIGEO_DPTO

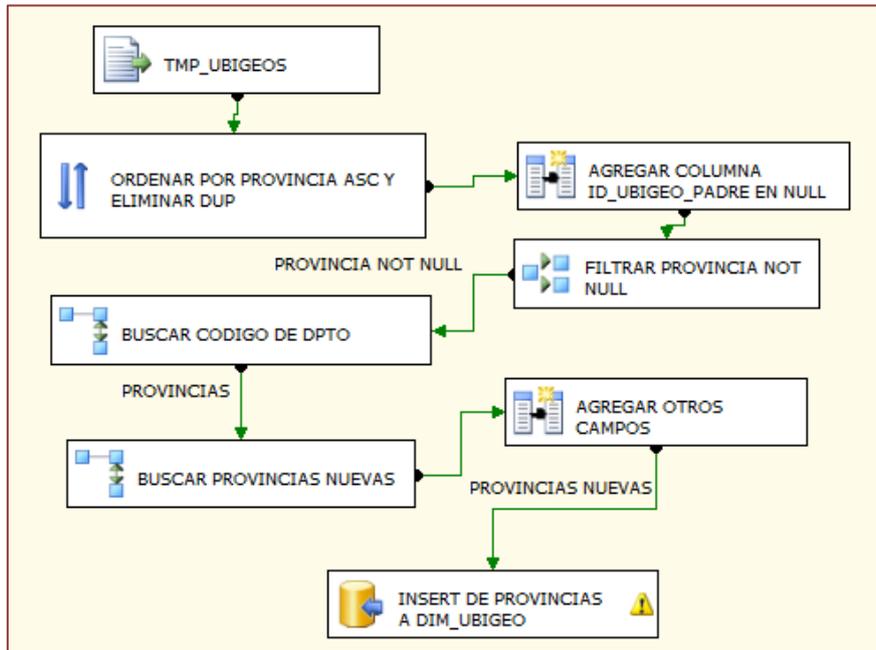
Ilustración X - RN_CW_DIM_UBIGEO_DPTO



Elaboración: los autores

- Flujo de Datos RN_CW_DIM_UBIGEO_PROV

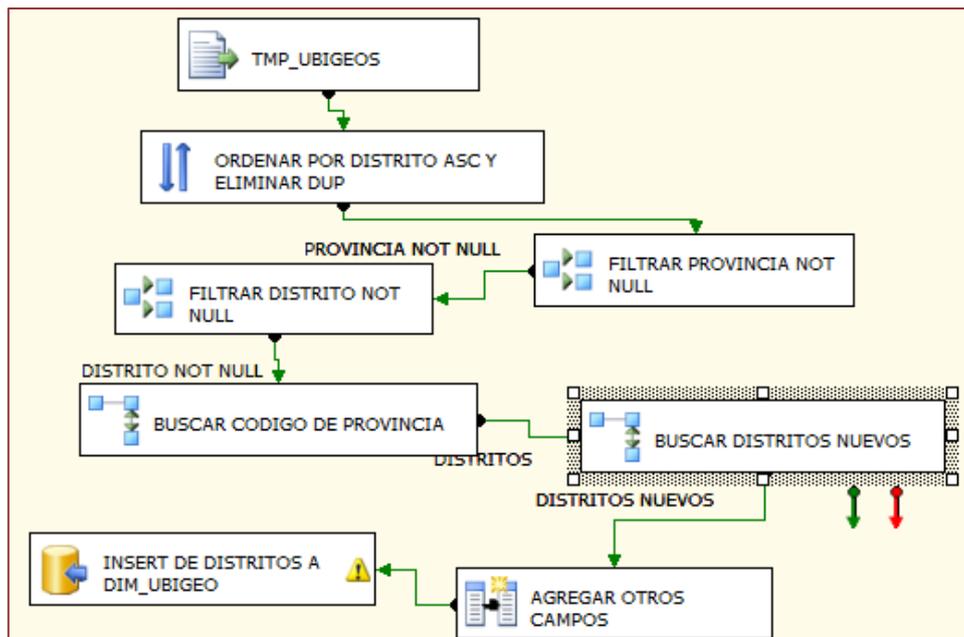
Ilustración XI - RN_CW_DIM_UBIGEO_PROV



Elaboración: los autores

- Flujo de Datos RN_CW_DIM_UBIGEO_DISTRITO

Ilustración XII - RN_CW_DIM_UBIGEO_DISTRITO

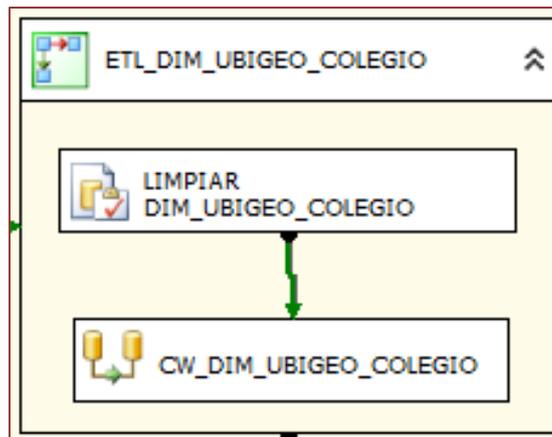


Elaboración: los autores

2.3 ETL_DIM_UBIGEO_COLEGIO

Este contenedor presenta una tarea SQL y un flujo de datos. La tarea SQL limpia toda la tabla DIM_UBIGEO_COLEGIO y luego el flujo es el que puebla la dicha tabla.

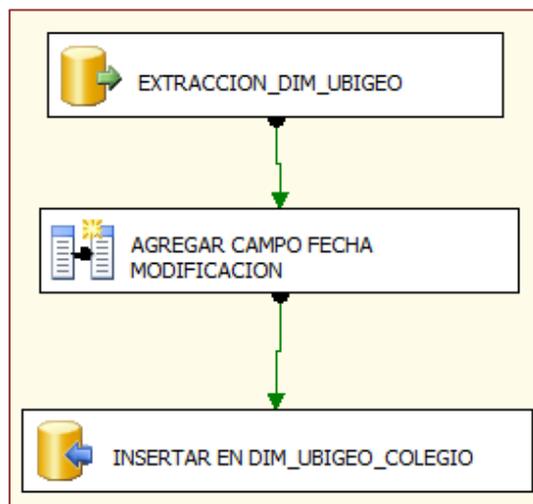
Ilustración XIII - ETL_DIM_UBIGEO_COLEGIO



Elaboración: los autores

- Flujo de Datos RN_CW_DIM_UBIGEO_DISTRITO

Ilustración XIV - ETL_DIM_UBIGEO_COLEGIO

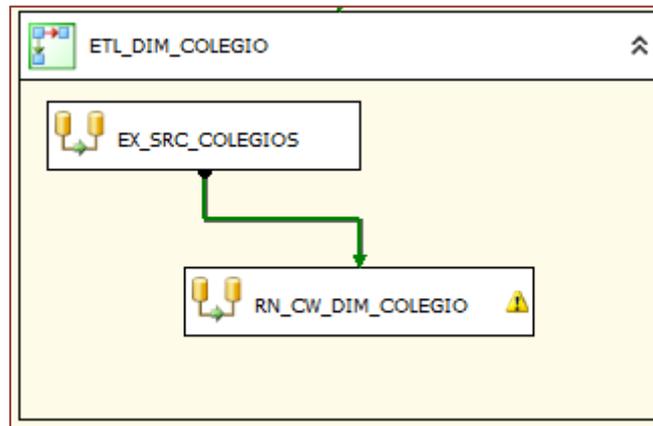


Elaboración: los autores

2.4 ETL_DIM_COLEGIO

Este contenedor presenta dos flujos de datos. La tarea SQL limpia toda la tabla DIM_UBIGEO_COLEGIO y luego el flujo es el que puebla la dicha tabla.

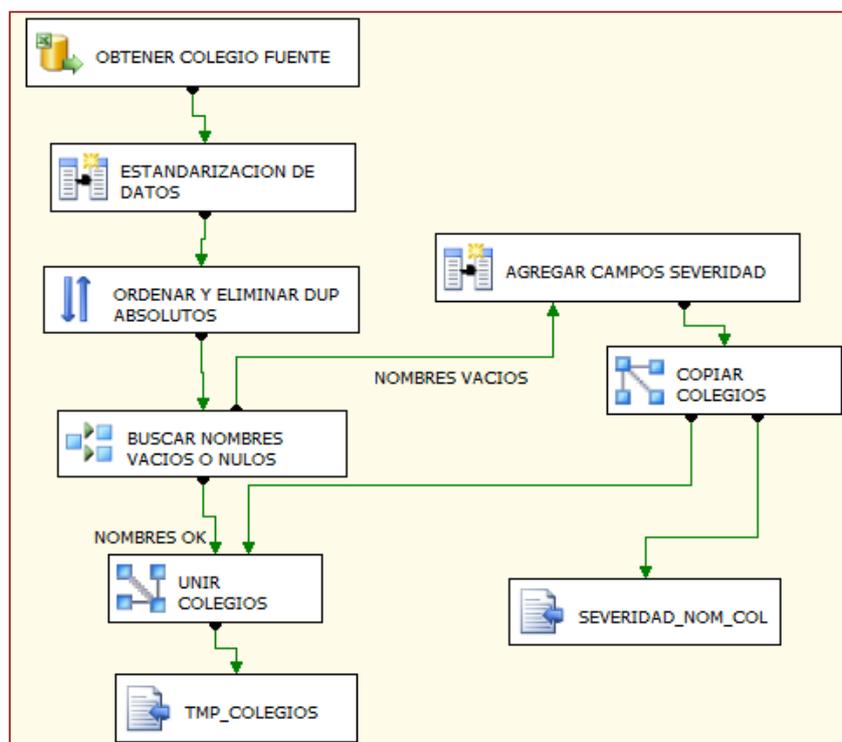
Ilustración XV - ETL_DIM_COLEGIO



Elaboración: los autores

- Flujo de Datos EX_SRC_COLEGIOS

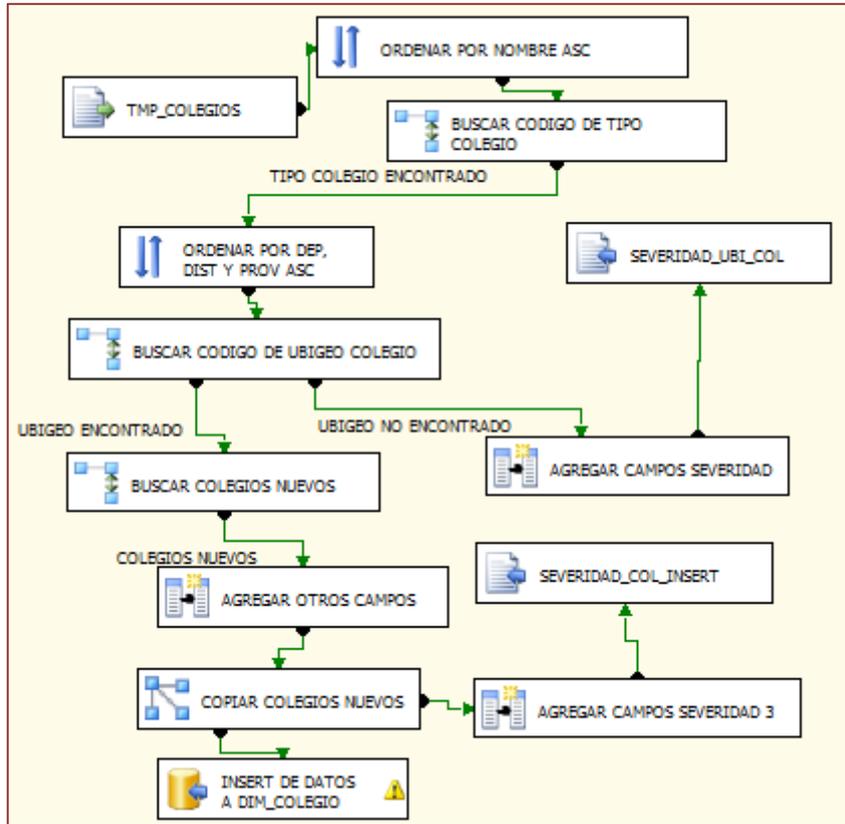
Ilustración XVI - EX_SRC_COLEGIOS



Elaboración: los autores

- Flujo de Datos RN_CW_DIM_COLEGIO

Ilustración XVII - RN_CW_DIM_COLEGIO

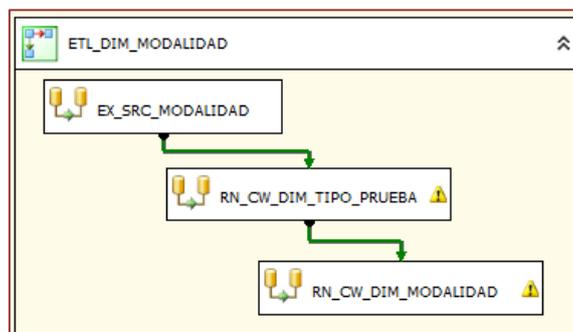


Elaboración: los autores

2.5 ETL_DIM_MODALIDAD

Este contenedor presenta dos flujos de datos. La tarea SQL limpia toda la tabla DIM_UBIGEO_COLEGIO y luego el flujo es el que puebla la dicha tabla.

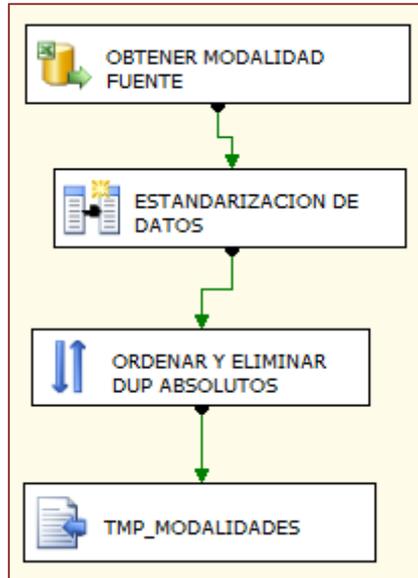
Ilustración XVIII - ETL_DIM_MODALIDAD



Elaboración: los autores

- Flujo de Datos EX_SRC_MODALIDAD

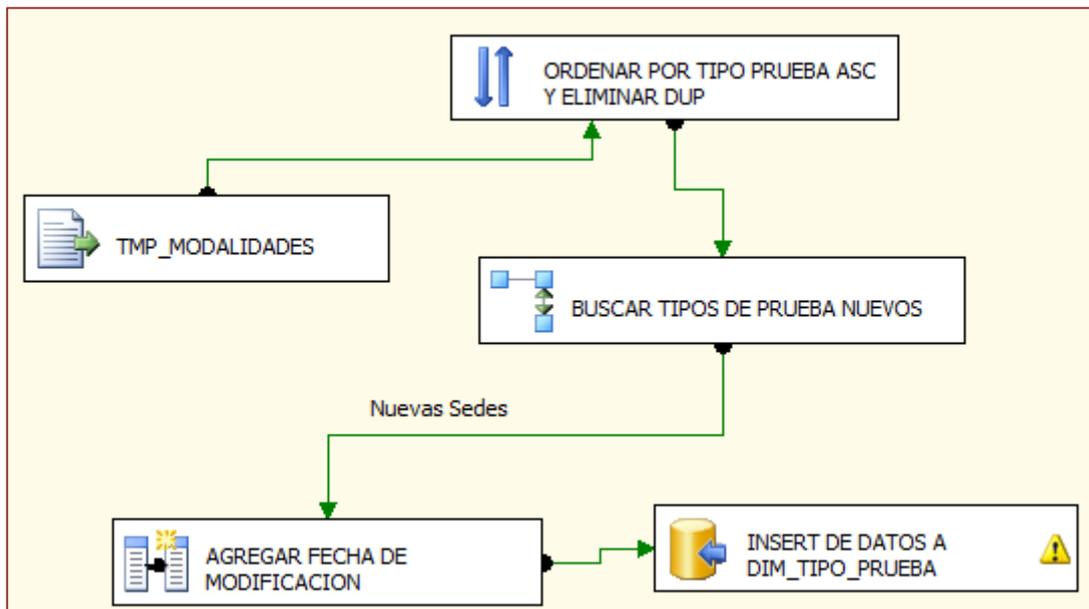
Ilustración XIX - EX_SRC_MODALIDAD



Elaboración: los autores

- Flujo de Datos RN_CW_DIM_TIPO_PRUEBA

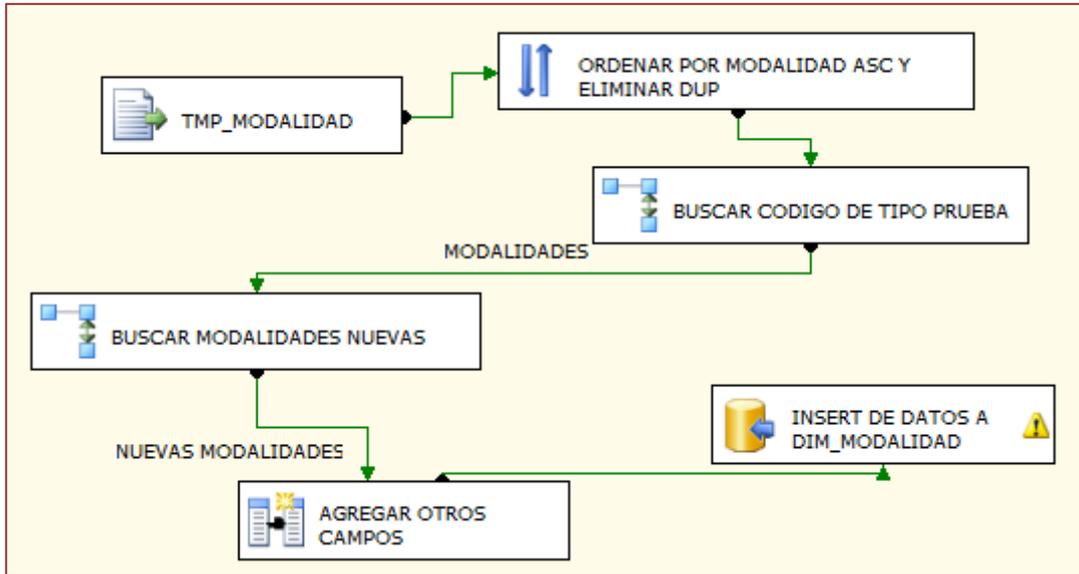
Ilustración XX - RN_CW_DIM_TIPO_PRUEBA



Elaboración: los autores

- Flujo de Datos RN_CW_DIM_MODALIDAD

Ilustración XXI - RN_CW_DIM_MODALIDAD



Elaboración: los autores

2.6 CARGA DE SEVERIDADES – Dimensiones SCD

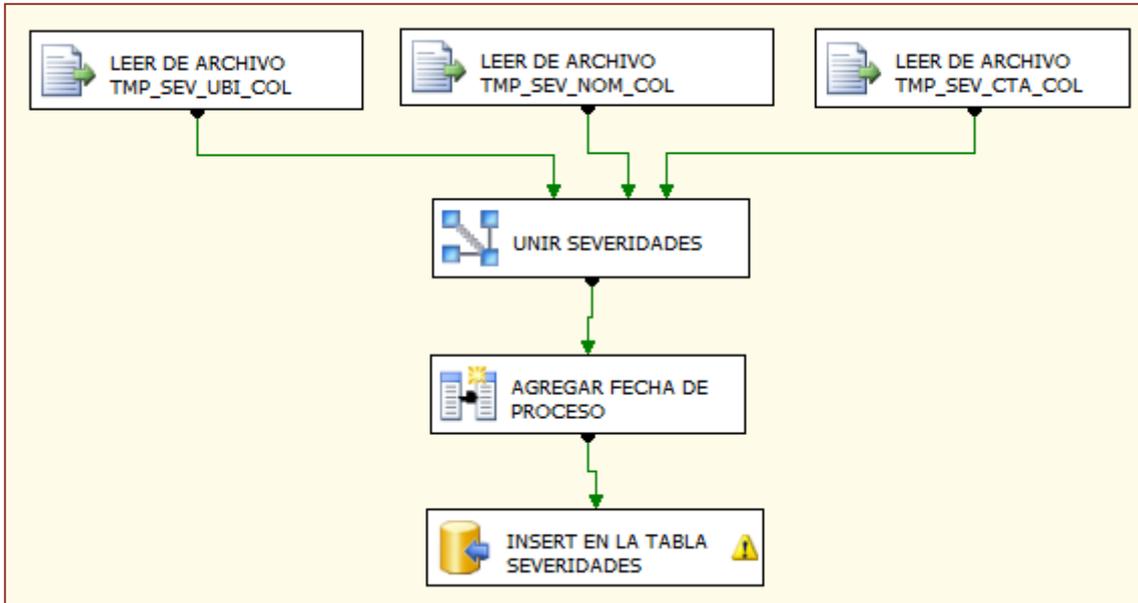
Ilustración XXII - CARGA DE SEVERIDADES



Elaboración: los autores

- Flujo de Datos CW_SEVERIDADES – Dimensiones SCD

Ilustración XXIII - CW_SEVERIDADES

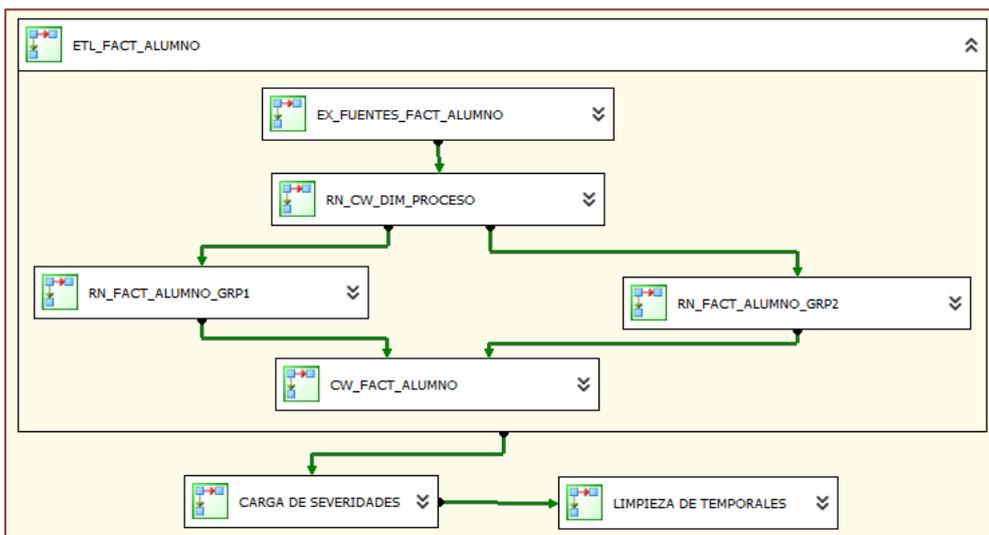


Elaboración: los autores

3. Paquete ETL_FACT_ALUMNO

El paquete ETL_FACT_ALUMNO presenta diferentes contenedores de secuencias, cada uno con un grupo de tareas de flujo de datos.

Ilustración XXIV - ETL_FACT_ALUMNO



Elaboración: los autores

3.1 EX_FUENTES_FACT_ALUMNO

Este contenedor presenta dos flujos de datos diferentes. Uno de extracción de postulantes y el otro de matriculados y no matriculados.

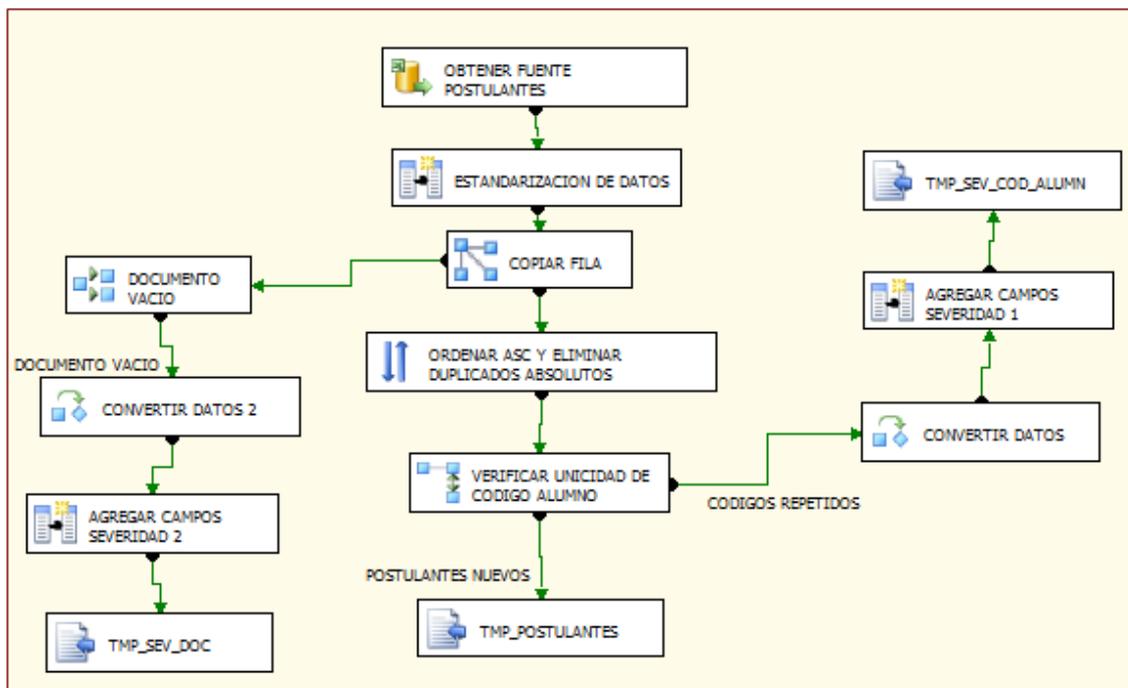
Ilustración XXV - EX_FUENTES_FACT_ALUMNO



Elaboración: los autores

- Flujo de Datos EX_SRC_POSTULANTES

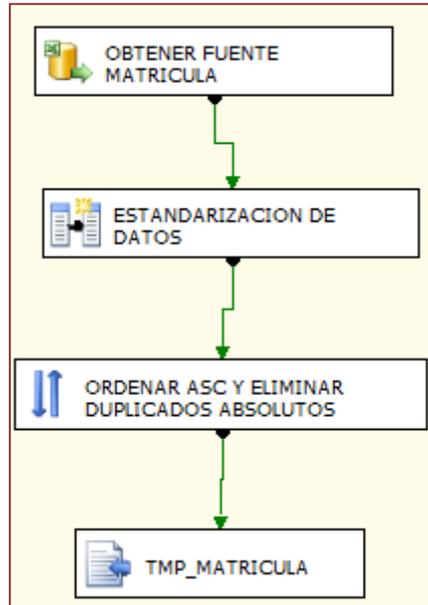
Ilustración XXVI - EX_SRC_POSTULANTES



Elaboración: los autores

- Flujo de Datos EX_SRC_MATRICULA

Ilustración XXVII - EX_SRC_MATRICULA

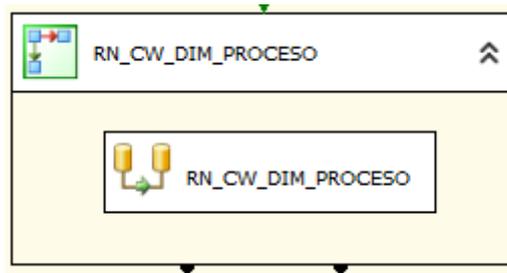


Elaboración: los autores

3.2 RN_CW_DIM_PROCESO

Este contenedor presenta un solo flujo de datos. Se utiliza un flujo de reglas de negocio y carga final a la tabla

Ilustración XXVIII - RN_CW_DIM_PROCESO



Elaboración: los autores

- Flujo de Datos RN_CW_DIM_PROCESO

Ilustración XXIX - RN_CW_DIM_PROCESO

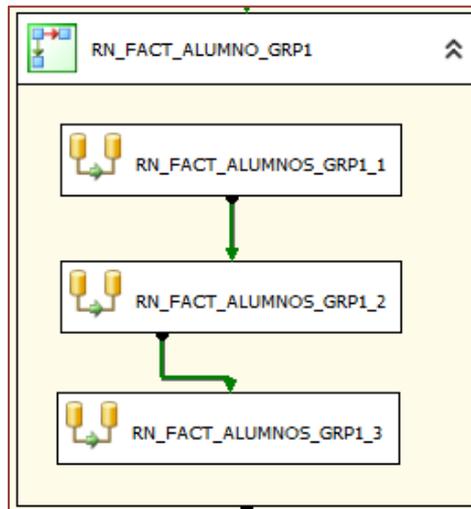


Elaboración: los autores

3.3 RN_FACT_ALUMNO_GRP1

Este contenedor presenta tres flujos de datos que podrían tomarse como uno solo. Todos son de reglas de negocio.

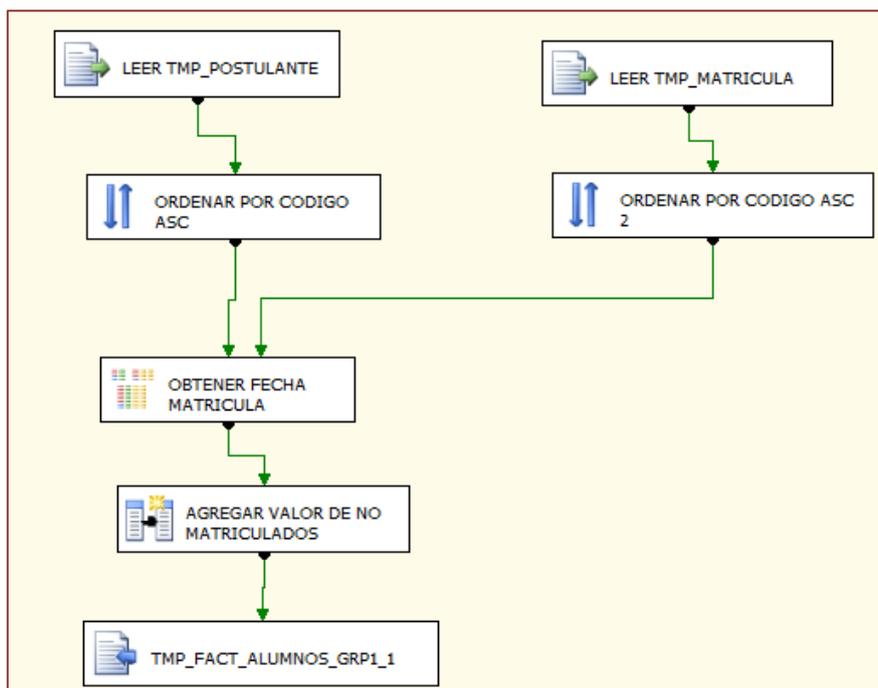
Ilustración XXX - RN_FACT_ALUMNO_GRP1



Elaboración: los autores

- Flujo de Datos RN_FACT_ALUMNOS_GRP1_1

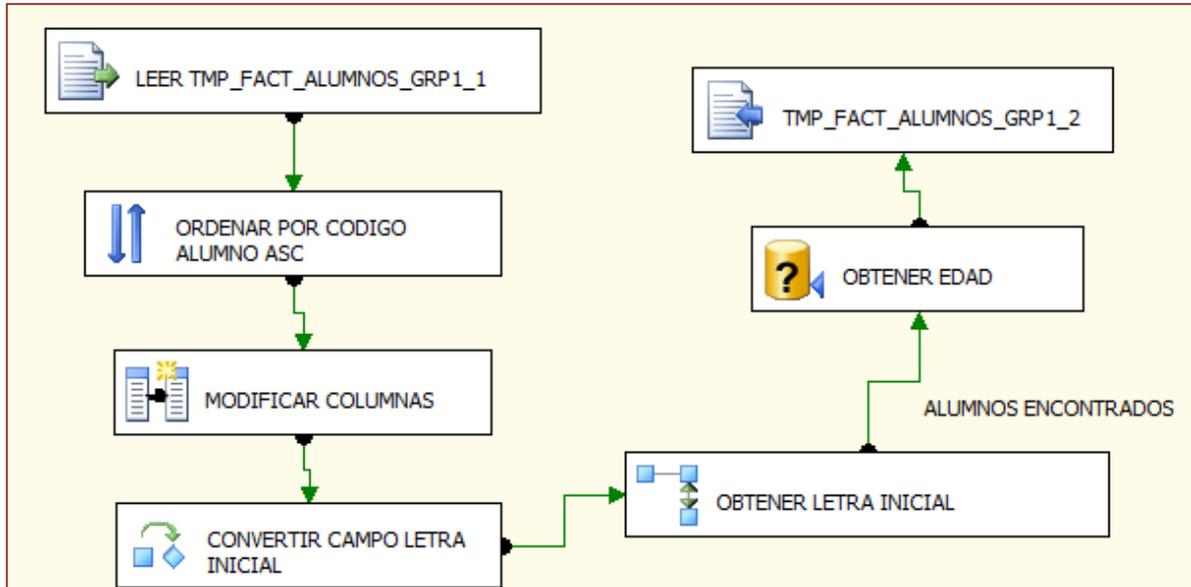
Ilustración XXXI - RN_FACT_ALUMNOS_GRP1_1



Elaboración: los autores

- Flujo de Datos RN_FACT_ALUMNOS_GRP1_2

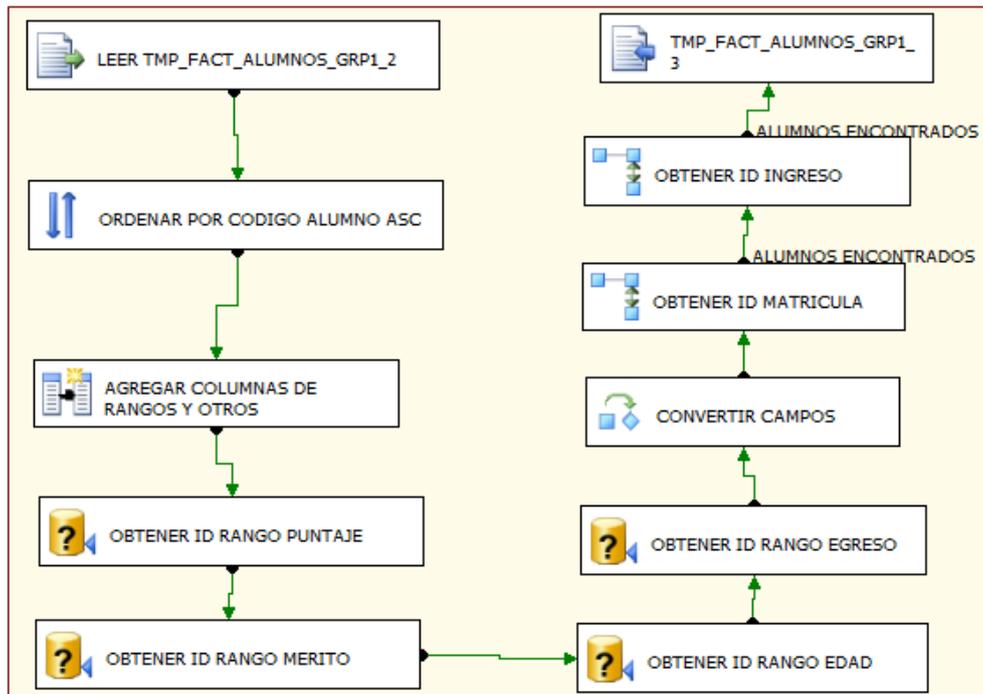
Ilustración XXXII - RN_FACT_ALUMNOS_GRP1_2



Elaboración: los autores

- Flujo de Datos RN_FACT_ALUMNOS_GRP1_3

Ilustración XXXIII - RN_FACT_ALUMNOS_GRP1_3

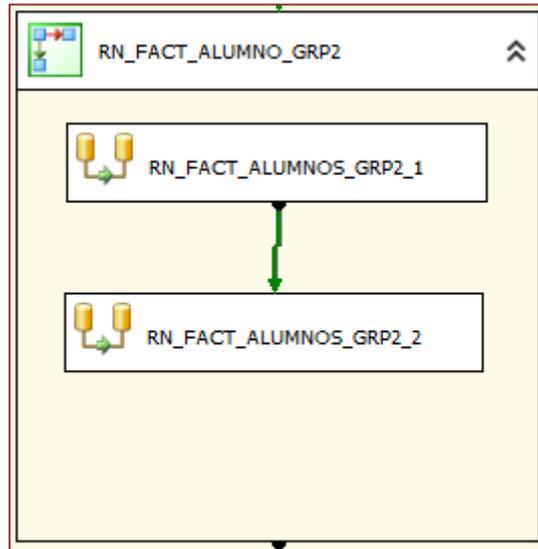


Elaboración: los autores

3.4 RN_FACT_ALUMNO_GRP2

Este contenedor presenta dos flujos de datos que podrían tomarse como uno solo. Todos son de reglas de negocio.

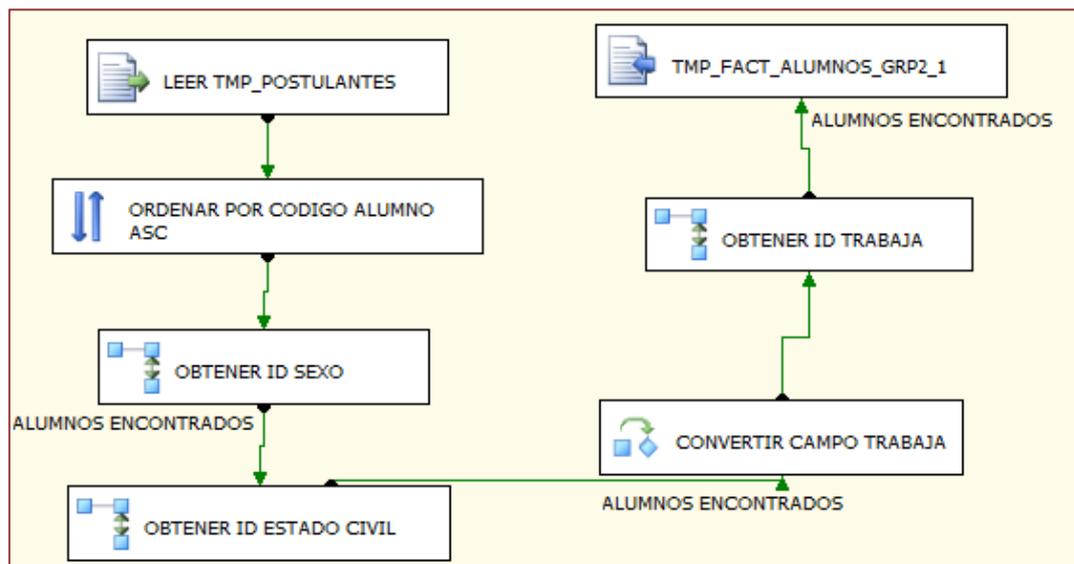
Ilustración XXXIV - RN_FACT_ALUMNO_GRP2



Elaboración: los autores

- Flujo de Datos RN_FACT_ALUMNOS_GRP2_1

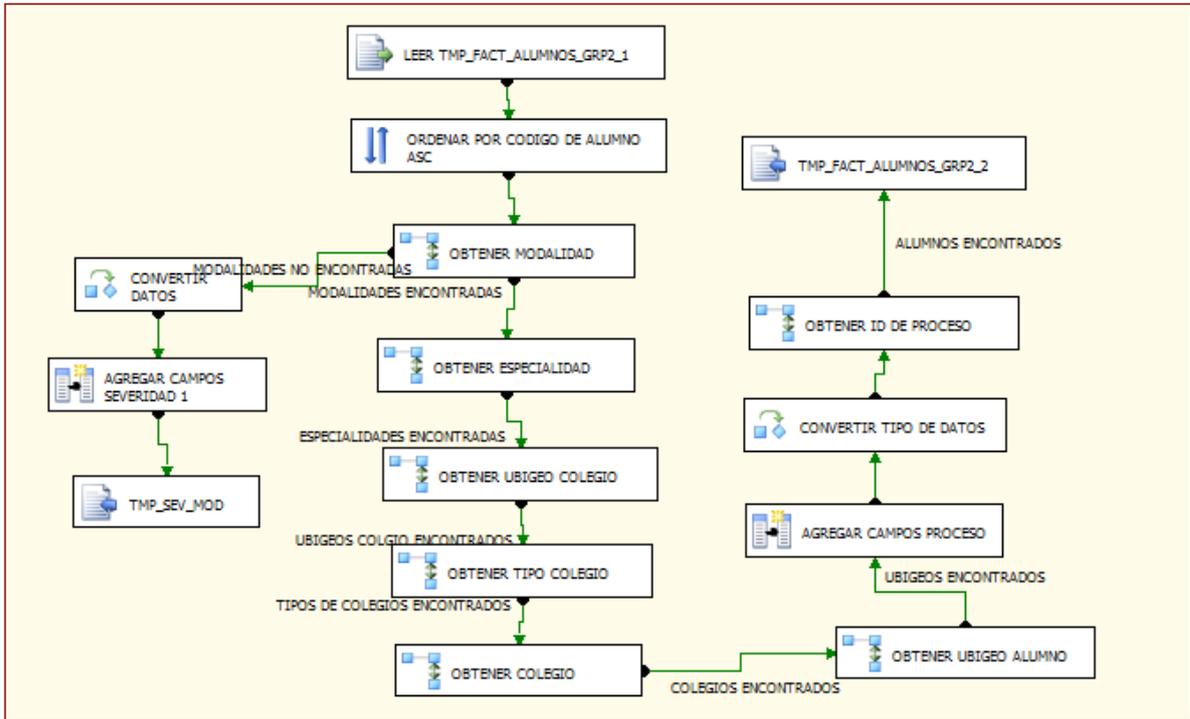
Ilustración XXXV - RN_FACT_ALUMNOS_GRP2_1



Elaboración: los autores

- Flujo de Datos RN_FACT_ALUMNOS_GRP2_2

Ilustración XXXVI - RN_FACT_ALUMNOS_GRP2_2

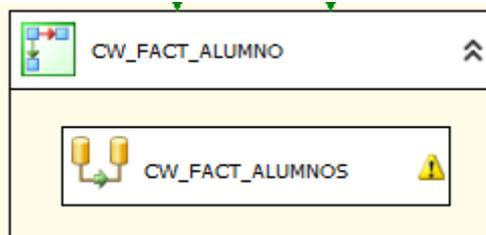


Elaboración: los autores

3.5 CW_FACT_ALUMNO

Este contenedor presenta un flujo de datos que se encarga de cargar la tabla final.

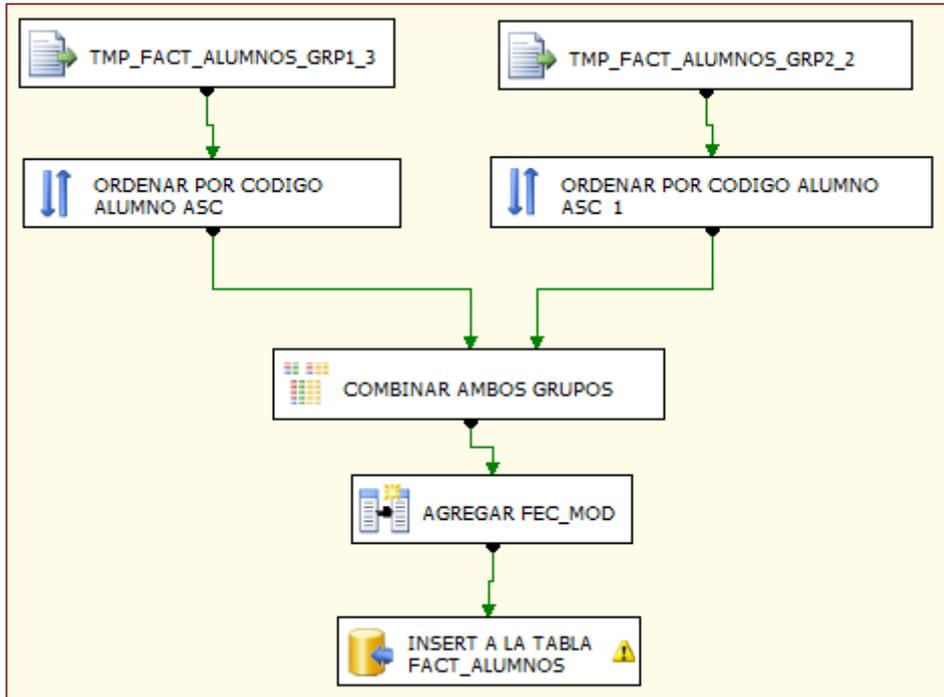
Ilustración XXXVII - CW_FACT_ALUMNO



Elaboración: los autores

- Flujo de Datos CW_FACT_ALUMNO

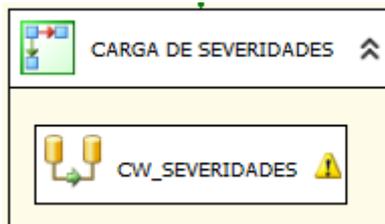
Ilustración XXXVIII - CW_FACT_ALUMNO



Elaboración: los autores

3.6 CARGA DE SEVERIDADES – FACT_ALUMNOS

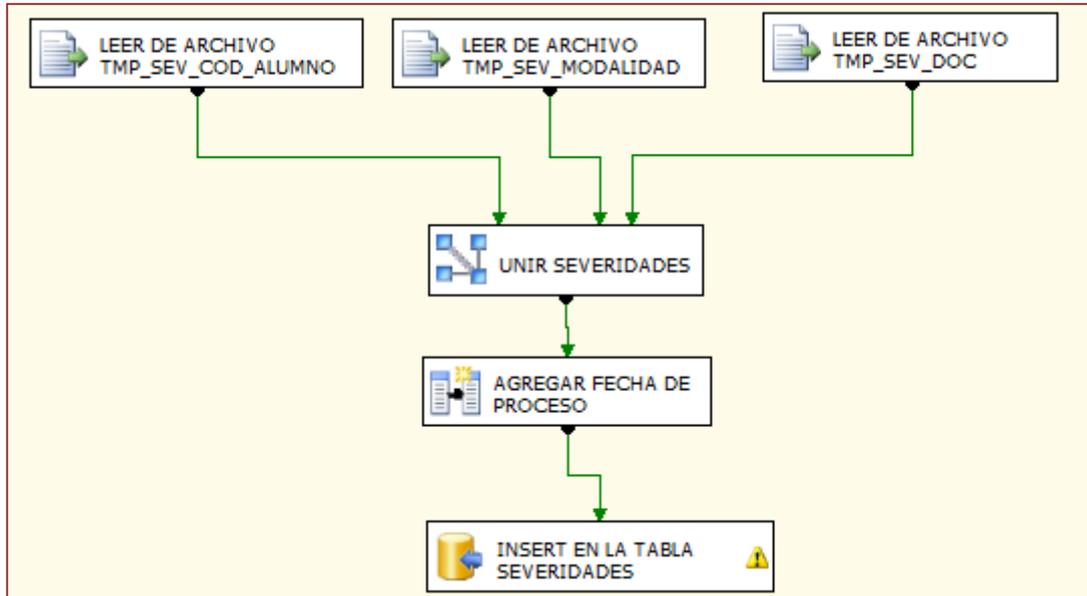
Ilustración XXXIX - CARGA DE SEVERIDADES



Elaboración: los autores

- Flujo de Datos CW_SEVERIDADES

Ilustración XL - CW_SEVERIDADES



Elaboración: los autores

3.7 LIMPIEZA DE TEMPORALES

En este contenedor se muestran dos Tareas de Sistemas de Archivos. Estas se utilizan para eliminar archivos temporales utilizados en los procesos ETL.

Ilustración XLI - LIMPIEZA DE TEMPORALES



Elaboración: los autores

ANEXO 16
DEFINICIÓN DE SEVERIDADES

INDICE

1. Definición	272
2. Modelo lógico	273
3. Diseño de tablas, campos e índices	273
4. Diseño físico	275
5. Severidades implementadas en la simulación	275

ÍNDICE DE ILUSTRACIONES

Ilustración I - Modelo Lógico	273
Ilustración II - Diseño físico	275

ÍNDICE DE TABLAS

Tabla I - Tablas de severidades	273
Tabla II - Tabla Severidades	274
Tabla III - Tabla Tipo_Severidad	274
Tabla IV - Severidades implementadas	275

1. Definición

La severidad es un concepto que se utilizó en el desarrollo de los procesos ETL, para el manejo de alertas de posibles inconsistencias de calidad, durante la carga de información al Datamart. Principalmente son utilizadas para poder darle estabilidad operativa al Datamart y como mecanismo de alerta a posibles problemas en la ejecución.

Una severidad puede ser de tres tipos: Severidad Tipo 1, Severidad Tipo 2 y Severidad Tipo 3. Cada una tiene un grado de importancia distinto, siendo la de Tipo 1 de mayor importancia y la Tipo 3 de menor.

Las severidades de Tipo 1 son aquellas que están relacionadas con la integridad referencial dentro del modelo dimensional del Datamart. Si en algún punto del flujo de datos, existe alguna inconsistencia de integridad referencial, la severidad Tipo 1 se activa y automáticamente el registro afectado se excluye de la carga y se inserta en la tabla SEVERIDADES.

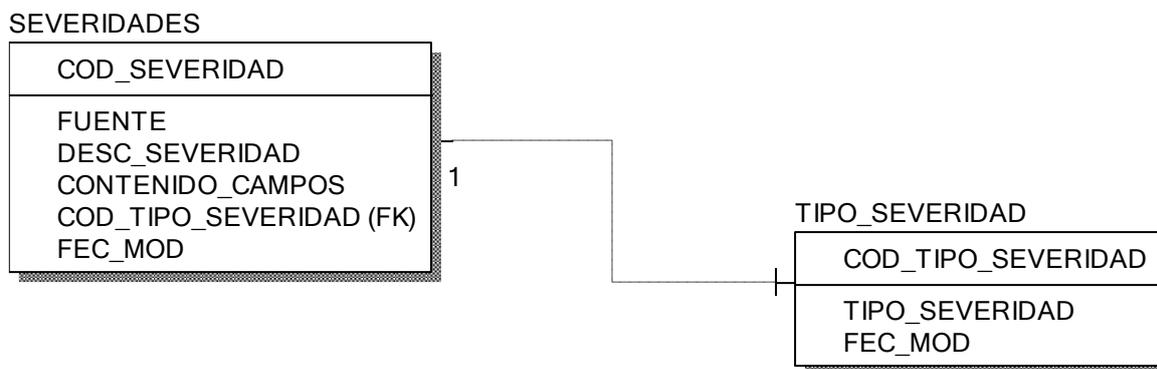
Las severidades de Tipo 2 son aquellas relacionadas a la completitud de un campo en particular. Si en el flujo de datos, algún registro posee uno o más campos vacíos cuando la definición de dicho campo dice lo contrario, el registro se carga normalmente, pero se genera una severidad Tipo 2.

Las severidades Tipo 3, son meramente informativas, es decir, son utilizadas en el caso de que alguna Dimensión SCD se actualice o cambien en el tiempo. Si en el flujo de datos, en alguna dimensión se inserta un registro nuevo, la severidad Tipo 3 se activa, registrando una alerta en la tabla de SEVERIDADES.

2. Modelo lógico

A continuación, se presenta el modelo lógico de las dos tablas de severidades. La tabla SEVERIDADES que almacena el detalle de la severidad y la tabla TIPO_SEVERIDAD que corresponde al tipo.

Ilustración I - Modelo Lógico



Elaboración: los autores

3. Diseño de tablas, campos e índices

El diseño de la tabla SEVERIDADES y la tabla TIPO_SEVERIDAD se realizó de la siguiente manera.

Tabla I - Tablas de severidades

Nombre de la Tabla	Descripción
SEVERIDADES	Tabla que almacena todas las incidencias con respecto a la calidad de la información.
TIPO_SEVERIDAD	Tabla que almacena los tres tipos de severidades (ENTIDAD REFERENCIAL, COMPLETITUD e INFORMATIVA)

Elaboración: los autores

- Tabla Severidades:

Tabla II - Tabla Severidades

Columnas de la tabla "SEVERIDADES"					
Nombre	Descripción	Tipo de dato	Opción Null	Es FK	Es PK
ID_SEVERIDAD	Identificador único de la severidad	int	IDENTITY	No	Sí
FUENTE	Nombre de la fuente de donde se genera la severidad	nvarchar(30)	NOT NULL	No	No
ID_TIPO_SEVERIDAD	Identificador único del tipo de severidad	smallint	NOT NULL	Sí	No
DESC_SEVERIDAD	Descripción de la severidad	nvarchar(100)	NOT NULL	No	No
CONTENIDO_CAMPOS	Contenido de los campos del registro que activó la severidad	nvarchar(400)	NULL	No	No
FEC_MOD	Última fecha de modificación de la tabla	smalldatetime	NOT NULL	No	No

Elaboración: los autores

- Tabla Tipo_Severidad:

Tabla III - Tabla Tipo_Severidad

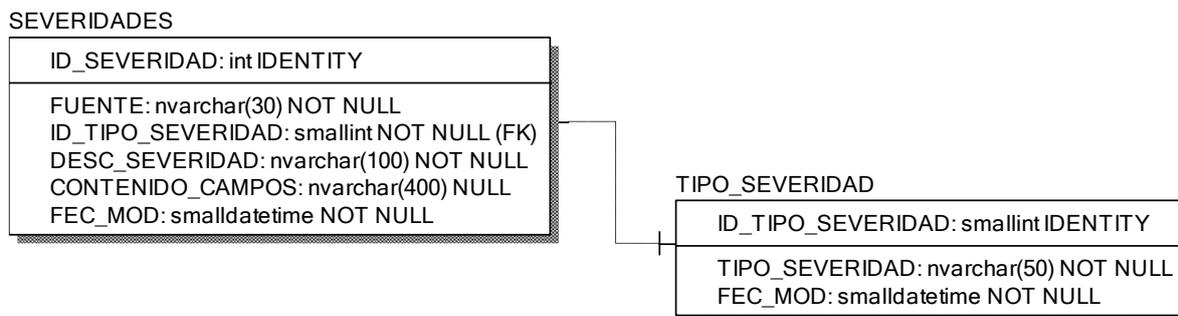
Columnas de la tabla "TIPO_SEVERIDAD"					
Nombre	Descripción	Tipo de dato	Opción Null	Es FK	Es PK
ID_TIPO_SEVERIDAD	Identificador único del tipo de severidad	smallint	IDENTITY	No	Sí
TIPO_SEVERIDAD	Tipo de severidad	nvarchar(50)	NOT NULL	No	No
FEC_MOD	Última fecha de modificación de la tabla	smalldatetime	NOT NULL	No	No

Elaboración: los autores

4. Diseño físico

A continuación, se presenta el diseño físico de las tablas de severidades. Para ambas tablas, se utilizó una llave secuencial auto numérica que inicia en 1 y se incluyó el campo FEC_MOD para registrar la fecha de actualización de la tabla.

Ilustración II - Diseño físico



Elaboración: los autores

5. Severidades implementadas en la simulación

Tabla IV - Severidades implementadas

Nombre	Tipo	Descripción	Paquete - Flujo
Nombre Colegio	Complejidad	Verifica que el nombre del colegio no venga vacío ni nulo.	ETL_DIMENSIONES_SCD / EX_SRC_COLEGIOS
Ubigeo Colegio	Identidad Referencial	Verifica que exista un ubigeo válido para el colegio	ETL_DIMENSIONES_SCD / RN_CW_DIM_COLEGIO
Colegios Nuevos	Informativa	Verifica la creación de nuevos colegios.	ETL_DIMENSIONES_SCD / RN_CW_DIM_COLEGIO
Número de Documento	Complejidad	Verifica que el número de documento no sea vacío ni nulo	ETL_FACT_ALUMNO / EX_SRC_POSTULANTES
Modalidad de Ingreso	Identidad Referencial	Verifica que exista una modalidad válida para el ingresante	ETL_FACT_ALUMNO / RN_FACT_ALUMNOS_GRP2_2

Elaboración: los autores

ANEXO 17
KPI'S – INDICADORES CLAVE DE RENDIMIENTO

ÍNDICE

	Página
1. Indicadores	278
2. Indicadores de eficiencia	279
3. Indicadores de matrícula realizada	284
4. Indicadores de matrículas no realizadas	288

ÍNDICE DE ILUSTRACIONES

Ilustración - I - Indicadores de eficiencia	278
Ilustración - II - Indicadores de matricula	278
Ilustración - III - Indicadores de matrículas no realizadas	279
Ilustración - IV - Eficiencia matricula	279
Ilustración - V - Eficiencia_Matricula_Arquitectura	280
Ilustración - VI - Eficiencia_Matricula_Aeronauticas	280
Ilustración - VII - Eficiencia_Matricula_Industrial	281
Ilustración - VIII - Eficiencia_Matricula_Civil	281
Ilustración - IX - Eficiencia_Matricula_Sistemas	282
Ilustración - X - Eficiencia_Matricula_Alimentarias	282
Ilustración - XI - Eficiencia_Matricula_Electronica	283
Ilustración - XII - Alumnos matriculados	284
Ilustración - XIII - Matriculados_Arquitectura	284
Ilustración - XIV - Matriculados_Aeronauticas	285
Ilustración - XV - Matriculados_Alimentarias	285
Ilustración - XVI - Matriculados_Sistemas	286
Ilustración - XVII - Matriculados_Industrial	286
Ilustración - XVIII - Matriculados_Electronica	287
Ilustración - XIX - Matriculados_Civil	287
Ilustración - XX - Alumnos_NO_Matriculados	288
Ilustración - XXI - Arquitectura_NO_Matriculados	288
Ilustración - XXII - Aeronauticas_NO_Matriculados	289
Ilustración - XXIII - Civil_NO_Matriculados	289
Ilustración - XXIV - Sistemas_NO_Matriculados	290
Ilustración - XXV - Alimentarias_NO_Matriculados	290
Ilustración - XXVI - Electronica_NO_Matriculados	291
Ilustración - XXVII - Industrial_NO_Matriculados	291

1. Indicadores

Se desarrollaron los siguientes indicadores claves de rendimiento:

Ilustración - I - Indicadores de eficiencia



Elaboración: los autores

Ilustración - II - Indicadores de matricula



Elaboración: los autores

Ilustración - III - Indicadores de matrículas no realizadas



Elaboración: los autores

2. Indicadores de eficiencia

Ilustración - IV - Eficiencia matricula

Nombre:

Grupo de medida asociado:

⌵ Expresión de valor

```
(([Measures].[Numero Postulantes],[Ingreso].[Ingreso].&[1],[Matricula].[Matricula].&[1])/([Measures].[Numero Postulantes],[Ingreso].[Ingreso].&[1]))
```

⌵ Expresión objetivo

0.85

⌵ Estado

Indicador de estado:

Expresión de estado:

```
case
  when kpivalue("Eficiencia_Matricula")>=.85 then 1
  when kpivalue("Eficiencia_Matricula")<.85 and
    kpivalue("Eficiencia_Matricula")>=.70 then 0
  else -1
end
```

Elaboración: los autores

Ilustración - V - Eficiencia_Matricula_Arquitectura

Nombre:
Eficiencia_Matricula_Arquitectura

Grupo de medida asociado:
Medidas Postulantes

Expresión de valor
$$\frac{([Measures].[Numero Postulantes], [Ingreso].[Ingreso].\&[1], [Matricula].[Matricula].\&[1], [Especialidad].[Carrera].\&[11])}{([Measures].[Numero Postulantes], [Ingreso].[Ingreso].\&[1], [Especialidad].[Carrera].\&[11])}$$

Expresión objetivo

Estado
Indicador de estado: Cilindro

Expresión de estado:

```
case
  when kpivalue("Eficiencia_Matricula_Arquitectura")>=.85 then 1
  when kpivalue("Eficiencia_Matricula_Arquitectura")<.85 and
    kpivalue("Eficiencia_Matricula_Arquitectura")>=.70 then 0
  else -1
end
```

Elaboración: los autores

Ilustración - VI - Eficiencia_Matricula_Aeronauticas

Nombre:
Eficiencia_Matricula_Aeronauticas

Grupo de medida asociado:
Medidas Postulantes

Expresión de valor
$$\frac{([Measures].[Numero Postulantes], [Ingreso].[Ingreso].\&[1], [Matricula].[Matricula].\&[1], [Especialidad].[Carrera].\&[12])}{([Measures].[Numero Postulantes], [Ingreso].[Ingreso].\&[1], [Especialidad].[Carrera].\&[12])}$$

Expresión objetivo

Estado
Indicador de estado: Cilindro

Expresión de estado:

```
case
  when kpivalue("Eficiencia_Matricula_Aeronauticas")>=.85 then 1
  when kpivalue("Eficiencia_Matricula_Aeronauticas")<.85 and
    kpivalue("Eficiencia_Matricula_Aeronauticas")>=.70 then 0
  else -1
end
```

Elaboración: los autores

Ilustración - VII - Eficiencia_Matricula_Industrial

Nombre:
Eficiencia_Matricula_Industrial

Grupo de medida asociado:
Medidas Postulantes

Expresión de valor
$$\frac{([Measures].[Numero Postulantes], [Ingreso].[Ingreso].\&[1], [Matricula].[Matricula].\&[1], [Especialidad].[Carrera].\&[17])}{([Measures].[Numero Postulantes], [Ingreso].[Ingreso].\&[1], [Especialidad].[Carrera].\&[17])}$$

Expresión objetivo

Estado
Indicador de estado: Cilindro

Expresión de estado:

```
case
when kpivalue("Eficiencia_Matricula_Industrial")>=.85 then 1
when kpivalue("Eficiencia_Matricula_Industrial")<.85 and
kpivalue("Eficiencia_Matricula_Industrial")>=.70 then 0
else -1
end
```

Elaboración: los autores

Ilustración - VIII - Eficiencia_Matricula_Civil

Nombre:
Eficiencia_Matricula_Civil

Grupo de medida asociado:
Medidas Postulantes

Expresión de valor
$$\frac{([Measures].[Numero Postulantes], [Ingreso].[Ingreso].\&[1], [Matricula].[Matricula].\&[1], [Especialidad].[Carrera].\&[13])}{([Measures].[Numero Postulantes], [Ingreso].[Ingreso].\&[1], [Especialidad].[Carrera].\&[13])}$$

Expresión objetivo

Estado
Indicador de estado: Cilindro

Expresión de estado:

```
case
when kpivalue("Eficiencia_Matricula_Civil")>=.85 then 1
when kpivalue("Eficiencia_Matricula_Civil")<.85 and
kpivalue("Eficiencia_Matricula_Civil")>=.70 then 0
else -1
end
```

Elaboración: los autores

Ilustración - IX - Eficiencia_Matricula_Sistemas

Nombre:
Eficiencia_Matricula_Sistemas

Grupo de medida asociado:
Medidas Postulantes

Expresión de valor
$$\frac{([Measures].[Numero Postulantes], [Ingreso].[Ingreso].\&[1], [Matricula].[Matricula].\&[1], [Especialidad].[Carrera].\&[14])}{([Measures].[Numero Postulantes], [Ingreso].[Ingreso].\&[1], [Especialidad].[Carrera].\&[14])}$$

Expresión objetivo

Estado

Indicador de estado: Cilindro

Expresión de estado:

```
case
when kpivalue ("Eficiencia_Matricula_Sistemas") >= .85 then 1
when kpivalue ("Eficiencia_Matricula_Sistemas") < .85 and
kpivalue ("Eficiencia_Matricula_Sistemas") >= .70 then 0
else -1
end
```

Elaboración: los autores

Ilustración - X - Eficiencia_Matricula_Alimentarias

Nombre:
Eficiencia_Matricula_Alimentarias

Grupo de medida asociado:
Medidas Postulantes

Expresión de valor
$$\frac{([Measures].[Numero Postulantes], [Ingreso].[Ingreso].\&[1], [Matricula].[Matricula].\&[1], [Especialidad].[Carrera].\&[16])}{([Measures].[Numero Postulantes], [Ingreso].[Ingreso].\&[1], [Especialidad].[Carrera].\&[16])}$$

Expresión objetivo

Estado

Indicador de estado: Cilindro

Expresión de estado:

```
case
when kpivalue ("Eficiencia_Matricula_Alimentarias") >= .85 then 1
when kpivalue ("Eficiencia_Matricula_Alimentarias") < .85 and
kpivalue ("Eficiencia_Matricula_Alimentarias") >= .70 then 0
else -1
end
```

Elaboración: los autores

Ilustración - XI - Eficiencia_Matricula_Electronica

Nombre:

Grupo de medida asociado:

⌵ Expresión de valor

```
(([Measures].[Numero Postulantes],[Ingreso].[Ingreso].&[1],[Matricula].[Matricula].&[1],[Especialidad].[Carrera].&[15])/([Measures].[Numero Postulantes],[Ingreso].[Ingreso].&[1],[Especialidad].[Carrera].&[15]))
```

⌵ Expresión objetivo

⌵ Estado

Indicador de estado:

Expresión de estado:

```
case  
  when kpivalue("Eficiencia_Matricula_Electronica")>=.85 then 1  
  when kpivalue("Eficiencia_Matricula_Electronica")<.85 and  
    kpivalue("Eficiencia_Matricula_Electronica")>=.70 then 0  
  else -1  
end
```

Elaboración: los autores

3. Indicadores de matrícula realizada

Ilustración - XII - Alumnos matriculados

Nombre:
AlumnosMatriculados

Grupo de medida asociado:
Medidas Postulantes

Expresión de valor
{[Measures].[Numero Postulantes],[Ingreso].[Ingreso].&[1],[Matricula].[Matricula].&[1]}

Expresión objetivo
752

Estado
Indicador de estado: Medidor

Expresión de estado:

```
case
when kpivalue("AlumnosMatriculados")>= 752 then 1
when kpivalue("AlumnosMatriculados")<752 and
kpivalue("AlumnosMatriculados")>=703 then 0
else -1
end
```

Elaboración: los autores

Ilustración - XIII - Matriculados_Arquitectura

Nombre:
Matriculados_Arquitectura

Grupo de medida asociado:
Medidas Postulantes

Expresión de valor
{[Measures].[Numero Postulantes],[Matricula].[Matricula].&[1],[Especialidad].[Carrera].&[1]}

Expresión objetivo

Estado
Indicador de estado: Formas

Expresión de estado:

```
case
when kpivalue("Matriculados_Arquitectura")>= 107 then 1
when kpivalue("Matriculados_Arquitectura")<107 and
kpivalue("Matriculados_Arquitectura")>=91 then 0
else -1
end
```

Elaboración: los autores

Ilustración - XIV - Matriculados_Aeronauticas

Nombre:

Grupo de medida asociado:

Expresión de valor

Expresión objetivo

Estado
Indicador de estado:

Expresión de estado:

```
case
  when kpivalue("Matriculados_Aeronauticas")>= 107 then 1
  when kpivalue("Matriculados_Aeronauticas")<107 and
    kpivalue("Matriculados_Aeronauticas")>=91 then 0
  else -1
end
```

Elaboración: los autores

Ilustración - XV - Matriculados_Alimentarias

Nombre:

Grupo de medida asociado:

Expresión de valor

Expresión objetivo

Estado
Indicador de estado:

Expresión de estado:

```
case
  when kpivalue("Matriculados_Alimentarias")>= 107 then 1
  when kpivalue("Matriculados_Alimentarias")<107 and
    kpivalue("Matriculados_Alimentarias")>=91 then 0
  else -1
end
```

Elaboración: los autores

Ilustración - XVI - Matriculados_Sistemas

Nombre:

Grupo de medida asociado:

⌵ Expresión de valor

```
([Measures].[Numero Postulantes],[Matricula].[Matricula].&[1],  
[Especialidad].[Carrera].&[14])
```

⌵ Expresión objetivo

⌵ Estado

Indicador de estado:

Expresión de estado:

```
case  
  when kpivalue("Matriculados_Sistemas")>= 107 then 1  
  when kpivalue("Matriculados_Sistemas")<107 and  
    kpivalue("Matriculados_Sistemas")>=91 then 0  
  else -1  
end
```

Elaboración: los autores

Ilustración - XVII - Matriculados_Industrial

Nombre:

Grupo de medida asociado:

⌵ Expresión de valor

```
([Measures].[Numero Postulantes],[Matricula].[Matricula].&[1],  
[Especialidad].[Carrera].&[17])
```

⌵ Expresión objetivo

⌵ Estado

Indicador de estado:

Expresión de estado:

```
case  
  when kpivalue("Matriculados_Industrial")>= 107 then 1  
  when kpivalue("Matriculados_Industrial")<107 and  
    kpivalue("Matriculados_Industrial")>=91 then 0  
  else -1  
end
```

Elaboración: los autores

Ilustración - XVIII - Matriculados_Electronica

Nombre:
Matriculados_Electronica

Grupo de medida asociado:
Medidas Postulantes

Expresión de valor
([Measures].[Numero Postulantes],[Matricula].[Matricula].&[1],[Especialidad].[Carrera].&[15])

Expresión objetivo

Estado
Indicador de estado: Formas

Expresión de estado:

```
case
  when kpivalue("Matriculados_Electronica")>= 107 then 1
  when kpivalue("Matriculados_Electronica")<107 and
    kpivalue("Matriculados_Electronica")>=91 then 0
  else -1
end
```

Elaboración: los autores

Ilustración - XIX - Matriculados_Civil

Nombre:
Matriculados_Civil

Grupo de medida asociado:
Medidas Postulantes

Expresión de valor
([Measures].[Numero Postulantes],[Matricula].[Matricula].&[1],[Especialidad].[Carrera].&[13])

Expresión objetivo

Estado
Indicador de estado: Formas

Expresión de estado:

```
case
  when kpivalue("Matriculados_Civil")>= 107 then 1
  when kpivalue("Matriculados_Civil")<107 and
    kpivalue("Matriculados_Civil")>=91 then 0
  else -1
end
```

Elaboración: los autores

4. Indicadores de matrículas no realizadas

Ilustración - XX - Alumnos_NO_Matriculados

Nombre:	<input type="text" value="Alumnos_NO_Matriculados"/>
Grupo de medida asociado:	<input type="text" value="Medidas Postulantes"/>
Expresión de valor	<pre>([Measures].[Numero Postulantes],[Ingreso].[Ingreso].&[1],[Matricula].[Matricula].&[2])</pre>
Expresión objetivo	<input type="text" value="100"/>
Estado	Indicador de estado: <input type="text" value="Medidor"/>
Expresión de estado:	<pre>case when kpivalue("Alumnos_NO_Matriculados")<=101 then 1 when kpivalue("Alumnos_NO_Matriculados")>101 and kpivalue("Alumnos_NO_Matriculados")<=213 then 0 else -1 end</pre>

Elaboración: los autores

Ilustración - XXI - Arquitectura_NO_Matriculados

Nombre:	<input type="text" value="Arquitectura_NO_Matriculados"/>
Grupo de medida asociado:	<input type="text" value="Medidas Postulantes"/>
Expresión de valor	<pre>([Measures].[Numero Postulantes],[Ingreso].[Ingreso].&[1],[Matricula].[Matricula].&[2],[Especialidad].[Carrera].&[11])</pre>
Expresión objetivo	<input type="text"/>
Estado	Indicador de estado: <input type="text" value="Señales viales"/>
Expresión de estado:	<pre>case when kpivalue("Arquitectura_NO_Matriculados")<=14 then 1 when kpivalue("Arquitectura_NO_Matriculados")>14 and kpivalue("Arquitectura_NO_Matriculados")<=30 then 0 else -1 end</pre>

Elaboración: los autores

Ilustración - XXII - Aeronauticas_NO_Matriculados

Nombre:

Grupo de medida asociado:

Expresión de valor

Expresión objetivo

Estado
Indicador de estado:

Expresión de estado:

```
case
when kpivalue("Aeronauticas_NO_Matriculados")<=14 then 1
when kpivalue("Aeronauticas_NO_Matriculados")>14 and
kpivalue("Aeronauticas_NO_Matriculados")<=30 then 0
else -1
end
```

Elaboración: los autores

Ilustración - XXIII - Civil_NO_Matriculados

Nombre:

Grupo de medida asociado:

Expresión de valor

Expresión objetivo

Estado
Indicador de estado:

Expresión de estado:

```
case
when kpivalue("Civil_NO_Matriculados")<=15 then 1
when kpivalue("Civil_NO_Matriculados")>15 and
kpivalue("Civil_NO_Matriculados")<=31 then 0
else -1
end
```

Elaboración: los autores

Ilustración - XXIV - Sistemas_NO_Matriculados

Nombre:

Grupo de medida asociado:

Expresión de valor

Expresión objetivo

Estado
Indicador de estado:

Expresión de estado:

```
case
when kpivalue ("Sistemas_NO_Matriculados") <=15 then 1
when kpivalue ("Sistemas_NO_Matriculados") >15 and
kpivalue ("Sistemas_NO_Matriculados") <=31 then 0
else -1
end
```

Elaboración: los autores

Ilustración - XXV - Alimentarias_NO_Matriculados

Nombre:

Grupo de medida asociado:

Expresión de valor

Expresión objetivo

Estado
Indicador de estado:

Expresión de estado:

```
case
when kpivalue ("Alimentarias_NO_Matriculados") <=14 then 1
when kpivalue ("Alimentarias_NO_Matriculados") >14 and
kpivalue ("Alimentarias_NO_Matriculados") <=30 then 0
else -1
end
```

Elaboración: los autores

Ilustración - XXVI - Electronica_NO_Matriculados

Nombre:

Grupo de medida asociado:

Expresión de valor

Expresión objetivo

Estado
Indicador de estado:

Expresión de estado:

```
case
  when kpivalue ("Electronica_NO_Matriculados")<=14 then 1
  when kpivalue ("Electronica_NO_Matriculados")>14 and
    kpivalue ("Electronica_NO_Matriculados")<=40 then 0
  else -1
end
```

Elaboración: los autores

Ilustración - XXVII - Industrial_NO_Matriculados

Nombre:

Grupo de medida asociado:

Expresión de valor

Expresión objetivo

Estado
Indicador de estado:

Expresión de estado:

```
case
  when kpivalue ("Industrial_NO_Matriculados")<=14 then 1
  when kpivalue ("Industrial_NO_Matriculados")>14 and
    kpivalue ("Industrial_NO_Matriculados")<=30 then 0
  else -1
end
```

Elaboración: los autores

ANEXO 18
REPORTES PRE-DEFINIDOS

ÍNDICE DE REPORTES

	Página
1. Indicador de eficiencia de matrícula	294
2. Indicador de alumnos matriculados	295
3. Indicador de alumnos no matriculados	296
4. Porcentaje de ingresantes matriculados y no matriculados por facultad	297
5. Porcentaje de ingresantes matriculados y no matriculados por carreras	298
6. Cantidad y distribución de porcentajes de alumnos matriculados por carrera	299
7. Porcentaje de matriculados y no matriculados por carreras	300
8. Distribución geográfica de postulantes (nacional)	301
9. Distribución geográfica de postulantes (Lima)	302
10. Rangos de mérito vs rangos de puntaje	304

ÍNDICE DE ILUSTRACIONES

	Página
Ilustración I - Eficiencia de matricula	294
Ilustración II – Alumnos matriculados	295
Ilustración III – Alumnos no matriculados	296
Ilustración IV – % de ingresantes mat. y no mat. x facultad	297
Ilustración V – % de ingresantes mat. y no mat. x carreras	298
Ilustración VI – Cant. y dist. de % de alumnos mat. por carrera	299
Ilustración VII – % de mat. y no mat. por carreras	300
Ilustración VIII – Distribución geográfica de postulantes (nacional)	301
Ilustración IX – Distribución geográfica de postulantes (Lima)	302
Ilustración X – Rangos de mérito vs rangos de puntaje	304

1. Indicador de eficiencia de matrícula

Ilustración I - Eficiencia de matrícula

EFICIENCIA DE MATRICULA		
AÑO	2014	
SEMESTRE	1	
Eficiencia_Matricula	Objetivo Eficiencia_Matricula	Estado Eficiencia_Matricula FIA
86%	85%	↑
EFICIENCIA DE MATRICULA	Arquitectura	Estado Eficiencia_Matricula_Arquitectura
	82%	▲
	Ciencias Aeronauticas	Estado Eficiencia_Matricula_Aeronauticas
	90%	●
	Ing. Civil	Estado Eficiencia_Matricula_Civil
	87%	●
	Ing. Sistemas	Estado Eficiencia_Matricula_Sistemas
	88%	●
Ing. Electronica	Estado Eficiencia_Matricula_Electronica	
86%	●	
Ing. Industrial	Estado Eficiencia_Matricula_Industrial	
89%	●	
Ing. Industrias Alimentarias	Estado Eficiencia_Matricula_Alimentarias	
78%	▲	

Elaboración: los autores

2. Indicador de alumnos matriculados

Ilustración II – Alumnos matriculados

ALUMNOS MATRICULADOS		
AlumnosMatriculados	Objetivo AlumnosMatriculados	Estado AlumnosMatriculados
730	752	
ALUMNOS MATRICULADOS	Arquitectura	Estado Matriculados_Arq. 
	Ciencias Aeronauticas	Estado Matriculados_Aeronaut. 
	Ing. Civil 113	Estado Matriculados_Civil 
	Ing. Sistemas 114	Estado Matriculados_Sistemas 
	Ing. Electronica 95	Estado Matriculados_Electronica 
	Ing. Industrias Alimentarias	Estado Matriculados_Aliment. 
	Ing. Industrial 116	Estado Matriculados_Industrial 

Elaboración: los autores

3. Indicador de alumnos no matriculados

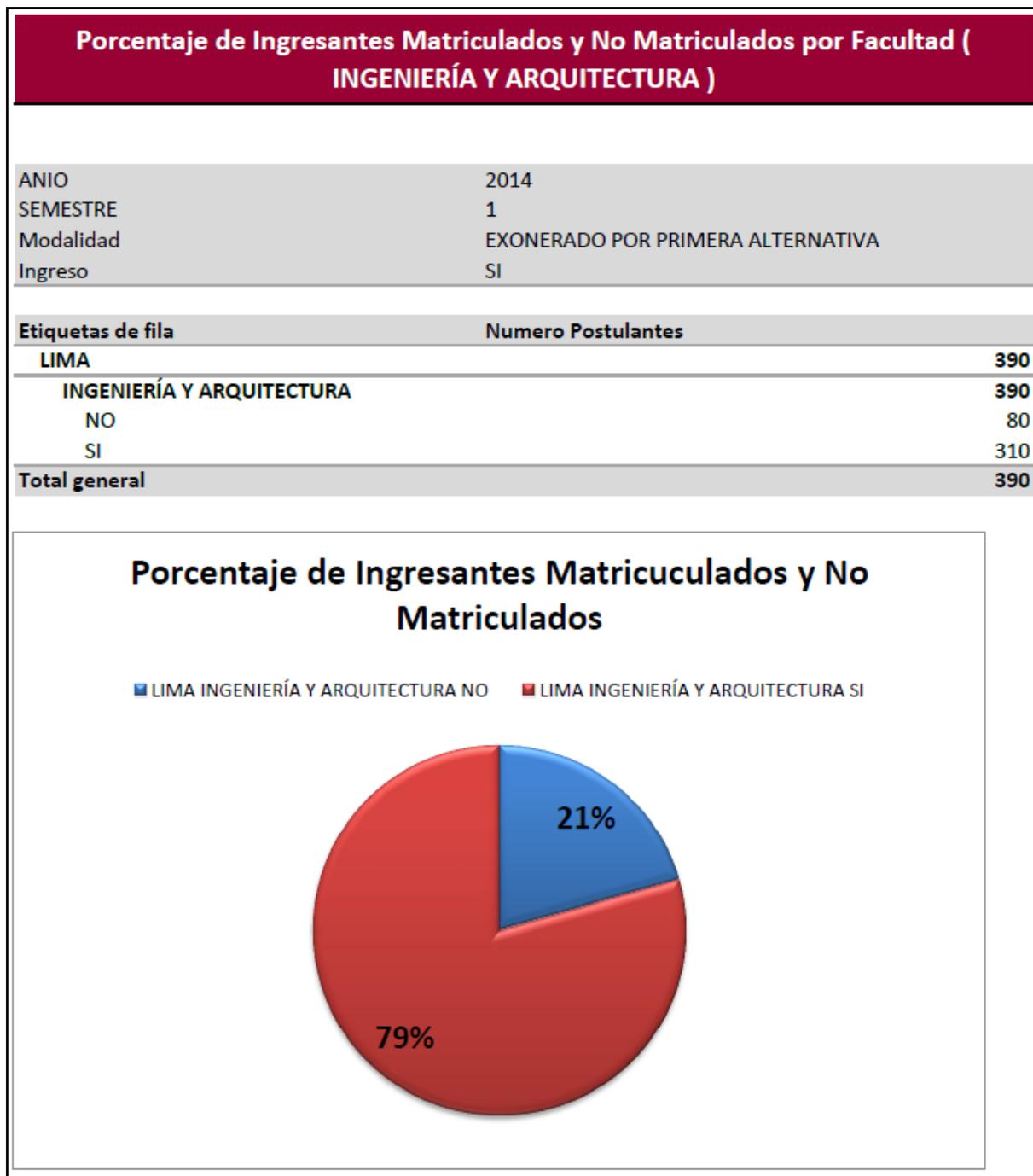
Ilustración III – Alumnos no matriculados

ALUMNOS NO MATRICULADOS			
Alumnos_NO_Matriculados	Objetivo Alumnos_NO_Matriculados	Estado	Alumnos_NO_Matriculados
120	100		
ALUMNOS NO MATRICULADOS	Arquitectura_NO_Matriculados	Estado Arquitectura_NO_Matriculados	
	23		
	Aeronauticas_NO_Matriculados	Estado Aeronauticas_NO_Matriculados	
	11		
	Civil_NO_Matriculados	Estado Civil_NO_Matriculados	
	17		
	Sistemas_NO_Matriculados	Estado Sistemas_NO_Matriculados	
	16		
Electronica_NO_Matriculados	Estado Electronica_NO_Matriculados		
15			
Alimentarias_NO_Matriculados	Estado Alimentarias_NO_Matriculados		
24			
Industrial_NO_Matriculados	Estado Industrial_NO_Matriculados		
14			

Elaboración: los autores

4. Porcentaje de ingresantes matriculados y no matriculados por facultad

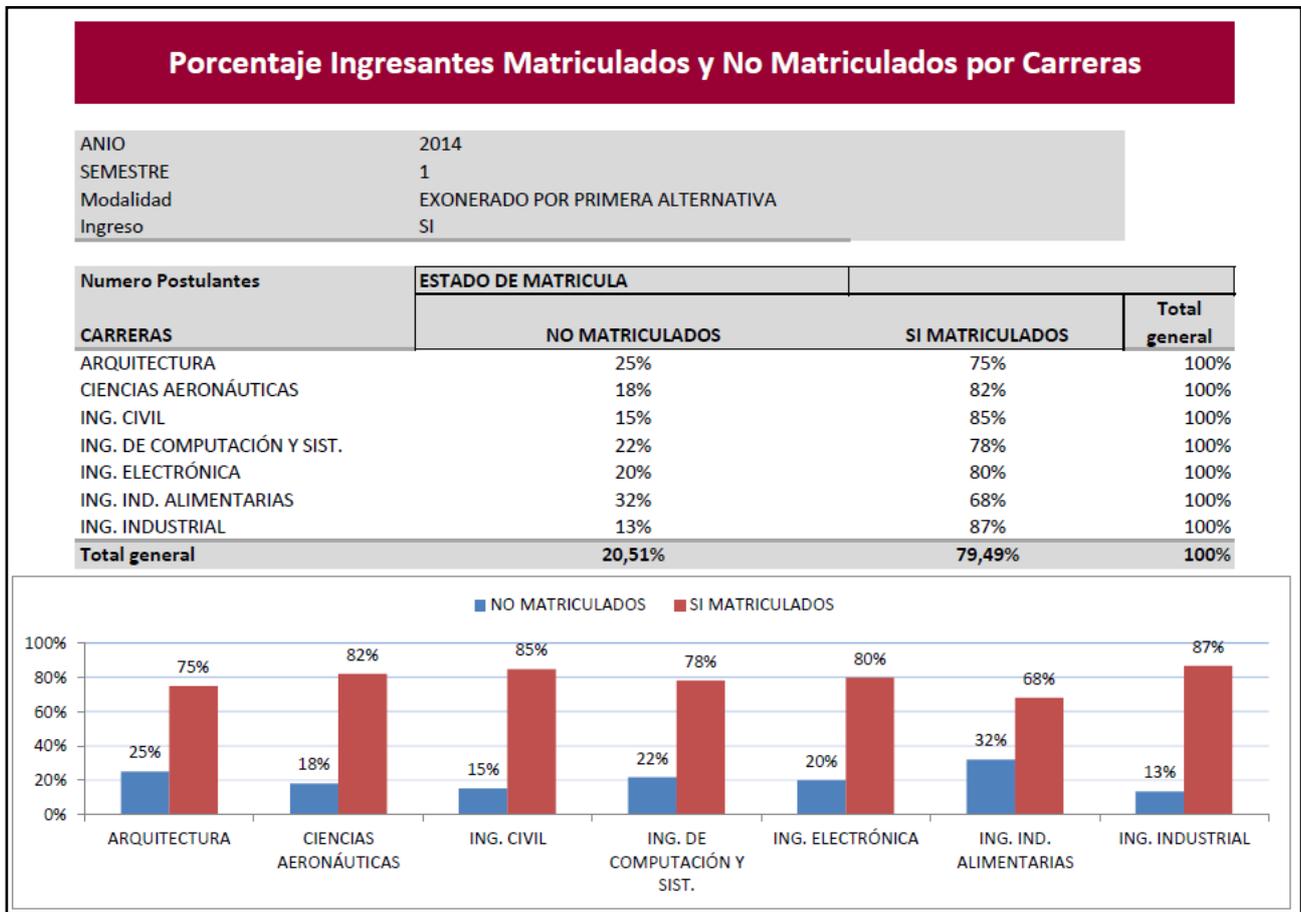
Ilustración IV – % de ingresantes mat. y no mat. x facultad



Elaboración: los autores

5. Porcentaje de ingresantes matriculados y no matriculados por carreras

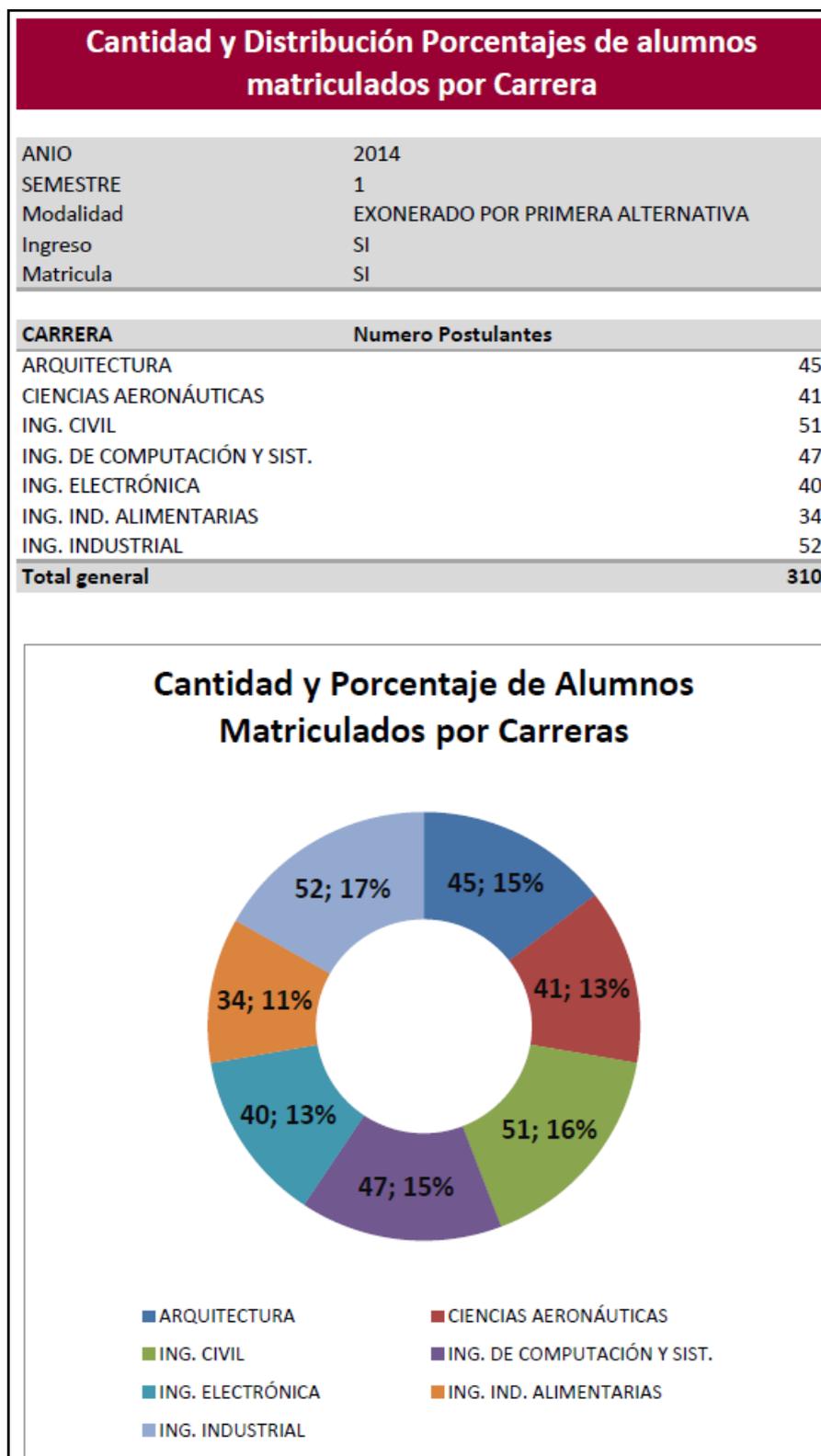
Ilustración V – % de ingresantes mat. y no mat. x carreras



Elaboración: los autores

6. Cantidad y distribución de porcentajes de alumnos matriculados por carrera

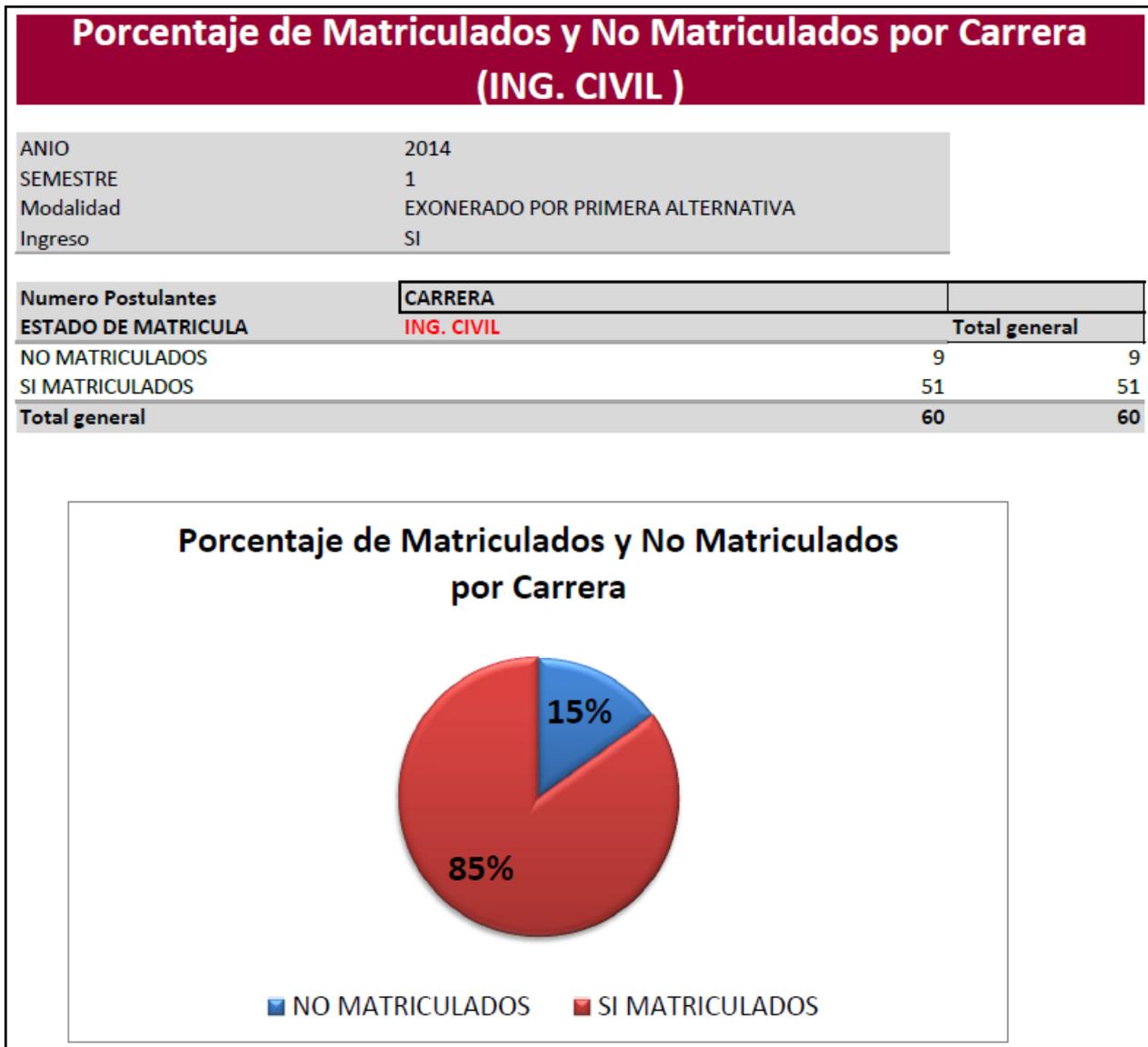
Ilustración VI – Cant. y dist. de % de alumnos mat. por carrera



Elaboración: los autores

7. Porcentaje de matriculados y no matriculados por carreras

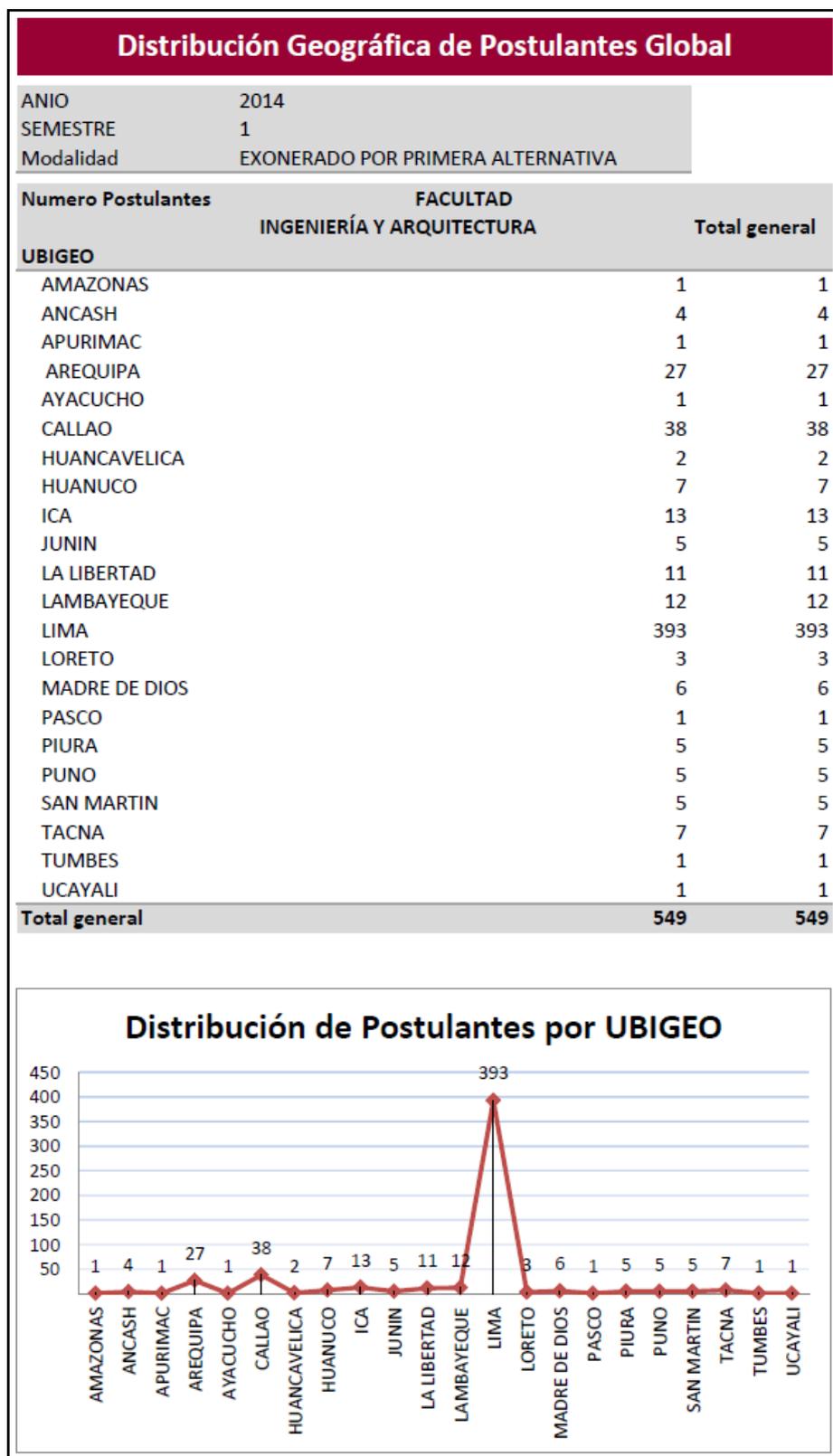
Ilustración VII – % de mat. y no mat. por carreras



Elaboración: los autores

8. Distribución geográfica de postulantes (nacional)

Ilustración VIII – Distribución geográfica de postulantes (nacional)



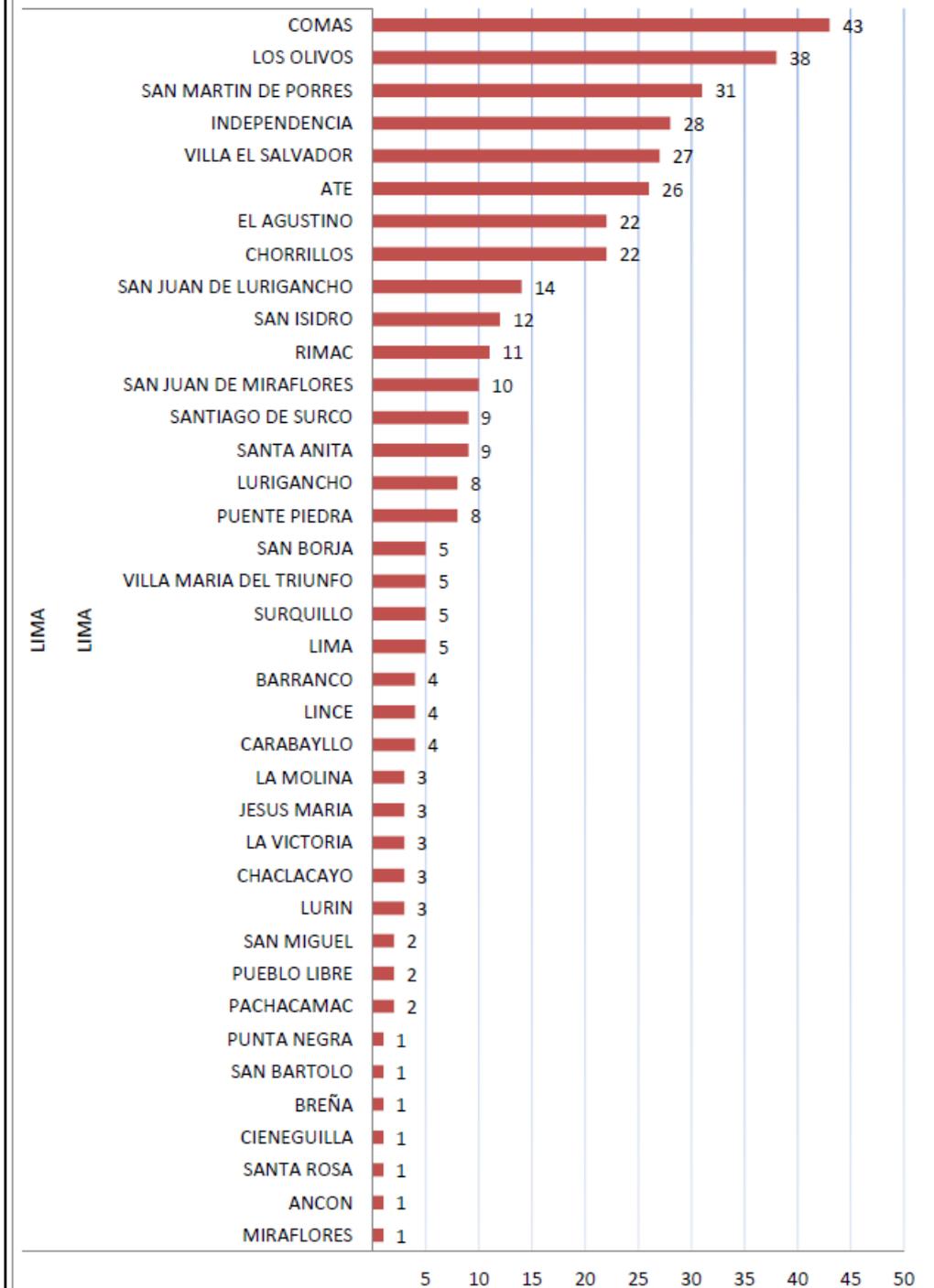
Elaboración: los autores

9. Distribución geográfica de postulantes (Lima)

Ilustración IX – Distribución geográfica de postulantes (Lima)

Distribución Geográfica de Postulantes Lima			
ANIO	2014		
SEMESTRE	1		
Modalidad	EXONERADO POR PRIMERA ALTERNATIVA		
Numero Postulantes	FACULTAD		
UBIGEO	INGENIERÍA Y ARQUITECTURA		Total general
LIMA		378	378
LIMA		378	378
MIRAFLORES		1	1
ANCON		1	1
SANTA ROSA		1	1
CIENEGUILLA		1	1
BREÑA		1	1
SAN BARTOLO		1	1
PUNTA NEGRA		1	1
PACHACAMAC		2	2
PUEBLO LIBRE		2	2
SAN MIGUEL		2	2
LURIN		3	3
CHACLACAYO		3	3
LA VICTORIA		3	3
JESUS MARIA		3	3
LA MOLINA		3	3
CARABAYLLO		4	4
LINCE		4	4
BARRANCO		4	4
LIMA		5	5
SURQUILLO		5	5
VILLA MARIA DEL TRIUNFO		5	5
SAN BORJA		5	5
PUENTE PIEDRA		8	8
LURIGANCHO		8	8
SANTA ANITA		9	9
SANTIAGO DE SURCO		9	9
SAN JUAN DE MIRAFLORES		10	10
RIMAC		11	11
SAN ISIDRO		12	12
SAN JUAN DE LURIGANCHO		14	14
CHORRILLOS		22	22
EL AGUSTINO		22	22
ATE		26	26
VILLA EL SALVADOR		27	27
INDEPENDENCIA		28	28
SAN MARTIN DE PORRES		31	31
LOS OLIVOS		38	38
COMAS		43	43
Total general		378	378

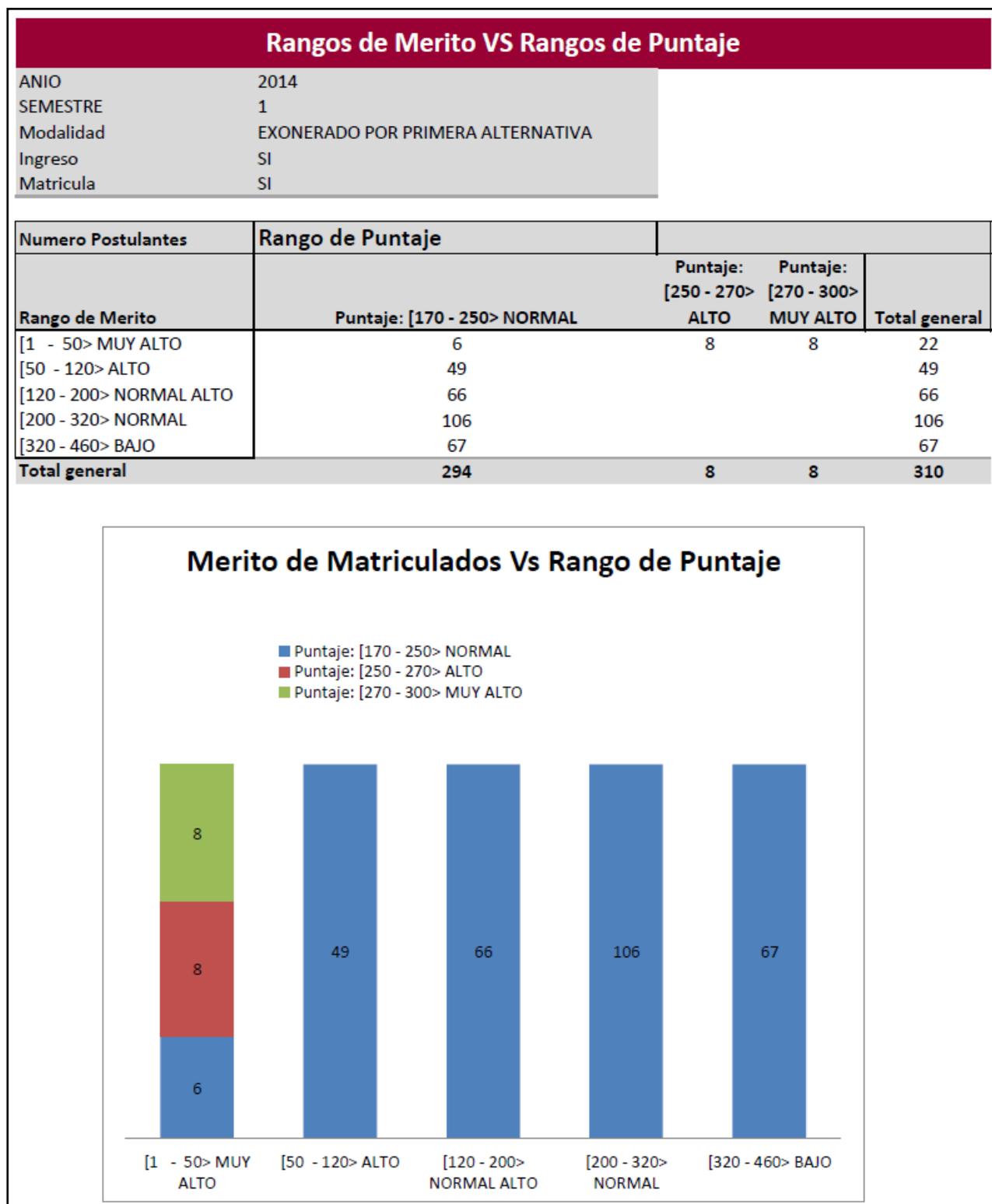
DISTRIBUCIÓN POSTULANTES - LIMA UBIGEO



Elaboración: los autores

10. Rangos de mérito vs rangos de puntaje

Ilustración X – Rangos de mérito vs rangos de puntaje



Elaboración: los autores