UNIVERSIDAD MAYOR DE SAN ANDRÉS
FACULTAD DE CIENCIAS PURAS Y NATURALES
CARRERA DE INFORMÁTICA

PROYECTO DE GRADO

"PLATAFORMA DE INFORMACIÓN ESTADÍSTICA SOCIOECONÓMICA Y ANTECEDENTES UNIVERSITARIOS APLICANDO INTELIGENCIA DE NEGOCIOS"
CASO: UNIVERSIDAD MAYOR DE SAN ANDRÉS

PARA OPTAR AL TITULO DE LICENCIATURA EN INFORMÁTICA
MENCIÓN: INGENIERÍA DE SISTEMAS INFORMÁTICOS

POSTULANTE: DANIELA VERY SANCHEZ

TUTOR METODOLÓGICO: LIC. JAVIER HUGO REYES PACHECO

ASESOR: LIC. ROBERTO VARGAS BLACUTT

LA PAZ – BOLIVIA
2015

UNIVERSIDAD MAYOR DE SAN ANDRÉS
FACULTAD DE CIENCIAS PURAS Y NATURALES
CARRERA DE INFORMÁTICA

LA CARRERA DE INFORMÁTICA DE LA FACULTAD DE CIENCIAS PURAS Y NATURALES PERTENECIENTE A LA UNIVERSIDAD MAYOR DE SAN ANDRÉS AUTORIZA EL USO DE LA INFORMACIÓN CONTENIDA EN ESTE DOCUMENTO SI LOS PROPÓSITOS SON ESTRICTAMENTE ACADÉMICOS.

LICENCIA DE USO

El usuario está autorizado a:

a) visualizar el documento mediante el uso de un ordenador o dispositivo móvil.
b) copiar, almacenar o imprimir si ha de ser de uso exclusivamente personal y privado.
c) copiar textualmente parte(s) de su contenido mencionando la fuente y/o haciendo la referencia correspondiente respetando normas de redacción e investigación.

El usuario no puede publicar, distribuir o realizar emisión o exhibición alguna de este material, sin la autorización correspondiente.

TODOS LOS DERECHOS RESERVADOS. EL USO NO AUTORIZADO DE LOS CONTENIDOS PUBLICADOS EN ESTE SITIO DERIVARA EN EL INICIO DE ACCIONES LEGALES CONTEMPLADOS EN LA LEY DE DERECHOS DE AUTOR.

Dedicatoria:

El presente proyecto va dedicado a Dios por todas las bendiciones que me obsequio, a mis papas Daniel Very, Albertina Sanchez, mis hermanas Mildred, Mayda y Patricia, mis Sobrinos Jhojan, Dorian, Tiago, Mateo, mi abuela Marcia Aranda, mi tia Elisabeth Landa, mi prima Marcia y mi cuñado Nivardo, quienes me apoyaron incondicionalmente en mi vida universitaria, brindándome su confianza y concejos, enseñándome a salir adelante y sobrepasar cualquier obstáculo hasta lograr mis sueños.

AGRADECIMIENTOS

Deseo expresar mi agradecimiento en primer lugar a mi querida familia, pues siempre expresaron la confianza que tenían en que yo culminaría la carrera y me alentaron a hacerlo.

A mi Tutor Lic. Javier Hugo Reyes Pacheco, por confiar en mi persona y en este proyecto, dando sugerencias y enseñanzas, guía y apoyo constante a mi tarea.

A mi Asesor Lic. Roberto Vargas Blacutt, por su disponibilidad en el proceso de revisión del proyecto, por las correcciones, observaciones y sugerencias durante el transcurso del presente proyecto.

Al Lic. Nicanor Delgado, por darme la oportunidad de aprender y trabajar en el desarrollo del presente proyecto.

Al Lic. Porfirio Mamani, por la gran colaboración en conocimientos en cuanto al desarrollo de este tipo de proyectos, guía y apoyo constante para seguir adelante en mi proyecto.

Finalmente agradezco a mis mejores amigos y amigas quienes me apoyaron moralmente para seguir adelante, brindándome su amistad incondicional.

Gracias a todos…!!!

RESUMEN

El presente proyecto de grado se ha desarrollado en la División de Sistemas de Información y Estadística, unidad dependiente de la Universidad Mayor de san Andrés, se pudo evidenciar que la información que procesa dicha unidad es de gran magnitud y de vital importancia para la gestión académica universitaria.

Se logró realizar el análisis, diseño e implementación de una herramienta de Inteligencia de Negocios, que concentra información histórica referente a datos sociales, económicos y académicos de los estudiantes de la Universidad Mayor de San Andrés desde la gestión 2010 hasta la 2015, con el fin de apoyar a la elaboración de políticas y proyectos con relación a la gestión académica.

La metodología empleada es The SAS Rapid Warehouse Methodology la cual establece etapas de construcción para Data Warehouse, las herramientas que se utilizó para la implantación de este tipo de solución son: Data Warehouse Builder, Oracle BI Administration Tool, Enterprise Manager Fusion Middleware y Analytics.

La herramienta implementa procesos de extracción transformación y carga de datos de 2 fuentes e datos, una aplicación informática OK y un sistema transaccional Sistema de Registro Universitario. El almacenamiento de esos datos en un Data Warehouse, la agregación de la información en Cubos OLAP, la implementación de reportes personalizados según requerimiento de los usurarios administradores de información estadística.

ABSTRACT

This graduation project has been developed in the Division of Information Systems and Statistics, part of the University of San Andrés unity was evident that the information processing unit that is large and vital to the academic management university.

It was possible to perform the analysis, design and implementation of a tool for Business Intelligence, which focuses historical information on social, economic and academic students of the Universidad Mayor de San Andrés data from management 2010 to 2015, to support the development of policies and projects concerning academic management.

The methodology used is The SAS Rapid Warehouse Methodology which establishes stages of construction for Data Warehouse, the tools used to implement this type of solution are: Data Warehouse Builder, Oracle BI Administration Tool, Enterprise Manager Fusion Middleware and Analytics.

The tool implements transformation processes of data extraction and loading of 2 sources and data, a computer application system OK and transactional University System Registry. The storage of such data in a data warehouse, the aggregation of information in OLAP cubes, implementation of customized reports as required by the usurious management of statistical information.

Contenido

CAPITULO I 2
MARCO INTRODUCTORIO 2
1.1 Introducción 2
1.2 Antecedentes 3
1.2.1 Universidad Mayor de San Andrés (U.M.S.A.) 3
1.2.2 División de Sistemas de Información y Estadística 4
1.2.3 División de Gestiones, Admisiones y Registros 4
1.2.4 Departamento de Bienestar Social 5
1.2.5 Trabajos Académicos Similares 6
1.3 Planteamiento del problema 6
1.4 Objetivos 8
1.4.1 Objetivo General 8
1.4.2 Objetivos Específicos 8
1.5 Justificación 9
1.5.1 Justificación Social 9
1.5.2 Justificación Económica 10
1.5.3 Justificación Científica 10
1.6 Técnicas y Herramientas 12
1.7 Límites y Alcances 12
CAPITULO II 13
MARCO TEÓRICO 13
2.1 Inteligencia de Negocios 13
2.2 Data Warehouse 14
2.3 Metadatos 15
2.4 Esquema relacional de una Data Warehouse o Data Mart 16
2.4.1 Esquema en Estrella 16
2.5 OLTP (Online Transactional Processing) 18
2.6 ETL (Extract, Transform and Load) 19
2.6.1 Extracción de Datos 19

2.6.2 Transformación de Datos 20
2.6.3 Cargado de Datos 20
2.7 Modelo multidimensional OLAP 21
2.7.1 OLAP (On Line Analytical Processing) 21
2.7.2 Cubo 22
2.7.3 Tabla Hecho (Tabla Fact) 24
2.7.4 Tabla Dimensión (Tabla Lock-up) 25
2.8 Metodología de Desarrollo 25
2.8.1 Rapid Warehousing Methodology 26
2.9 Lenguaje Unificado de modelos para Inteligencia de Negocios 29
2.10 Norma ISO/IEC 25012 para Calidad de Datos 30
2.11 Herramientas de Desarrollo 33
2.11.1 Oracle Warehouse Builder 33
2.11.2 Oracle BI Administration Tool 33
2.11.3 Enterprise Manager Fusion Middleware 34
2.11.4 Analytics 35
CAPITULO III 36
MARCO METODOLÓGICO 36
3.1 Construcción de Metadatos 36
3.1.1 Estructura de Metadato del Modelo de Data Warehouse 37
3.1.2 Estructura de Metadatos Tablas Hechos 37
3.1.3 Estructura de Metadatos Tablas Dimensiones 38
3.2 Construcción de la Base de Datos del Data Warehouse 39
3.3 Construcción de Cubos 41
CAPITULO IV 42
MARCO APLICATIVO 42
4.1 Definición de Objetivos 42
4.2 Definición de Requerimientos de Información 42
4.2.1 Requerimientos de los Usuarios 42
4.2.2 Requerimientos de Información 44
4.2.3 Análisis de los Orígenes de Datos 47

4.3 Diseño y Modelación 49
4.3.1 Arquitectura de la Plataforma Estadística 49
4.3.2 Capa Origen de Datos 50
4.3.3 Capa de Integración de Datos 53
4.3.4 Capa de Almacén de Datos 55
4.3.5 Capa de Agregación de Datos 62
4.3.6 Capa de Presentación de la Información 64
4.4 Implementación 65
4.4.1 Implementación Capa de Integración de Datos 65
4.4.2 Implementación de Capa de Agregación de Datos 70
4.4.3 Implementación Capa de Presentación de Datos 71
4.5 Revisión 75
4.5.1 Análisis Calidad de Datos 75
4.5.2 Procedimientos de Pruebas 78
4.5.3 Cumplimiento de Requerimientos 79
CAPITULO V 82
CONCLUSIONES Y RECOMENDACIONES 82
5.1 Conclusiones 82
5.2 Recomendaciones 83
BIBLIOGRAFÍA 84
ANEXOS 86

ÍNDICE DE FIGURAS

Figura 1: Cadena de valor de la información dentro de Business Intelligence 14

Figura 2: Definición de Data Warehouse 15

Figura 3: Ejemplo de Esquema en Estrella 17

Figura 4: Estructura de Procesos ETL 19

Figura 5: Esquema de un Cubo 23

Figura 6: Fases de la Metodología Rapid Warehousing Methodology 26

Figura 7: Proceso de ETL, utilizando la herramienta Oracle Warehouse Builder 33

Figura 8: Arquitectura de Oracle Business Intelligence 34

Figura 9: Herramienta Analytics 35

Figura 10: Esquema de Elaboración de Metadatos 36

Figura 11: Esquema de Elaboración de Metadatos 41

Figura 12: Arquitectura del Plataforma de Información Estadística 49

Figura 13: Modelo Relacional SRU, Estudiante 50

Figura 14: Modelo Relacional SRU, Persona 51

Figura 15: Estructura de Datos en la Aplicación OK 52

Figura 16: Proceso de Extracción, Transformación y Carga de datos (ETL) 53

Figura 17: Script del proceso de ETL, herramienta OWB 54

Figura 18: Proceso de Construcción del Esquema Estrella del Data Warehouse 59

Figura 19: Modelo en Estrella H_Datos_Socioeconomicos 60

Figura 20: Modelo en Estrella H_Seguro_Universitario 60

Figura 21: Modelo en Estrella H_Vida_Historica 61

Figura 22: Modelo en Estrella H_Postulantes_Titulo 61

Figura 23: Hecho H_Datos_Socioeconomicos 62

Figura 24: Hecho H_Vida_Historica 63
Figura 25: Cubo OLAP de datos Socioeconomicos 63
Figura 26: Cubo OLAP de datos Antecedentes Académicos 64
Figura 27: Interacción de los usuarios con la plataforma estadística 64
Figura 28: Arquitectura Física del Almacén de Datos 65
Figura 29: Consulta Sql para la extracción de datos 66
Figura 30: Tablas hecho y dimensiones construidas a partir de consultas 67
Figura 31: Mapeo de Transformación del Hecho Datos Socioeconomicos 67
Figura 32: Código de Transformación 68
Figura 33: Código de procedimientos de transformación 69
Figura 34: Mapeo para Carga de datos al hecho Datos Socioeconomicos 70
Figura 35: Agregación de medidas en las tablas hecho 71
Figura 36: Página principal del Data Warehouse 72
Figura 37: Indicador del crecimiento poblacional estudiantil 72
Figura 38: Indicadores de estudiantes matriculados por carrera 73
Figura 39: Cantidad de estudiantes matriculados por Género 73
Figura 40: Reporte de titulados por modalidad 74
Figura 41: Reporte de titulados por modalidad 74

ÍNDICE DE TABLAS

Tabla 1: Estructura de Metadato del Modelo Data Warehouse 37

Tabla 2: Estructura de Metadato de Hechos 38

Tabla 3: Estructura de Metadato de Dimensiones 39

Tabla 4: Matriz de Requerimientos de Datos Socioeconomicos 44

Tabla 5: Matriz de Requerimientos de Datos Socioeconomicos 45

Tabla 6: Matriz de Requerimientos de Datos Vida Histórica Académica 46

Tabla 7: Diccionario de Dim_Persona 55

Tabla 8: Diccionario de Dim_Estudiante 55

Tabla 9: Diccionario de H_Datos_Socioeconomicos 56

Tabla 10: Diccionario de H_Vida_Historica 57

Tabla 11: Diccionario de H_Seguro_Universitario 58

Tabla 12: Diccionario de H_Postulantes_Titulo 58

Tabla 13: Permisos a Usuarios 71

Tabla 14: Calculo de métricas NA (t) 75

Tabla 15: Calculo de métricas NR (t) 76

Tabla 16: Calculo de métricas NVN (t) 76

Tabla 17: Calculo de métricas NVE (t) 76

Tabla 18: Caso de Prueba de ETL 78

Tabla 19: Caso de Prueba cubo OLAP 79

Tabla 20: Cumplimiento de requerimientos de información 80

CAPITULO I

MARCO INTRODUCTORIO

1.1 Introducción

En la actualidad la Estadística es probablemente uno de los métodos científicos más utilizado en todos los campos del conocimiento humano, puede responder a muchas necesidades que la sociedad actual nos plantea, es así que se ocupa de la recolección, análisis e interpretación de datos, con el objetivo de representar la realidad y predecir su futuro, en este sentido los datos que son la materia prima para el análisis estadístico deben ser significativos y representativos, generando información que permita la construcción de indicadores estadísticos relevantes para medir el desarrollo de una determinada organización o institución tanto públicas como privadas.

Es por ello que observando el gran crecimiento de datos en las empresas e instituciones, la cual conlleva también al incremento de volúmenes de información, siendo estos activos de vital importancia para el análisis estadístico se presenta la necesidad de realizar una mejor administración e interpretación de este conjunto de datos, coadyuvando así al apoyo en el proceso de cumplimiento de objetivos estratégicos de una organización.

En este sentido la tecnología de información que da soporte al manejo, administración, gestión y control de la información como un instrumento estratégico, es el área de Inteligencia de Negocios (Business Intelligence) permitiendo el análisis avanzado de los datos desde una perspectiva histórica, vigente y predictiva, de esta manera descubriendo información a través de modelos, esquemas y aplicaciones. Un componente importante dentro de inteligencia de negocios es la herramienta Data Warehouse que consiste en almacenar una réplica masiva de los datos disponibles en un SGBDR (Sistemas Gestores de Bases de Datos Relacionales) el cual mediante procesos de ETL (Extracción, Transformación y Carga) generan metadatos, estos sirven de insumo para el data warehouse de donde se extraen las consultas para posterior toma de decisiones.

Sin embargo la institución pública no es ajena a la necesidad de este tipo de tecnologías, la Universidad Mayor de San Andrés (U.M.S.A.) genera grandes volúmenes de datos relacionados con las características sociales y académicas de los estudiantes universitarios, por lo indicado se necesita el desarrollo de herramientas basadas en Inteligencia de Negocios que ayuden al cumplimiento de objetivos y metas, con respecto a la gestión académica universitaria.

Por lo anteriormente descrito este proyecto de grado propone a través de la tecnología inteligencia de negocios, la elaboración e implementación de una herramienta Data Warehouse que ayude al procesamiento de información estadística en base a características ya mencionadas, permitiendo así que la División de Sistemas de información y Estadística proporcione datos estadísticos confiables y fidedignos al departamento de Bienestar Social, División de Gestiones, Admisiones y Registros, y otras unidades dependientes de la U.M.S.A., las cuales requieren información para realizar un amplio análisis e interpretación sobre la situación actual de la población estudiantil universitaria, como consecuencia a esto ejecutar proyectos y políticas para una mejor gestión académica.

1.2 Antecedentes

1.2.1 Universidad Mayor de San Andrés (U.M.S.A.)

- **Misión Institucional**

  La Universidad Mayor de San Andrés, es una institución autónoma, responsable de la generación y difusión de conocimiento científico, en la formación de profesionales idóneos de reconocida calidad y la revalorización de conocimientos ancestrales para la construcción de una sociedad justa, desarrollada, productiva, inclusiva y competitiva a nivel local, regional y nacional.

- **Visión Institucional**

La Universidad Mayor de San Andrés es una Institución estratégica y autónoma con reconocimiento internacional. Líder nacional en la generación del conocimiento técnico-científico y la formación de profesionales idóneos con excelencia académica.

La U.M.S.A. cuenta con una estructura organizacional que privilegia la gestión académica científica eficiente, fortalecida con alianzas estratégicas interinstitucionales.

1.2.2 División de Sistemas de Información y Estadística

La División de Sistemas de Información y Estadística es dependiente del Departamento de Tecnologías de Información y Comunicación de la U.M.S.A., la cual mantiene relaciones interinstitucionales.

Gestiona el procesamiento de información estadística de la población estudiantil, sobre cuya base serán desarrollados indicadores estadísticos.

Desarrolla e implementa sistemas para la administración del registro universitario y seguimiento académico de los estudiantes.

1.2.3 División de Gestiones, Admisiones y Registros

- **Misión de la Unidad**

  Coadyuvar en el cumplimiento de la actividad académica de la Universidad Mayor de San Andrés cumpliendo con todas las políticas y la gestión eficaz y eficiente de disposiciones referidas al Régimen Académico Estudiantil, viabilizando procesos y trámites inherentes a este ámbito.

- **Visión de la Unidad**

  Brindar servicios con calidad y eficiencia administrativa, con la implementación de políticas, reglamentos y procesos automatizados integrados que garanticen la gestión y administración de los procesos, la realización de trámites con el registro, seguimiento y control oportuno.

- **Función Principal de la Unidad**

  Viabilizar procesos de Administración Académica, referidos al Régimen Estudiantil, de acuerdo a programación, requisitos, procedimientos y normas vigentes.

1.2.4 Departamento de Bienestar Social

- **Misión del Departamento**

  Contribuye al mejoramiento de la calidad de vida de estudiantes de la metrópoli, municipios y sedes rurales universitarias, ofertando beneficios y servicios sociales integrales y con calidad, mediante la planificación y ejecución de actividades dirigidos a un bienestar biopsicosocial.

- **Visión del Departamento**

  Estudiantes universitarios de la U.M.S.A. formados de manera integral en un bienestar socializado, basto y diversa cantidad de programas y proyectos sociales integrales e inclusivos, psicológicos, de becas, de salud, deportivos, recreativos, nutricionales, que permiten una calidad de vida equitativa efectiva y eficiente.

- **Objetivos del Departamento**

  Lograr un ambiente dirigido a construir un bienestar socializado y calidad de vida del estudiante universitario de la metrópoli, municipios y sedes rurales universitarias, a través de la creación y ejecución de programas y proyectos de beneficios, becas y servicios sociales integrales, salud, alimentarios, psicológicos, deportivo-recreativos y educativos, dirigidos a estudiantes de sectores empobrecidos y postergados económica y socialmente, para contar con oportunidades equitativas y desarrollo integral en el proceso de profesionalización.

  Propiciar una estructura organizativa administrativa que permita la operativización de las políticas de Bienestar Estudiantil en concordancia con las

políticas institucionales y ejes centrales de: Educación, Investigación, Administración, Gestión y de apoyo a la interacción social, dirigidos a coadyuvar el proceso de profesionalización.

- **Unidades Dependientes de Departamento**

  Las unidades dependientes del Departamento de Bienestar Social son:

  - División de Trabajo Social
  - División Salud
  - Sección de Becas Académicas
  - Sección Deportes
  - Coordinación Centros Infantiles

1.2.5 Trabajos Académicos Similares

En la carrera de Informática de la U.M.S.A., en relación al proyecto podemos mencionar los siguientes trabajos.

- Data Warehouse como apoyo a la toma de decisiones del Servicio Nacional de Impuestos Internos. (T. 555).
- Sistema de Soporte a la Toma de Decisiones respecto a la formulación y ejecución presupuestaria del Gobierno Autónomo Municipal de La Paz. (T. 2416).
- Data Warehouse en la integración de la información de apoyo a la toma de decisiones aplicada AASANA (T. 554).

1.3 Planteamiento del problema

El conocimiento de las características sociales y académicas de la población estudiantil de la Universidad Mayor de San Andrés constituyen un factor fundamental para la

gestión académica universitaria cuyo propósito es contribuir a la formulación de nuevas políticas y planificación de proyectos en el área de la educación universitaria.

Estos factores son reflejados a través de indicadores estadísticos socioeconómicos e historiales académicos universitarios, entre los factores se hallan: datos personales, características de vivienda y colegio de egreso, áreas de trabajo, datos universitarios.

Todos estos datos están desagregados por facultad y carrera los mismos existen como base informativa, sin embargo se pudo observar que no es suficiente esta clasificación de información, evitando así que se realice un amplio análisis e interpretación sobre la situación académica de los estudiantes.

Con relación a lo mencionado se realizó entrevistas con la División de Gestiones, Admisiones y Registros, y Departamento de Bienestar Social donde se pudo evidenciar la problemática de falta de información estadística confiable, oportuna y precisa, manifestando sus necesidades a obtener indicadores estadísticos más descriptivos con relación a la población estudiantil universitaria.

Actualmente la Universidad Mayor de San Andrés, tiene implementado el Sistema de Registro Universitario (S.R.U.) y la aplicación informática "OK", que permiten la gestión de los datos estadísticos, referente a la información socioeconómica y vida universitaria de los estudiantes en la U.M.S.A., sin embargo, presentan una serie de problemáticas al momento del análisis de información contenida en sus diferentes herramientas informáticas.

Entre los principales problemas identificados podemos señalar:

- Datos estadísticos imprecisos, debido a las distintas herramientas informáticas proveedoras. Cuando la información socioeconómica respecto a la población universitaria es requerida, el personal debe realizar comprobaciones de datos entre la bases de datos del S.R.U., Hojas de Cálculo Excel, Aplicación Informática "OK" y libros estadísticos para verificar la exactitud de los datos. Lo

cual ocasiona una incertidumbre en la autenticidad de dicha información y retardo en el proceso.

- La información histórica no está almacena en una estructura diseñada para la agregación histórica, se encuentra en las bases de datos del SRU, aplicación informática "OK" y en Hojas de Cálculo Excel. Lo que ocasiona dispersión y poca comprensión de los datos estadísticos.

- Complejos procesos y herramientas obsoletas para la obtención de información según el crecimiento de datos, lo que ocasiona retrasos en la toma de decisiones a autoridades de la U.M.S.A. (Rectorado, Decanato, Jefes de carrera, etc.).

Estas problemáticas ocasionan que la Universidad Mayor de San Andrés presente deficiencias y restricciones en el apoyo a la toma de decisiones con relación a la gestión académica universitaria.

1.4 Objetivos

1.4.1 Objetivo General

Implementar una plataforma tecnológica basada en inteligencia de negocios, que permita la centralización, procesamiento y administración de datos históricos con respecto a los antecedentes socioeconómicos y vida histórica universitaria, con el fin de brindar información de apoyo, permitiendo así dar solución a las necesidades estratégicas con relación a la gestión académica universitaria en la Universidad Mayor de San Andrés.

1.4.2 Objetivos Específicos

Identificando los problemas expuestos anteriormente se han establecido los siguientes objetivos específicos.

- Realizar un análisis detallado de la estructura de la base de datos relacional del Sistema de Registro Universitario y la organización de datos de la aplicación informática OK, estableciendo formas y niveles de relación entre ellos.

- Realizar un análisis de datos requeridos, según necesidades, a partir de las fuentes proveedoras de información.
- Diseñar técnicas que permitan la extracción de los datos a partir de las fuentes de información, implementando procesos de limpieza de datos para garantizar la calidad de datos, para su almacenamiento en el Data Warehouse.
- Modelar la Plataforma de Información Estadística a partir de un análisis multidimensional aplicando estándares y simultáneamente siguiendo una metodología de desarrollo de este tipo de soluciones de software.
- Diseñar una estructura para el almacenamiento de datos que permita la integración de información histórica, a partir de datos almacenados en el Sistema de Registro Universitario y herramientas informáticas usadas.
- Generar una solución que permita la agregación de la información de acuerdo a requerimientos del usuario.
- Generar reportes: cuadros y gráficos estadísticos e indicadores clave de rendimiento para el análisis de los datos contenidos en el Data Warehouse.
- Generar indicadores estadísticos en cuanto el crecimiento poblacional estudiantil con diferentes características, descripción de situación social y académica de estudiantes nuevos y antiguos, realizar las distintitas relaciones entre variables almacenadas en la base de datos de data Warehouse.

1.5 Justificación

1.5.1 Justificación Social

El presente proyecto contribuye al proceso de toma de decisiones con respecto a la gestión académica universitaria, colaborando de esta manera al personal administrador a presentar información estadística de calidad.

Esta herramienta tecnológica que se espera implementar también cumplirá una labor social importante, por lo que dependiendo del tipo de información que exponga, la Universidad Mayor de San Andrés podrá participar de convenios interinstitucionales (transporte, guardería, becas, etc.).

Las instituciones públicas y privadas de Bolivia también resultaran beneficiadas dado que se podrá presentar información requerida según sus necesidades en cuanto a datos socioeconómicos y vida histórica universitaria de la población estudiantil en la U.M.S.A.

Teniendo de esta manera una justificación social conveniente, tanto para los usuarios de la herramienta tecnológica como para los destinatarios de las acciones que se programen con los resultados.

1.5.2 Justificación Económica

El presente proyecto permitirá que la Universidad Mayor de San Andrés realice una mejor administración de recursos económicos, permitiendo estrategias de distribución presupuestaria en beneficio a cada facultad y carrera según la cantidad y característica de población estudiantil perteneciente a ellas.

1.5.3 Justificación Científica

La determinación de metodologías, técnicas y conceptos en el desarrollo de un proyecto basado en inteligencia de negocios, sin lugar a duda es de gran importancia, el empleo de las mismas nos ayudara en el proceso de recolección de información, análisis y desarrollo, cumpliendo así con todos los requerimientos de la institución.

La Inteligencia de Negocios se refiere al proceso de convertir datos en información, información en conocimientos y conocimiento en acciones estratégicas para gestionar una institución, algunas de las herramientas para implementar esta tecnología informática son: Data Warehouse, Cubos OLAP, Data Mining, Sistemas de Soporte de Decisiones, etc.

Según Ralph Kimball[1] "La construcción de un Data Warehouse es transformar datos en conocimiento para ello es necesario ensamblar datos existentes en distintas fuentes teniendo en cuenta las etapas a seguir para obtener un resultado que es un conjunto de datos analíticos que queda almacenado en un depósito especial que es el Data Warehouse o el Data Mart, una vez almacenado los datos estratégicos y tácticos están preparados para realizar un estudio para la toma de decisiones".

La metodología a emplear en este proyecto es Rapid Warehousing Methodology propuesta por SAS Institute, utilizada para el desarrollo de productos de inteligencia de negocios, la misma está enfocada al desarrollo de procesos iterativos e incrementales dividido en cinco fases:

- Definición de Objetos.
- Definición de Requerimientos de Información.
- Diseño y Modelización.
- Implementación.
- Revisión.

Se utilizara como herramienta de modelado el Lenguaje Unificado de Modelado UML y extensiones para modelado multidimensional.

Por lo mencionado el presente proyecto tiene bases teóricas reconocidas como valederas, en el entendido que "La teoría se define como un conjunto de conceptos sistemáticamente interrelacionados, definiciones y proposiciones que sirven para explicar y predecir fenómenos"[2]. Por lo tanto este proyecto es base para incrementar la investigación o análisis en los diversos sistemas operacionales, donde se puede aplicar inteligencia de negocios.

[1] Ralph Kimball: **The Data Warehouse Toolkit, Ed. John Wiley, 1998.**
[2] Naghi N.: **Mohammad, Metodología de la Investigación, Pág. 56**

1.6 Técnicas y Herramientas

Las herramientas a utilizar son propias de la corporación ORACLE, ya que la Universidad Mayor de San Andrés cuenta con las licencias respectivas de los productos. Específicamente las herramientas de desarrollo son:

- Data Warehouse Builder
- Oracle BI Administration Tool
- Enterprise Manager Fusion Middleware
- Analytics

1.7 Límites y Alcances

El presente proyecto abordara las etapas de definición de objetos, definición de requerimientos, diseño, modelado, implementación y revisión del proyecto como lo establece la metodología.

El proyecto se limitara a interactuar con el Sistema de Registro Universitario con relación a los datos y procesos inmiscuidos en el mismo.

CAPITULO II

MARCO TEÓRICO

2.1 Inteligencia de Negocios

La Inteligencia de Negocios BI (Business Intelligence) es una herramienta bajo la cual diferentes tipos de organizaciones, pueden soportar la toma de decisiones basadas en información precisa y oportuna; garantizando la generación del conocimiento necesario que permita escoger la alternativa que sea más conveniente para el éxito de la empresa.

Dentro del marco de beneficios que representa una solución de inteligencia de negocios podemos mencionar que esta nos permite:

Manejar el crecimiento.- El reto para las empresas es evolucionar, es crecer y esto significa "cambio". Que tan ágiles son mis procesos para enfrentar los cambios y las necesidades puntuales de la empresa.

Control de costos.- El manejo de costos es el detonador que fuerza a muchas empresas a considerar una solución de inteligencia de negocios, para tener la capacidad de medir gastos y ver esto a un nivel de detalle que identifique la línea de negocio, producto, centro de costo, entre otras.

Indicadores de gestión.- Los indicadores de desempeño me permiten representar medidas enfocadas al desempeño organizacional con la capacidad de representar la estrategia organizacional en objetivos, métricas, iniciativas y tareas dirigidas a un grupo y/o individuos en la organización. Dentro de las capacidades funcionales de los indicadores de gestión podemos mencionar: el monitoreo, análisis y la administración.

"Monitorea" lo procesos críticos de negocio y las actividades utilizando métricas que me den una alerta sobre un problema potencial o alguna gestión que debo realizar. "Analiza" la raíz de los problemas explorando la información de múltiples perspectivas en varios niveles de detalles. "Administra" los recursos y procesos para dirigir la toma

de decisiones, optimizar el desempeño. Esto nos permite tener una visión global de la empresa con la capacidad de dirigir la organización en la dirección correcta.

Figura 1: Cadena de valor de la información dentro de Business Intelligence
Fuente: Modelo jerárquico de Inteligencia de Negocio, Alejandro Peña Aranda

2.2 Data Warehouse

El Data Warehouse es una herramienta de información que se basa en un cúmulo de información, tanto detallada como resumida, que proviene de datos que se encuentran en bases de datos operativas y de otras fuentes externas (datos externos a la propia compañía o datos antiguos contenidos en otro tipo de soporte, etc.).

Un Data Warehouse debe tener las siguientes características:

Orientado a temas.- Los datos en la base de datos están organizados de manera que todos los elementos de datos relativos al mismo evento u objeto del mundo real queden unidos entre sí.

Variante en el tiempo.- Los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se puedan generar reflejen esas variaciones.

No volátil.- La información no se modifica ni se elimina, una vez almacenado un dato, éste se convierte en información de *sólo lectura*, y se mantiene para futuras consultas.

Integrado.- La base de datos contiene los datos de todos los sistemas operacionales de la organización, dichos datos deben ser consistentes.

Figura 2: Definición de Data Warehouse
Fuente: Internet

2.3 Metadatos

Los metadatos, son conocimientos acerca de datos, están compuestos por definiciones de los elementos de datos en el depósito, dichos metadatos son almacenados en un repositorio que maneja la herramienta. Las razones por las que son necesarios los metadatos en un Data Warehouse, son las siguientes:

Puede ser necesario involucrar diversos tipos de fuentes de datos, cada una de las cuales define sus datos en diferentes formas. El metadato provee una forma consistente para describir las estructuras de los datos.

Es importante describir los cambios que son hechos a los datos cuando están siendo transformados. Algunas herramientas permiten crear metadatos durante el proceso de transformación.

Los usuarios que utilizan los datos del Data Warehouse, necesitan tener una clara explicación del manejo de los diferentes campos, hechos, niveles y dimensiones.

Una de las partes importantes de los metadatos, es realizar una descripción entendible de los datos al usuario, tales descripciones son almacenadas en la herramienta en una base de datos conocida también como repositorio de datos.

Los metadatos, describen los datos en una forma clara y no ambigua. Esta es la clase de información que un programa de computación necesita para procesar correctamente los datos, tales como:

- Nombre de campos, tablas y bases de datos.
- Nombre de niveles o jerarquías.
- Nombre de dimensiones, cubos y bases de datos OLAP.
- Tipos de datos.
- Longitud de campos.
- Índices y relaciones.
- Llaves primarias y foráneas.

2.4 Esquema relacional de una Data Warehouse o Data Mart

El modelos multidimensional de datos se puede instrumentar por un esquema relacional, donde las dimensiones del cubo son modelados como relaciones de dimensiones, Los cubos son modelados como funciones del producto cartesiano sobre las dimensiones de las medidas de datos.

En un Data Warehouse relacional encontramos esquemas en estrella y copo de nieve que premian por encima de todas las consultas.

2.4.1 Esquema en Estrella

En las bases de datos usadas para *data warehousing*, un esquema en estrella es un modelo de datos que tiene una tabla de hechos (o *tabla fact*) que contiene los datos para el análisis, rodeada de las tablas de dimensiones. Este aspecto, de tabla de hechos (o

central) más grande rodeada de radios o tablas más pequeñas es lo que asemeja a una estrella, dándole nombre a este tipo de construcciones.

Las tablas de dimensiones tendrán siempre una clave primaria simple, mientras que en la tabla de hechos, la clave principal estará compuesta por las claves principales de las tablas dimensionales.

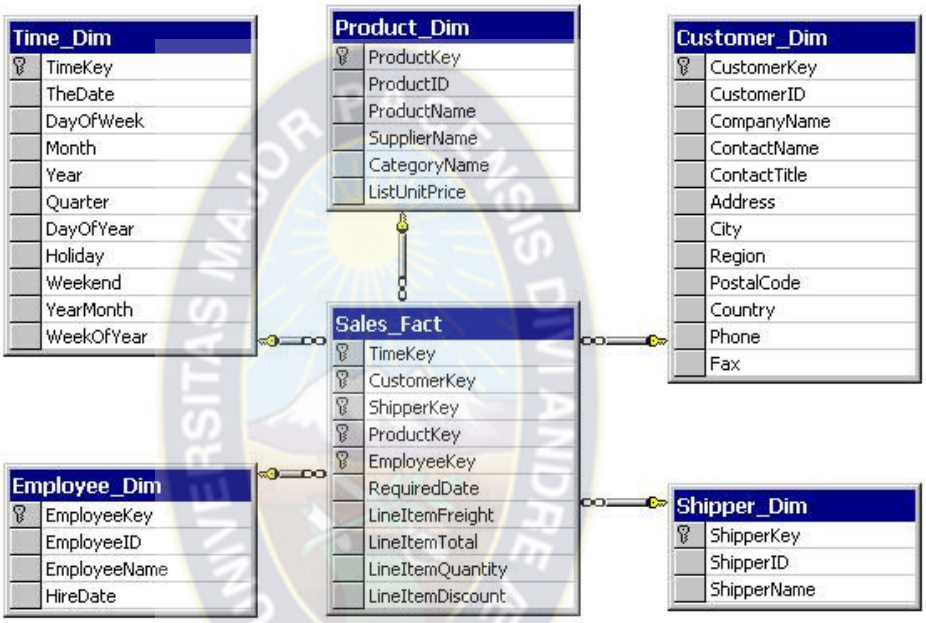


Figura 3: Ejemplo de Esquema en Estrella
Fuente: Internet

Este esquema es ideal por su simplicidad y velocidad para ser usado en análisis multidimensionales (OLAP, Data Marts, EIS). Permite acceder tanto a datos agregados como de detalle.

El diseño de esquemas en estrella permite implementar la funcionalidad de una base de datos multidimensional utilizando una clásica base de datos relacional (más extendidas que las multidimensionales).

Otra razón para utilizar los esquemas en estrella es su simplicidad desde el punto de vista del usuario final. Las consultas no son complicadas, ya que las condiciones y las uniones (*JOIN*) necesarias sólo involucran a la tabla de hechos y a las de dimensiones,

no haciendo falta que se encadenen uniones y condiciones a dos o más niveles como ocurriría en un esquema en copo de nieve. En la mayoría de los casos son preferibles los de estrellas por su simplicidad respecto a los de copo de nieve por ser más fáciles de manejar.

Finalmente, es la opción con mejor rendimiento y velocidad pues permite indexar las dimensiones de forma individualizada sin que repercuta en el rendimiento de la base de datos en su conjunto.

2.5 OLTP (Online Transactional Processing)

Los sistemas OLTP están diseñados para gestionar un gran número de peticiones concurrentes sobre sus bases de datos, y que los usuarios puedan insertar, modificar, borrar y consultar dichos datos. Están enfocados a que cada operación (*transacción*) trabaje con pequeñas cantidades de filas, y a que ofrezcan una respuesta rápida. Habitualmente utilizan sistemas de bases de datos relacionales para gestionar los datos, y suelen estar altamente normalizados. En ellos es muy importante la integridad de los datos, y deben cumplir las propiedades ACID (*Atomicity, Consistency, Isolation, Durability*):

- **Atomicidad**: una operación, o se realiza por completo o no se realiza, nunca debe quedar a medias.
- **Consistencia**: sólo se ejecutan las operaciones que cumplen las reglas de integridad de la base de datos.
- **Aislamiento** (*Isolation*): una operación no puede afectar a otras, dos transacciones sobre los mismos datos son independientes y no generan errores entre sí.
- **Durabilidad**: una vez realizada una operación, ésta es persistente y no se puede deshacer.

2.6 ETL (Extract, Transform and Load)

Es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos**,** data mart o data warehouse para analizar, o en otro sistema operacional para apoyar un proceso de negocio. En síntesis, las funciones específicas de los ETL son tres:

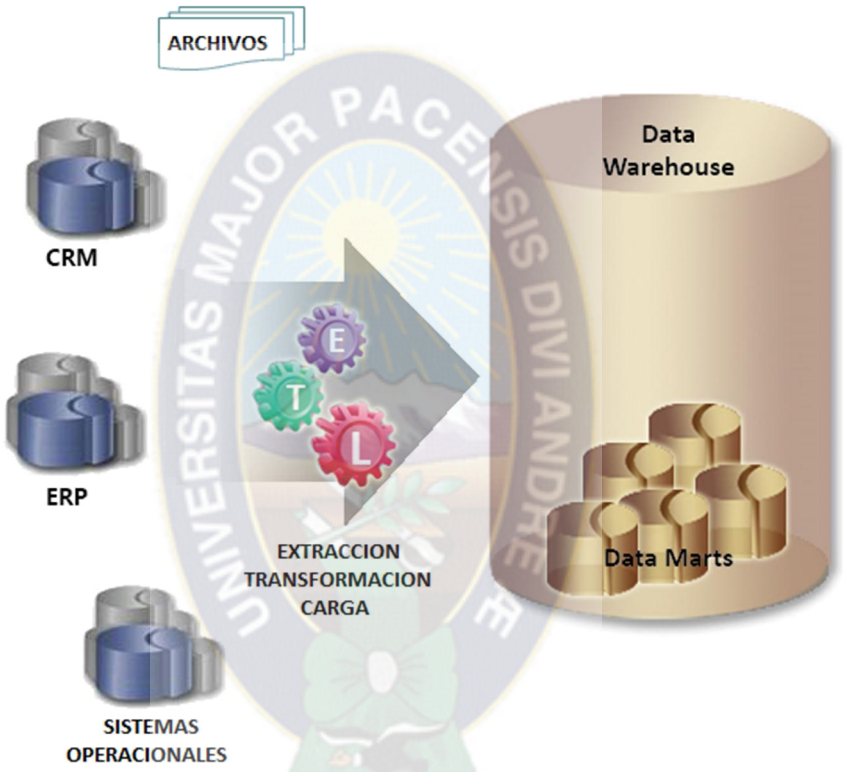


Figura 4: Estructura de Procesos ETL
Fuente: Internet

- Extracción.
- Transformación.
- Carga.

2.6.1 Extracción de Datos

Es aquí, en donde, basándose en las necesidades y requisitos del usuario, se exploran las diversas fuentes OLTP que se tengan a disposición, y se extrae la información que se considere relevante al caso. Si se encuentran en un sistema propietario o fuentes

externas, ya sean textuales, hipertextuales, hojas de cálculos, etc., la obtención de los mismos puede ser un tanto más dificultoso, debido a que, por ejemplo, se tendrán que realizar cambios de formato y/o volcado de información a partir de alguna herramienta específica. Una vez que los datos son seleccionados y extraídos, se guardan en un almacenamiento intermedio, lo cual permite, entre otras ventajas:

- Manipular los datos sin interrumpir ni paralizar los OLTP, ni tampoco el DW.
- Almacenar y gestionar los metadatos que se generarán en los procesos ETL.
- Facilitar la integración de las diversas fuentes, internas y externas.

2.6.2 Transformación de Datos

Esta función es la encargada de convertir aquellos datos inconsistentes en un conjunto de datos compatibles y congruentes, para que puedan ser cargados en el DW. Estas acciones se llevan a cabo, debido a que pueden existir diferentes fuentes de información, y es vital conciliar un formato y forma única, definiendo estándares, para que todos los datos que ingresarán al DW estén integrados. Los casos más comunes en los que se deberá realizar integración, son los siguientes:

- Codificación.
- Medida de atributos.
- Convenciones de nombramiento.
- Fuentes múltiples.

Además de lo antes mencionado, esta función se encarga de realizar el proceso de Limpieza de Datos (Data Cleansing).

2.6.3 Cargado de Datos

Este proceso es el responsable de cargar la estructura de datos del DW con:

- Aquellos datos que han sido transformados y que residen en el almacenamiento intermedio.
- Aquellos datos de los OLTP que tienen correspondencia directa con el depósito de datos.

Se debe tener en cuenta, que los datos antes de moverse al almacén de datos, deben ser analizados con el propósito de asegurar su calidad, ya que este es un factor clave, que no debe dejarse de lado.

2.7 Modelo multidimensional OLAP

2.7.1 OLAP (On Line Analytical Processing)

El procesamiento analítico en línea OLAP (Online Analytic Processing), es la componente más poderosa de los Data Warehouse, ya que es el motor de consultas especializado de la bodega. Las herramientas OLAP, son una tecnología de software para análisis en línea, administración y ejecución de consultas, que permiten inferir información del comportamiento del negocio.

Su principal objetivo es el de brindar rápidas respuestas a complejas preguntas, para interpretar la situación del negocio y tomar decisiones. Cabe destacar que lo que es realmente interesante en OLAP, no es la ejecución de simples consultas tradicionales, sino la posibilidad de utilizar operadores tales como drill-up, drill-down, etc, para explotar profundamente la información.

Además, a través de este tipo de herramientas, se puede analizar el negocio desde diferentes escenarios históricos, y proyectar como se ha venido comportando y evolucionando en un ambiente multidimensional, o sea, mediante la combinación de diferentes perspectivas, temas de interés o dimensiones. Esto permite deducir tendencias, por medio del descubrimiento de relaciones entre las perspectivas que a simple vista no se podrían encontrar sencillamente.

Las herramientas OLAP requieren que los datos estén organizados dentro del depósito en forma multidimensional, por lo cual es que utilizan los cubos multidimensionales.

Además de las características ya descritas, se pueden enumerar las siguientes:

- Permite recolectar y organizar la información analítica necesaria para los usuarios y disponer de ella en diversos formatos, tales como tablas, gráficos, reportes, etc.
- Soporta análisis complejos de grandes volúmenes de datos.
- Complementa las actividades de otras herramientas que requieran procesamiento analítico en línea.
- Presenta al usuario una visión multidimensional de los datos (matricial) para cada tema de interés del negocio.
- Permite definir de forma flexible las dimensiones que se quieren analizar, sus restricciones, jerarquías y combinaciones.
- No tiene limitaciones con respecto al número máximo de dimensiones permitidas.
- Permite a los usuarios, analizar la información basándose en más criterios que un análisis de forma tradicional.
- Al contar con muestras grandes, se pueden explorar mejor los datos en busca de respuestas.
- Permiten realizar agregaciones y combinaciones de los datos de maneras complejas y específicas, con el fin de realizar análisis más estratégicos.

2.7.2 Cubo

El proceso analítico en línea OLAP efectúa el almacenamiento lógico de los datos en arreglos o matrices multidimensionales denominadas cubos.

El cubo contiene los datos que son requeridos por los usuarios; organiza la información dentro de dimensiones y medidas en una estructura multidimensional para soportar las consultas que tiene los usuarios acerca de los datos de su compañía. Además proporcionan un mecanismo para la consulta de datos con rapidez y con una respuesta

uniforme ante diferentes cantidades de datos que contenga el cubo o por la complejidad de una consulta.

El cubo se rige por una jerarquía que se establece según los requerimientos de información.

Figura 5: Esquema de un Cubo
Fuente: Kimball, 98

Exponiendo en términos matemáticos el concepto de cubo OLAP se puede decir que es una representación abstracta de la proyección de una relación de un RDBMS (Sistema administrador de bases de datos relacionales). Dada una relación de orden N, se considera la posibilidad de una proyección que dispone de los campos **X, Y, Z** como clave de la relación y de **W** como atributo residual. Categorizando esto como una función se tiene que:

$$\mathbf{W: (X, Y, Z) \rightarrow W}$$

Los atributos **X, Y, Z** se corresponden con los ejes del cubo, mientras que el valor de **W** devuelto por cada tripleta **(X, Y, Z)** se corresponde con el dato o elemento que se rellena en cada celda del cubo.

Debido a que los dispositivos de salida (monitores, impresoras) sólo cuentan con dos dimensiones, no pueden caracterizar fácilmente cuatro dimensiones, es más práctico proyectar "rebanadas" o secciones de los datos del cubo (se dice proyectar en el sentido clásico vector analítico de reducción dimensional, no en el sentido de SQL, aunque los dos conceptos son claramente análogos), tales como la expresión:

$$\mathbf{W: (X, Y) \rightarrow W}$$

Aunque no se conserve la clave del cubo (al faltar el parámetro **Z**), puede tener algún significado semántico, sin embargo, también puede que una sección de la representación funcional con tres parámetros para un determinado valor de **Z** también resulte de interés.

La motivación que hay tras OLAP vuelve a mostrar de nuevo el paradigma de los informes de tablas cruzadas del sistema de gestión de base de datos. Se puede desear una visualización al estilo de una hoja de cálculo, donde los valores de **X** se encuentran en la fila **$1**, los valores de **Y** aparecen en la columna **$A**, y los valores de **W: (X, Y) → W** se encuentran en las celdas individuales a partir de la celda **$B2** y desde ahí, hacia abajo y hacia la derecha. Si bien se puede utilizar el **Lenguaje de Manipulación de Datos** (o DML) de SQL para mostrar las tuplas **(X, Y, W)**, este formato de salida no es tan deseable como la alternativa de tablas cruzadas. El primer método requiere que se realice una búsqueda lineal para cada par **(X,Y)** dado, para determinar el correspondiente valor de W, mientras que el segundo permite realizar una búsqueda más convenientemente permitiendo localizar el valor W en la intersección de la columna **X** apropiada con la fila **Y** correspondiente.

2.7.3 Tabla Hecho (Tabla Fact)

Es la tabla central en un esquema dimensional. Es en ella donde se almacenan las mediciones numéricas del negocio. Estas medidas se hacen sobre el grano, o unidad básica de la tabla.

El grano o la granularidad de la tabla queda determinada por el nivel de detalle que se almacenará en la tabla. Cada medida es tomada de la intersección de las dimensiones que la definen.

Idealmente está compuesta por valores numéricos, continuamente evaluados y aditivos. La razón de estas características es que así se facilita que los miles de registros que involucran una consulta sean comprimidos en unas pocas líneas en un set de respuesta.

La clave de la tabla fact recibe el nombre de clave compuesta o concatenada debido a que se forma de la composición (o concatenación) de las llaves primarias de las tablas dimensionales a las que está unida.

Así entonces, se distinguen dos tipos de columnas en una tabla fact: columnas fact y columnas key. Donde la columna fact es la que almacena alguna medida de negocio y una columna key forma parte de la clave compuesta de la tabla.

2.7.4 Tabla Dimensión (Tabla Lock-up)

Estas tablas son las que se conectan a la tabla fact, son las que alimentan a la tabla fact. Una tabla lock_up almacena un conjunto de valores que están relacionados a una dimensión particular.

Tablas lock_up no contienen hechos, en su lugar los valores en las tablas lock_up son los elementos que determinan la estructura de las dimensiones. Así entonces, en ellas existe el detalle de los valores de la dimensión respectiva.

Una tabla lock_up está compuesta de una primary key que identifica unívocamente una fila en la tabla junto con un conjunto de atributos, y dependiendo del diseño del modelo multidimensional puede existir una foreign key que determina su relación con otra tabla lock_up.

Para decidir si un campo de datos es un atributo o un hecho se analiza la variación de la medida a través del tiempo. Si varía continuamente implicaría tomarlo como un hecho, caso contrario será un atributo.

Los atributos dimensionales son un rol determinante en un data warehouse. Ellos son la fuente de todas las necesidades que debieran cubrirse. Esto significa que la base de datos será tan buena como lo sean los atributos dimensionales, mientras más descriptivos, manejables y de buena calidad, mejor será el data warehouse.

2.8 Metodología de Desarrollo

2.8.1 Rapid Warehousing Methodology

Es una metodología propuesta por SAS Institute[3]. Esta metodología es iterativa, y está basada en el desarrollo incremental de almacén de datos dividido en cinco fases como se puede apreciar en la Figura 6.

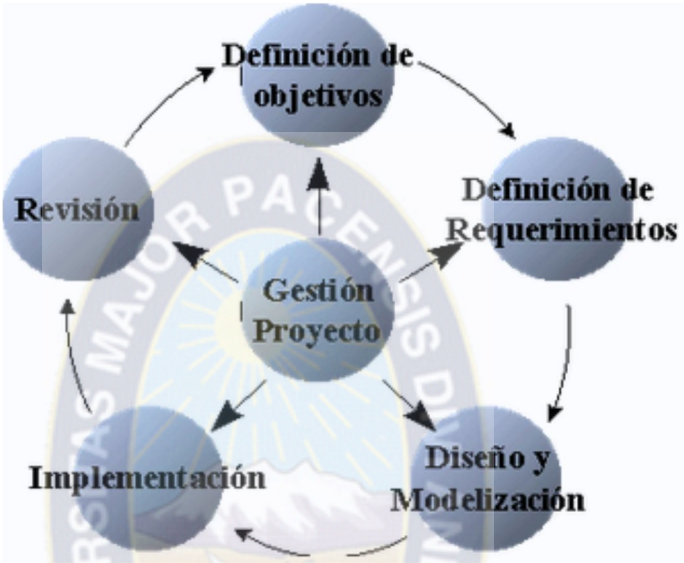


Figura 6: Fases de la Metodología Rapid Warehousing Methodology
Fuente: SAS Institute

- Definición de Objetos.
- Definición de Requerimientos de Información.
- Diseño y Modelización.
- Implementación.
- Revisión.

2.8.1.1 Definición de Objetos

En efecto, como punto de arranque de todo, es preciso "vender la idea" a los usuarios finales de un Data Warehouse. Esto es así, por ser una idea bastante novedosa y sobre la que pueden surgir recelos de su efectividad. Estos recelos se pueden eliminar

[3] Instituto de Sistemas de Análisis Estadístico (Statistical Analisys Sistems Intitute).

comenzando por un pequeño módulo, del cual se valoren los beneficios posteriores, para iniciar progresivamente el desarrollo de nuevos módulos, cada uno con un coste unitario cada vez más reducido, pero sin embargo con unos beneficios distribuidos cada vez mayores por poder cada vez incluir más información.

El simple hecho de realizar un informe de necesidades previas en el que se enumeren la situación de los datos entre los diversos sistemas operacionales, puede ser un hecho decisivo para emprender un proyecto de este tipo. Muchas veces la información existente se encuentra tan poco normalizada, existen tantas discrepancias entre estos sistemas, que el abordar un Data Warehouse en el que se limpien estos datos y se normalicen pueden aportar un valor intangible: "*la calidad y fiabilidad de la información*".

La venta de esta idea no sólo se ha de realizar frente a la Dirección sino que es preciso realizarla a todos los niveles: a la Dirección, Gerencia e incluso al área de Desarrollo.

Tras esta venta de la idea, comienzan dos fases similares al análisis de requisitos del sistema (ARS según abreviaturas de la metodología MÉTRICA): la definición de objetivos y requerimientos de información, en el que se analicen las necesidades del comprador.

2.8.1.2 Definición de Requerimientos de Información

En esta fase se definirá el equipo de proyecto que debe estar compuesto por representantes del departamento informático y de los departamentos usuarios del Data Warehouse además de la figura de jefe de proyecto.

Se definirá el alcance del sistema y cuáles son las funciones que el Data Warehouse realizará como suministrador de información de negocio estratégica para la empresa. Se definirán así mismo, los parámetros que permitan evaluar el éxito del proyecto.

Durante esta fase se mantendrán sucesivas entrevistas con los representantes del departamento usuario final y los representantes del departamento de informática. Se

realizará el estudio de los sistemas de información existentes, que ayudaran a comprender las carencias actuales y futuras que deben ser resueltas en el diseño del Data Warehouse

Asimismo, en esta fase el equipo de proyecto debe ser capaz de validar el proceso de entrevistas y reforzar la orientación de negocio del proyecto. Al finalizar esta fase se obtendrá el documento de definición de requerimientos en el que se reflejarán no solo las necesidades de información de los usuarios, sino cual será la estrategia y arquitectura de implantación del Data Warehouse.

2.8.1.3 Diseño y Modelización

Los requerimientos de información identificados durante la anterior fase proporcionarán las bases para realizar el diseño y la modelización del Data Warehouse.

En esta fase se identificarán las fuentes de los datos (sistema operacional, fuentes externas,..) y las transformaciones necesarias para, a partir de dichas fuentes, obtener el modelo lógico de datos del Data Warehouse. Este modelo estará formado por entidades y relaciones que permitirán resolver las necesidades de negocio de la organización.

El modelo lógico se traducirá posteriormente en el modelo físico de datos que se almacenará en el Data Warehouse y que definirá la arquitectura de almacenamiento del Data Warehouse adaptándose al tipo de explotación que se realice del mismo.

La mayor parte estas definiciones de los datos del Data Warehouse estarán almacenadas en los metadatos y formarán parte del mismo.

2.8.1.4 Implementación

La implantación de un Data Warehouse lleva implícitos los siguientes pasos:

1. Extracción de los datos del sistema operacional y transformación de los mismos.
2. Carga de los datos validados en el Data Warehouse. Esta carga deberá ser planificada con una periodicidad que se adaptará a las necesidades de refresco detectadas durante las fases de diseño del nuevo sistema.

3. Explotación del Data Warehouse mediante diversas técnicas dependiendo del tipo de aplicación que se dé a los datos:
    - Query & Reporting
    - On-line analytical processing (OLAP)
    - Executive Information System (EIS) o Información de gestión
    - Decision Support Systems (DSS)
    - Visualización de la información

La información necesaria para mantener el control sobre los datos se almacena en los metadatos técnicos (cuando describen las características físicas de los datos) y de negocio (cuando describen cómo se usan esos datos). Dichos metadatos deberán ser accesibles por los usuarios finales que permitirán en todo momento tanto al usuario, como al administrador que deberá además tener la facultad de modificarlos según varíen las necesidades de información.

Con la finalización de esta fase se obtendrá un Data Warehouse disponible para su uso por parte de los usuarios finales y el departamento de informática.

2.8.1.5 Revisión

La construcción del Data Warehouse no finaliza con la implantación del mismo, sino que es una tarea iterativa en la que se trata de incrementar su alcance aprendiendo de las experiencias anteriores.

Después de implantarse, debería realizarse una revisión del Data Warehouse planteando preguntas que permitan, después de los seis o nueve meses posteriores a su puesta en marcha, definir cuáles serían los aspectos a mejorar o potenciar en función de la utilización que se haga del nuevo sistema.

2.9 Lenguaje Unificado de modelos para Inteligencia de Negocios

El perfil de UML para el diseño conceptual de almacenes de datos según el modelado multidimensional permite representar las principales propiedades multidimensionales a

un nivel conceptual, como son las relaciones muchos-a-muchos entre hechos y dimensiones, las dimensiones degeneradas, las jerarquías múltiples y de camino alternativo, etc.

El perfil de UML se encuentra definido formalmente mediante reglas expresadas con Object Constraint Lenguage que defienden el correcto uso de los nuevos elementos de modelado, con lo que se evita un uso arbitrario del perfil. Se establece el proceso de diseño en tres niveles.

Nivel 1: Definición del modelo, un paquete representa un esquema estrella de un modelo multidimensional, En este nivel, una dependencia entre dos paquetes indica que los esquemas estrella comparten al menos una dimensión.

Nivel 2: Definición de un esquema estrella. Un paquete representa un hecho o una dimensión de esquema estrella. En este nivel, una dependencia entre dos paquetes de dimensión indica que las dimensiones comparten al menos un nivel en sus correspondientes jerarquías.

Nivel 3: Definición de un hecho o dimensión. Se compone de un conjunto de clases que representan los niveles jerárquicos en un paquete de dimensión o el esquema estrella completo en el caso de un paquete de hecho.

2.10 Norma ISO/IEC 25012 para Calidad de Datos

El modelo de Calidad de Datos representa los cimientos sobre los cuales se construye un sistema para la evaluación de un producto de datos. En un modelo de Calidad de Datos se establecen las características de Calidad de Datos que se deben tener en cuenta a la hora de evaluar las propiedades de un producto de datos determinado.

La Calidad del Producto de Datos se puede entender como el grado en que los datos satisfacen los requisitos definidos por la organización a la que pertenece el producto. Son precisamente estos requisitos los que se encuentran reflejados en el modelo de

Calidad de Datos mediante sus características (Precisión, Completitud, Consistencia, Credibilidad, Actualidad, Accesibilidad, etc.).

El modelo de Calidad de Producto de Datos definido por el estándar ISO/IEC 25012 se encuentra compuesto por las 15 características, los cuales están clasificados en dos categorías:

Calidad de Datos Inherente: Se refiere al grado con el que las características de calidad de los datos tienen el potencial intrínseco para satisfacer las necesidades establecidas y necesarias cuando los datos son utilizados bajo condiciones específicas. Desde el punto de vista inherente, la Calidad de Datos se refiere a los mismos datos, en particular a:

Valores de dominios de datos y posibles restricciones (e.g., Reglas de Negocio gobernando la calidad requerida por las características en una aplicación dada).

Relaciones entre valores de datos (e.g., Consistencia).

Metadatos.

Calidad de Datos Dependiente del Sistema: Se refiere al grado con el que la Calidad de Datos es alcanzada y preservada a través de un sistema informático cuando los datos son utilizados bajo condiciones específicas.

Las características dentro de la Calidad de Datos Inherente son:

Precisión: Grado en el que los datos representan correctamente el verdadero valor del atributo deseado de un concepto o evento en un contexto de uso específico.

Tiene dos principales aspectos:

Precisión Sintáctica: cercanía de los valores de los datos a un conjunto de valores definidos en un dominio considerado sintácticamente correcto.

Precisión Semántica: cercanía de los valores de los datos a un conjunto de valores definidos en un dominio considerado semánticamente correcto.

Completitud: Grado en el que los datos asociados con una entidad tienen valores para todos los atributos esperados e instancias de entidades relacionadas en un contexto de uso específico.

Consistencia: Grado en el que los datos están libres de contradicción y son coherentes con otros datos en un contexto de uso específico. Puede ser analizada en datos que se refieran tanto a una como a varias entidades comparables.

Credibilidad: Grado en el que los datos tienen atributos que se consideran ciertos y creíbles en un contexto de uso específico. La credibilidad incluye el concepto de autenticidad (la veracidad de los orígenes de datos, atribuciones, compromisos).

Actualidad: Grado en el que los datos tienen atributos que tienen la edad correcta en un contexto de uso específico.

Las características dentro de la Calidad de Datos Inherente y Dependiente del Sistema

Accesibilidad: Grado en el que los datos pueden ser accedidos en un contexto específico, particularmente por personas que necesiten tecnologías de apoyo o una configuración especial por algún tipo de discapacidad.

Conformidad: Grado en el que los datos tienen atributos que se adhieren a estándares, convenciones o normativas vigentes y reglas similares referentes a la calidad de datos en un contexto de uso específico.

Confidencialidad: Grado en el que los datos tienen atributos que aseguran que los datos son sólo accedidos e interpretados por usuarios autorizados en un contexto de uso específico. La confidencialidad es un aspecto de la seguridad de la información (junto con la disponibilidad y la integridad) definida como en ISO/IEC 13335-1:2004.

Eficiencia: Grado en el que los datos tienen atributos que pueden ser procesados y proporcionados con los niveles de rendimiento esperados mediante el uso de cantidades y tipos adecuados de recursos en un contexto de uso específico.

Exactitud: Grado en el que los datos tienen atributos que son exactos o proporcionan discernimiento en un contexto de uso específico.

Comprensibilidad: Grado en el que los datos tienen atributos que permiten ser leídos e interpretados por los usuarios y son expresados utilizando lenguajes, símbolos y unidades apropiados en un contexto de uso específico. Cierta información sobre la comprensibilidad puede ser expresada mediante metadatos

2.11 Herramientas de Desarrollo

2.11.1 Oracle Warehouse Builder

Es una herramienta para la creación del esquema del DW, definición de jerarquías y medidas, mapeo de las fuentes de información, calendarización, ejecución y mantenimiento de las actividades de ETL y herramientas para mejorar la calidad de la información.

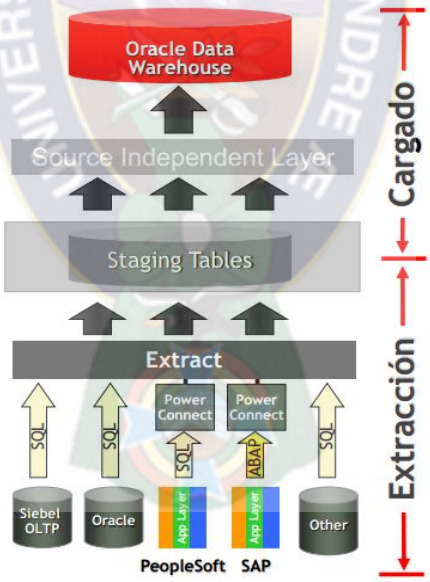


Figura 7: Proceso de ETL, utilizando la herramienta Oracle Warehouse Builder
Fuente: Oracle, Oracle Data Warehouse Builder

2.11.2 Oracle BI Administration Tool

Es la herramienta que nos permite crear y administrar los repositorios de datos que se utilizaran dentro de los catálogos en el Oracle Business Intelligence para crear nuestros

reportes, gráficos y cuadros de trabajo o Dashboards. Desde esta aplicación se definirán las tres (3) capas necesarias para completar la creación de nuestros metadatos, las cuales son: capa Física, capa del Modelo de Negocio y Mapeo y por último la capa de Presentación.

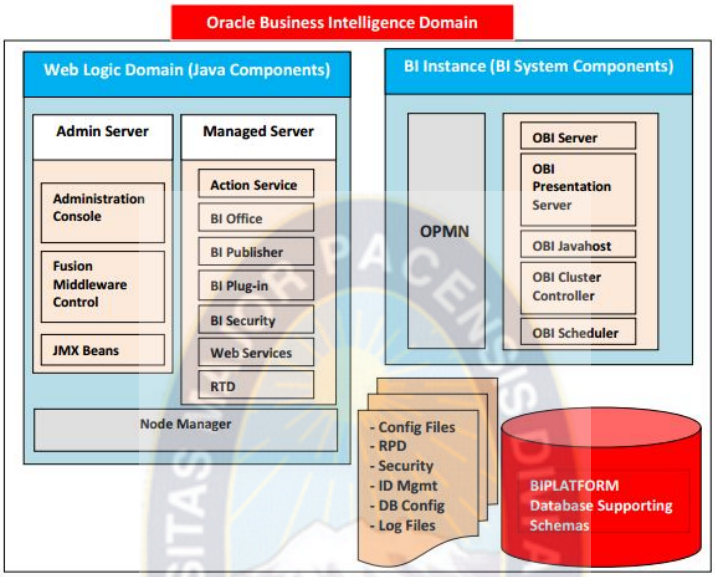


Figura 8: Arquitectura de Oracle Business Intelligence
Fuente: Oracle, Business Intelligence

2.11.3 Enterprise Manager Fusion Middleware

Esta herramienta Fusion Middleware Control organiza una amplia variedad de datos de rendimiento y de funciones administrativas en distintas páginas iníciales basadas en web. Estas páginas iníciales facilitan la localización de los datos de supervisión más importantes y las funciones administrativas más habituales de un componente, todo desde el explorador web.

Fusion Middleware Control es la herramienta de gestión de nivel superior para Web Center Content con una instancia de Content Server y se puede utilizar para:

- Desplegar, anular el despliegue y volver a desplegar Web Center Content con una instancia de Content Server.
- Configurar servicios backend.

- Configurar la gestión de seguridad.
- Controlar el ciclo de vida del proceso.
- Exportar e importar datos.
- Acceder a archivos log y gestionar la configuración de los logs.
- Gestionar migraciones.
- Supervisar el rendimiento.
- Diagnosticar problemas de tiempo de ejecución.

2.11.4 Analytics

Es herramienta de Analítica Web. Ofrece información agrupada del tráfico que llega a los sitios web según la audiencia, la adquisición, el comportamiento y las conversiones que se llevan a cabo en el sitio web. Se pueden obtener informes como el seguimiento de usuarios exclusivos, el rendimiento del segmento de usuarios, los resultados de las diferentes campañas de marketing online, las sesiones por fuentes de tráfico, tasas de rebote, duración de las sesiones, contenidos visitados, conversiones, etc.

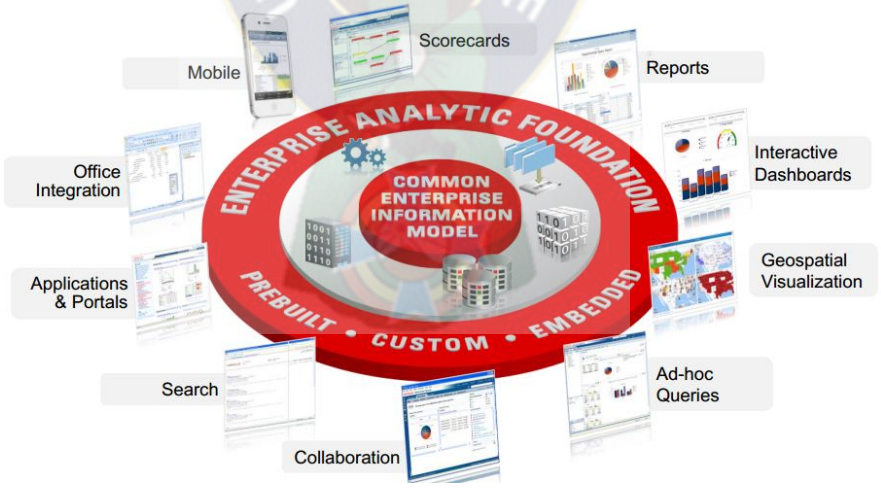


Figura 9: Herramienta Analytics
Fuente: Oracle, Business Intelligence

CAPITULO III

MARCO METODOLÓGICO

3.1 Construcción de Metadatos

Los metadatos cumplen un rol muy importante, ya que representan la información de los datos almacenados en el Data Warehouse, es por ello que el modelo de la Plataforma de Información Estadística se estructura a partir de sus metadatos, previo análisis de requerimientos de los usuarios y fuente de datos de donde se extraerá la información. Posteriormente se identificarán la estructura de las tablas hecho y dimensiones como se muestra en la Figura 10.

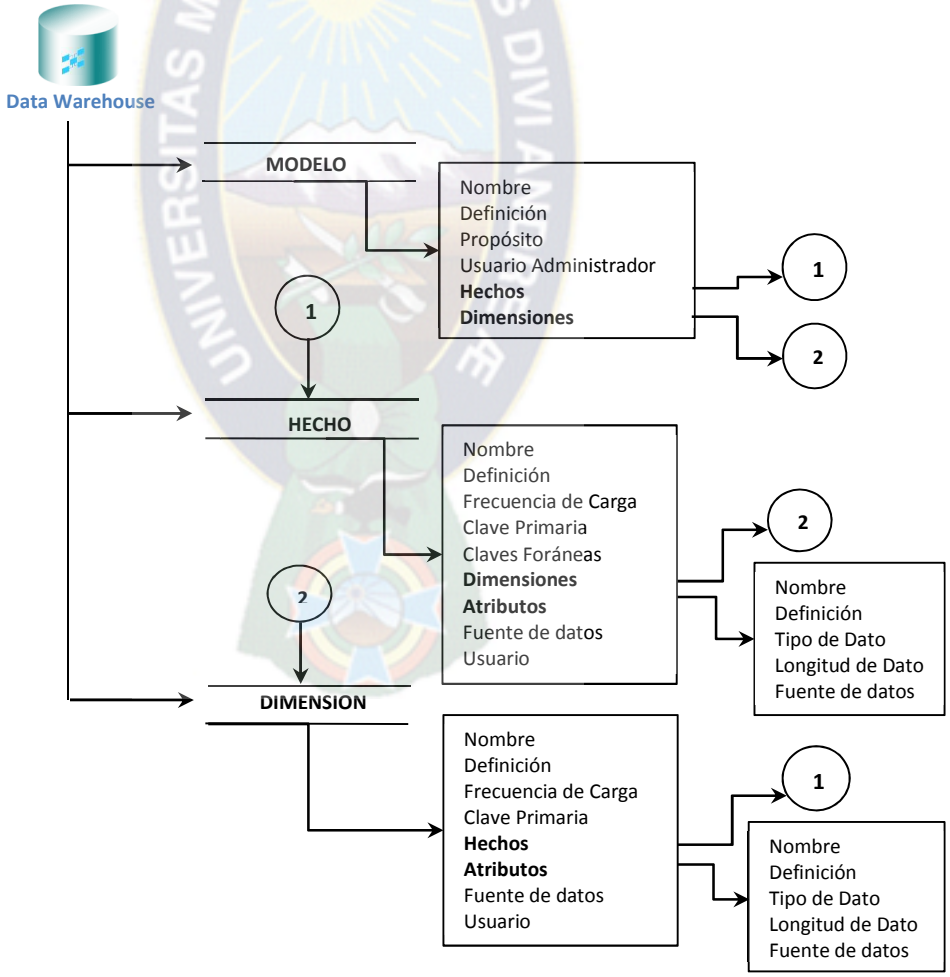


Figura 10: Esquema de Elaboración de Metadatos
Fuente: Adaptación de Data Modeling Techniques for Data Warehousing.
IBM International Technical Support Organization

3.1.1 Estructura de Metadato del Modelo de Data Warehouse

En el Data Warehouse, el modelo de datos es utilizado continuamente como base, porque los usuarios constantemente consultan el modelo para determinar los datos que ellos requieren utilizar para analizar algún aspecto de la institución, en este contexto se muestra en el Tabla 1 la estructura del Metadato del Data Warehouse de la Plataforma de Información Estadística.

Metadatos del Modelo de Data Warehouse	
Nombre:	Identificara el Modelo del Data Warehouse por "Plataforma de Información Estadística", esto para diferenciar de otros modelos que existan.
Definición:	La Plataforma de Información Estadística modelara una base de datos que almacenara información con respecto a datos socioeconómicos y antecedentes universitarios de los estudiantes de la U.M.S.A.
Propósito:	El propósito de implementar la Plataforma de Información Estadística, es con el fin de brindar información de apoyo, para realizar una mejor gestión académica universitaria.
Usuario Administrador:	Persona responsable que realizara las operaciones de actualizaciones y mantenimiento del almacén de datos.
Hechos:	El modelo estará construido por 4 tablas hecho, las siguientes son: Datos Socioeconomicos, Vida Histórica en la U.M.S.A., Seguro Universitario, Postulantes a Titulación.
Dimensiones:	El modelo estará conformado por 36 tablas dimensiones, a continuación mencionaremos las tablas más representativas: Persona, Estudiante, País Nacionalidad, Ocupación Padres, Vivienda, Colegio, Trabajo, Carrera, Facultad, Sede, las descripciones de las restantes dimensiones se encuentran en el

Tabla 1: Estructura de Metadato del Modelo Data Warehouse
Fuente: Elaboración Propia

3.1.2 Estructura de Metadatos Tablas Hechos

Los metadatos de las tablas hechos describen la estructura que comprenderán cada una de las ellas dentro del diseño del almacén de datos, en el Tabla 2 se observa la estructura

que tiene un hecho dentro del Data Warehouse de la Plataforma de Información Estadística.

Metadatos de Hechos	
Nombre:	Las 4 tablas hechos ya mencionadas se identificaran en el modelo según tema de estudio, con los siguientes nombres: H_Datos_Socioeconomicos, H_Vida_Historica, H_Seguro_Universitario, H_Postulantes_Titulo.
Definición:	Cada tabla hecho modelara datos agrupados y ordenados según el tema de interés que se requiere analizar.
Frecuencia de Carga:	La actualización o carga de datos de las tablas hechos se realizara cada semestre de una gestión.
Clave Primaria:	Cada tabla hecho obtendrá como propiedad una clave primaria para el control de cada registro que se almacene en esta tabla.
Claves Foráneas:	Cada tabla hecho obtendrá como propiedad claves foráneas que referenciaran a las dimensiones con las que se encuentran asociadas.
Dimensiones:	Nombres de las dimensiones asociadas a cada tabla hecho.
Atributos:	Lista de atributos donde se informa el nombre, tipo de dato, longitud del dato y la fuente de dato, cada tabla hecho tiene determinados atributos que varían continuamente, según análisis de necesidad.
Fuente de Datos:	La primera carga de datos a las tablas hecho se realizara de a través de las siguientes fuentes de datos: Aplicación Informática OK y Sistema de Registro Universitario.
Usuario Administrador:	Persona responsable que realizara las operaciones de actualizaciones y mantenimiento de las tablas hecho.

Tabla 2: Estructura de Metadato de Hechos
Fuente: Elaboración Propia

3.1.3 Estructura de Metadatos Tablas Dimensiones

Los metadatos de las tablas dimensiones describen la estructura que comprenderán cada una de las ellas dentro del diseño del almacén de datos, en el Tabla 3 se observa la estructura que tiene una dimensión dentro del Data Warehouse de la Plataforma de Información Estadística.

Metadatos de Dimensiones	
Nombre:	Las 36 tablas dimensiones ya mencionadas se identificaran en el modelo con los siguientes nombres: Dim_Persona, Dim_Estudiante, Dim_Pais_Nacionalidad, Dim_Ocupacion_Padre, Dim_Vivienda, Dim_Colegio, Dim_Trabajo_Ocupacion, Dim_Carrera_Programa, Dim_Facultad, Dim_Sede_Localidad, las descripciones de las restantes dimensiones se encuentran en el
Definición:	Cada tabla dimensión modelara datos ordenados, esos proveerán de información descriptiva a las tablas hechos, según necesidad de información que requiere la estructura de cada tablas hecho.
Frecuencia de Carga:	La actualización o carga de datos de las tablas dimensiones se realizara según modificaciones que se realicen en los formularios estadísticos o Sistema de Registro Universitario, excepto las Dim_Persona y Dim_Estudiante que tendrán actualizaciones por semestre en cada gestión.
Clave Primaria:	Cada tabla dimensión obtendrá como propiedad una clave primaria para el control de cada registro que se almacene en esta tabla.
Hechos:	Nombres de las hechos asociados a cada tabla dimensión.
Atributos:	Lista de atributos donde se informa el nombre, tipo de dato, longitud del dato y la fuente de dato, cada tabla dimensión tiene determinados atributos que pueden ser constantes o discretos, según análisis de necesidad.
Fuente de Datos:	La carga de datos de las tablas dimensión se realizara a través del Sistema de Registro Universitario.
Usuario Administrador:	Persona responsable que realizara las operaciones de actualizaciones y mantenimiento de las tablas dimensión.

Tabla 3: Estructura de Metadato de Dimensiones
Fuente: Elaboración Propia

3.2 Construcción de la Base de Datos del Data Warehouse

Para la construcción o diseño de la base de Datos del Data Warehouse se aplicara el modelo multidimensional esquema en estrella, con el que se modelará las tablas Hechos con sus respectivas dimensiones referenciadas mediante llaves foráneas.

Siguiendo la estructura del metadato para el diseño del Data Warehouse se indicó cuatro tablas Hechos, cada una relacionada con sus correspondientes dimensiones según requerimientos de información, mencionadas tablas Hecho son:

- **Tabla Hecho Datos Socioeconómicos:** Almacena toda la información relacionada con datos personales, sociales y económicos de los estudiantes universitarios, con el objetivo de construir indicadores estadísticos socioeconómicos y a partir de estos adquirir conocimiento sobre las características sociales de la población estudiantil de la U.M.S.A. La tabla hecho Datos Socioeconomicos depende de las dimensiones persona, género, estado civil, país nacionalidad y origen, provincia, departamento de nacimiento, grado académico, ocupación del padre y madre, características de vivienda, características de colegio, ocupación del estudiante, las cuales proveen de información a la tabla hecho.

- **Tabla Hecho Vida Histórica U.M.S.A.:** Contiene toda la información relacionada con datos académicos de matriculación de los estudiantes universitarios, con el objetivo de realizar un análisis estadístico, brindando de esta manera información de apoyo para realizar una apropiada gestión académica universitaria. La tabla hecho Vida Histórica depende de las dimensiones persona, estudiante, carrera, facultad, tipo de ingreso a la U.M.S.A, sede, matrícula, estado, nivel, tipo y grado del estudiante, las cuales proveen de información a la tabla hecho.

- **Tabla Hecho Seguro Universitario:** Guarda toda la información relacionada con datos del seguro social medico universitario PROMES, con el fin de realizar un análisis estadístico más completo y descriptivo sobre los estudiantes universitarios, asociando el resultado de análisis con datos de las tablas hecho Datos Socioeconómico y Vida Histórica. La tabla hecho Seguro Universitario depende de las dimensiones persona, estudiante, carrera, facultad y sede.

- **Tabla Hecho Postulación a Titulo:** Colecciona todos los datos relacionados con estudiantes que presentan sus trámites o solicitudes para optar a una titulación universitaria, a objeto de contar con parámetros de información sobre los profesionales que genera la U.M.S.A. La tabla hecho Postulantes a Titulo

depende de las dimensiones persona, tipo de titulación, mención de la titulación, carrera, facultad y sede.

Cabe mencionar que las tablas hecho compartirán más de una dimensión, con el objeto de relacionar la los datos de las tablas hechos y así obtener información más representativa, respecto a las dimensiones no podrán tener una dependencia con otra dimensión, respetando así la estructura del esquema en estrella.

El Data Warehouse almacenara datos que serán previamente analizados en el proceso de extracción, transformación y carga (ETL), según requerimientos y necesidades para la construcción de las tablas hecho ya mencionadas.

3.3 Construcción de Cubos

Para la construcción de los cubos multidimensionales se abstraerá datos del data warehouse, relacionando cada cubo según requerimientos de los usuarios. Una vez preparado el cubo especificaremos el uso de las dimensiones, seleccionando el tipo de relación y la granularidad de atributos, del resultado de este proceso dependerá la presentación de las vistas que el usuario administre.

Figura 11: Esquema de Elaboración de Metadatos
Fuente: Elaboración Propia

CAPITULO IV

MARCO APLICATIVO

4.1 Definición de Objetivos

Los objetivos del presente proyecto se definen en el capítulo I, sin embargo mencionaremos algunos de ellos.

- Realizar un análisis detallado de la estructura de la base de datos relacional del Sistema de Registro Universitario y la organización de datos de la aplicación informática OK, estableciendo formas y niveles de relación entre ellos.
- Diseñar una estructura para el almacenamiento de datos que permita la integración de información histórica, a partir de datos almacenados en el Sistema de Registro Universitario y herramientas informáticas usadas.
- Generar una solución que permita la agregación de la información de acuerdo a requerimientos del usuario.

Para llegar a cumplir cada uno de los objetivos mencionados se tomara en cuenta los parámetros que de definen en el Anexo A, los cuales se definieron previo análisis de requerimiento de cada uno de los usuarios.

4.2 Definición de Requerimientos de Información

De acuerdo a la metodología se definen dos tipos de requerimientos; Requerimientos de información, requerimientos funcionales e identificación de datos en las bases operacionales.

4.2.1 Requerimientos de los Usuarios

- **Usuarios que trabajan con la información**

  Personal de la División de Sistemas de Información y Estadística que se encargan de la elaboración de información estadística con relación a datos

socioeconómicos y antecedentes universitarios tanto actuales como históricos, gestionando las bases de datos y otras fuentes de información.

A continuación se detalla los requerimientos para los usuarios que trabajan en la elaboración de la información estadística:

- ✓ La plataforma estadística debe almacenar información histórica sobre datos socioeconómicos desde la gestión 2010 hasta la gestión 2015.
- ✓ La plataforma estadística tendrá que guardar información histórica sobre antecedentes universitarios desde la gestión 1990 hasta la gestión 2015.
- ✓ La plataforma estadística debe coleccionar información histórica sobre estudiantes postulantes a titulación y estudiantes con seguro universitario desde la gestión 2008 hasta la gestión 2015.
- ✓ El sistema debe mostrar información contenida en los siguientes sistemas y aplicaciones que recopilan la información acerca los temas mencionados: Sistema de Registro Universitario (SRU), aplicación informática OK.
- ✓ La información debe ser desplegado en un entorno que permita la personalización de reportes.
- ✓ La información almacenada debe provenir de datos oficiales por cada gestión.

- **Usuarios tomadores de decisión**

Está conformado por personal encargado de la toma de decisiones, con algún cargo jerárquico dentro de la institución (Rector, jefes de departamentos, divisiones), y personas externas a la institución.

- ✓ Los datos deben ser oficiales.
- ✓ La información estadística debe ser confiable y fidedigna.

- ✓ Se debe proporcionar reportes con información estadística completa, que contenga la mayor descripción entre unión de variables contenidas en el almacén de datos.

4.2.2 Requerimientos de Información

A partir de las entrevistas realizadas con los dos tipos de usuarios se pudo identificar las siguientes necesidades de información, dando esto lugar a tomar en cuenta los datos necesarios que serán cargados en el Data Warehouse de la Plataforma de Información Estadística.

Unidades y Departamentos	Requerimiento	Solución
Departamento de Bienestar Social, División de Sistemas de Información y Estadística.	Conocer información socioeconómica de los estudiantes de la U.M.S.A. es de vital importancia para el departamento de Bienestar Social debido a los proyectos que ejecutan sobre Beca Trabaja, Beca Comedor, Guardería Infantil, proyectos de análisis de características sociales de los universitarios, etc. Debido a las necesidades que tienen otras unidades, departamentos e instituciones ajenas a la U.M.S.A., se requiere realizar un análisis estadístico sobre datos sociales y económicos de la población estudiantil. La información estadística debe ser analizada a nivel carrera, facultad y universidad.	• Total estudiantes por lugar de nacimiento (país o departamento). • Total estudiantes por género, estado civil, edad (tomando en cuenta intervalos de medición). • Total estudiantes relacionado con la ocupación de los padres. • Total estudiantes relacionado con características de vivienda que habita y zona. • Total estudiantes relacionada con estudios realizados a nivel colegio. • Total estudiantes relacionada con ocupación y área laboral del estudiante.

Tabla 4: Matriz de Requerimientos de Datos Socioeconomicos
Fuente: Elaboración Propia

Unidades y Departamentos	Requerimiento	Solución
Departamento de Bienestar Social, División de Sistemas de Información y Estadística.	Obtener información estadística socioeconómica más descriptiva con análisis más complejos, los datos deben ser analizados a nivel carrera y facultad.	• Perfiles sociales de los estudiantes, relacionando las variables ocupación laboral, características de vivienda, características del colegio de egreso y ocupación del estudiante.
Departamento de Bienestar Social.	Conocer información para ejecutar proyectos de Guardería Infantil, etc.	• Total estudiantes según número de hermanos, número de hijos y dependientes.
Div. de Sistemas de Información y Estadística.	Conocer información sobre estudiantes que se benefician con el seguro universitario PROMES, analizar datos a nivel carrera, facultad y sedes.	• Total estudiantes con seguro universitario.
Div. de Sistemas de Información y Estadística.	Conocer información sobre el crecimiento poblacional, comparaciones poblacionales e implementar porcentajes de medición, todo esto como resultado de la necesidad de información solicitada por unidades, departamentos y autoridades internas y externas a la U.M.S.A.	• Crecimiento poblacional estudiantil comparando variables de género, estado civil, lugar de nacimiento y otras variables. • Comparación de crecimiento poblacional estudiantil entre la ciudad de La Paz, el Alto y sedes. • porcentajes a cada variable de

Tabla 5: Matriz de Requerimientos de Datos Socioeconomicos
Fuente: Elaboración Propia

Unidades y Departamentos	Requerimiento	Solución
Dpto. de Bienestar Social, Div. de Sistemas de Información y Estadística, Div. de Gestiones, Admisiones y Registros.	Gestionar el crecimiento poblacional estudiantil según emisión de matrícula, analizar a nivel carrera facultad y U.M.S.A.	• Total estudiantes matriculados antiguos. • Total estudiantes matriculados nuevos. • Total estudiantes por modalidad de ingreso a la U.M.S.A., evaluar por carrera y facultad.
Div. de Sistemas de Información y Estadística, Div. de Gestiones, Admisiones y Registros.	Generar información sobre los años de estudio que lleva un universitario en una carrera, facultad y U.M.S.A.	• Total estudiantes según años de permanencia en la U.M.S.A, tomando en cuenta intervalos (1 año, 2 años, 3 años, 4 años, 5 - 6 años, 7-8 - 9 años, 10 – 11 años, > 11 años).
Div. de Sistemas de Información y Estadística, Div. de Gestiones, Admisiones y Registros.	Conocer información sobre los antecedentes universitarios, es muy significativo para realizar una mejor gestión académica, en consecuencia a este análisis poder ejecutar proyectos y políticas en beneficios a los estudiantes de la U.M.S.A.	• Cantidad de estudiantes por sedes de la U.M.S.A. • Cantidad aproximada de estudiantes que abandonan su carrera. • Cantidad de estudiantes que conllevan carreras paralelas. • Cantidad de estudiantes que realizan estudios en pregrado y continúan a nivel postgrado de la U.M.S.A. • Número de Titulados por menciones.

Tabla 6: Matriz de Requerimientos de Datos Vida Histórica Académica
Fuente: Elaboración Propia

4.2.3 Análisis de los Orígenes de Datos

Realizando todo el análisis de requerimientos se identificó 3 fuentes de información, en los cuales la Plataforma Estadística deberá ser capaz de unificar todas estas para un posterior análisis.

- **Sistema de Registro Universitario**

  Antiguamente el Sistema de Matriculación Jiwasana, realizaba la generación y control de las matrículas universitarias ya desde la gestión 2005 para pregrado y gestión 2008 para postgrado, debido a los problemas que contenía este sistema, se realizó una reingeniería para integrar sistemas e incorporar nuevos procesos al mismo.

  En la actualidad el mencionado sistema fue renombrado como Sistemas de Registro Universitario (SRU) en donde se realizaron muchos cambios, mejorando así su estructura y su desarrollo tecnológico, para poder de esta manera presentar una mejor solución informática con relación a la gestión académica de los estudiantes matriculados en la U.M.S.A., el objetivo principal es el registro y control de la matriculación universitaria tanto de pregrado como de postgrado, registro de estudiantes al seguro social universitario PROMES, registro de tramites de Gestiones, y el control de colegiaturas de postgrado.

  Es así que realizando la gestión de matriculación de estudiantes se toman también en cuenta los datos sociales y económicos de los estudiantes, llenando los formularios de Registro Universitario (Formulario 01) y Encuesta para Titulación de pregrado (Formulario 03) los cuales se muestran en el Anexo B.

  Granularidad del Sistema de Registro Universitario

  El Sistema de Registro Universitario está desarrollado en Java y la base de datos se encuentra almacenada en PostgresSQL.

  El tipo de información almacenada en el sistema es:

- Información relativa a Personas-Funcionarios: Datos personales y académicos de funcionarios de la U.M.S.A.
- Información referente a Persona – Estudiante: Datos sociales, económicos y académicos de estudiantes de la U.M.S.A.

- **Aplicación Informática OK**

  La Aplicación Informática OK que es implementada desde a gestión 1992, fue diseñada específicamente para el almacenamiento de datos estadísticos de la población estudiantil de la Universidad Mayor de San Andrés, desde entonces es la herramienta que procesa los reportes e indicadores estadísticos, la estructura de inicio de dicha aplicación a sufrido muchos cambios según cambios del sistema del cual depende y también según necesidades de los usuarios.

  Granularidad de la Aplicación Informática OK

  La aplicación OK está desarrollado en FoxPro, y su base de datos está en Access.

  El nivel de información que administra esta aplicación es:

  - Información estadística socioeconómica: Almacena información social y económica de los estudiantes de la U.M.S.A., cada una de ellas contempladas en variables.
  - Información académica: Información de los estudiantes relacionada con la matriculación.

- **Hojas de cálculo Excel**

  Las hojas de cálculo Excel, donde se almacena los datos estadísticos desde la gestión 1992, son los reportes que se administran para sacar información estadística de manera general (nivel de carrera, facultad y U.M.S.A.), es decir el nivel de detalle de sus datos son incompletos ya que representa cierto grupo de las variables estadísticas de manera concreta y no muestra a detalle la información de un estudiante.

Es por ello que la información almacena en las hojas Excel se tomara solo como referencia para el cálculo de los totales de cada variable estadística.

4.3 Diseño y Modelación

En esta fase se describen las fuentes de datos SRU y Aplicación OK, de igual forma las transformaciones necesarias para que a partir de mencionadas fuentes, se obtenga el modelo lógico de datos del Data Warehouse.

El modelo lógico se traducirá posteriormente en el modelo físico de datos la cual definirá la arquitectura de almacenamiento del Data Warehouse, adaptándose a la explotación de datos de consultas OLAP.

4.3.1 Arquitectura de la Plataforma Estadística

Visualizando la arquitectura que comprende una herramienta de inteligencia de negocios se propone el siguiente esquema de acuerdo al nivel de resumen de los datos, implementando así 5 capas.

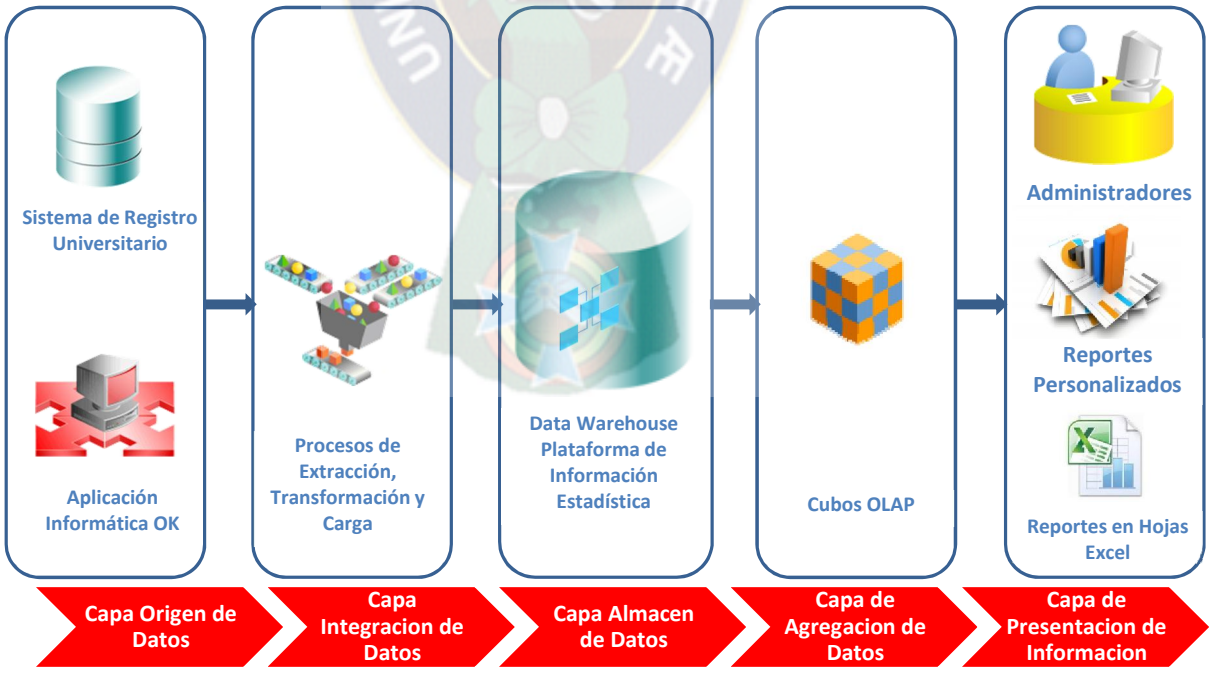


Figura 12: Arquitectura del Plataforma de Información Estadística
Fuente: Elaboración Propia

4.3.2 Capa Origen de Datos

Se definen las fuentes de origen de los datos que proporcionaran información al Data Warehouse, con relación al proyecto se identificó el Sistema de Registro Universitario y la aplicación informática OK.

Fuente Origen de Datos: Sistema de Registro Universitario

Se analiza el modelo relacional del sistema.

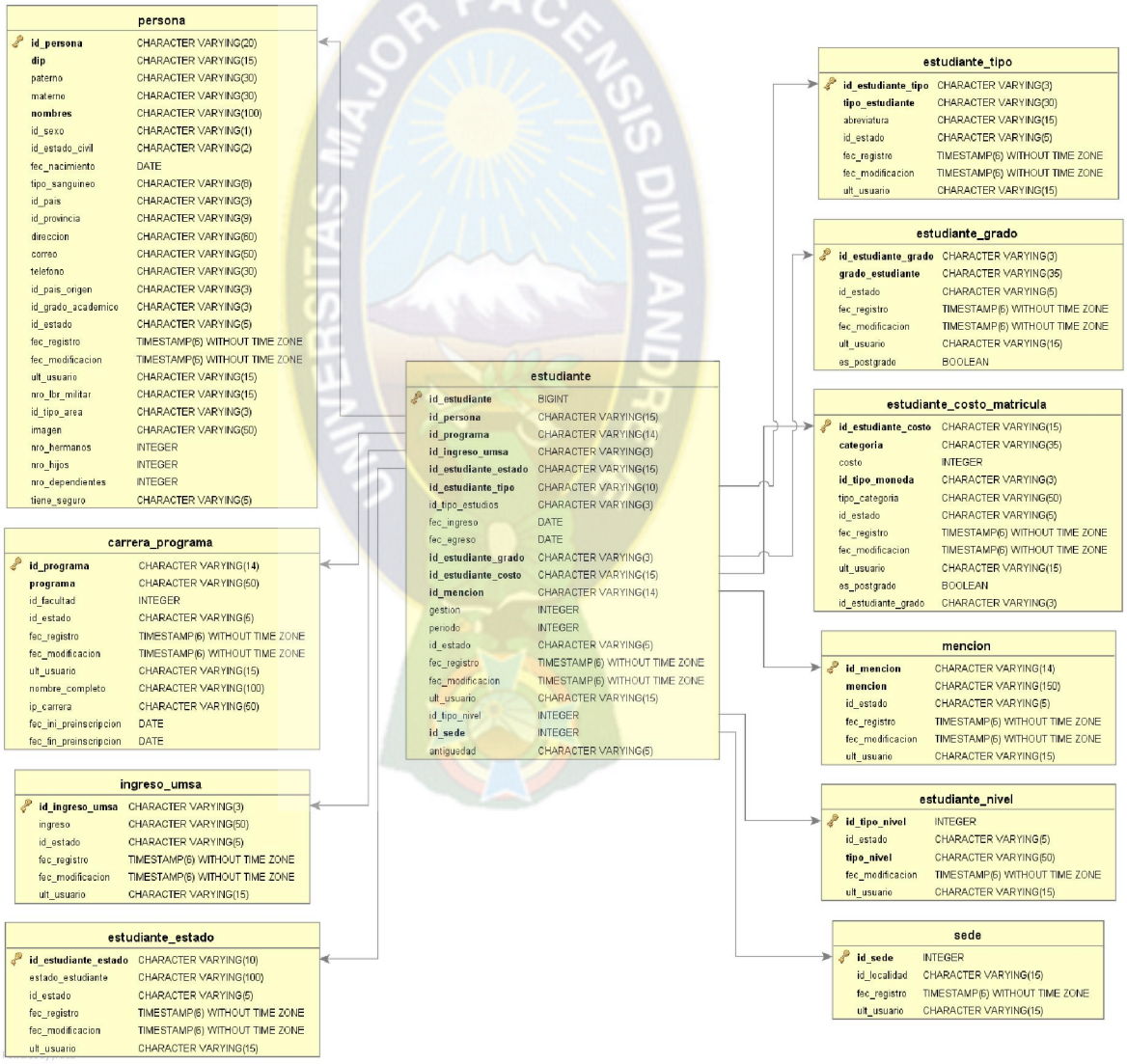


Figura 13: Modelo Relacional SRU, Estudiante
Fuente: BD Relacional Fuente de Origen Postgres

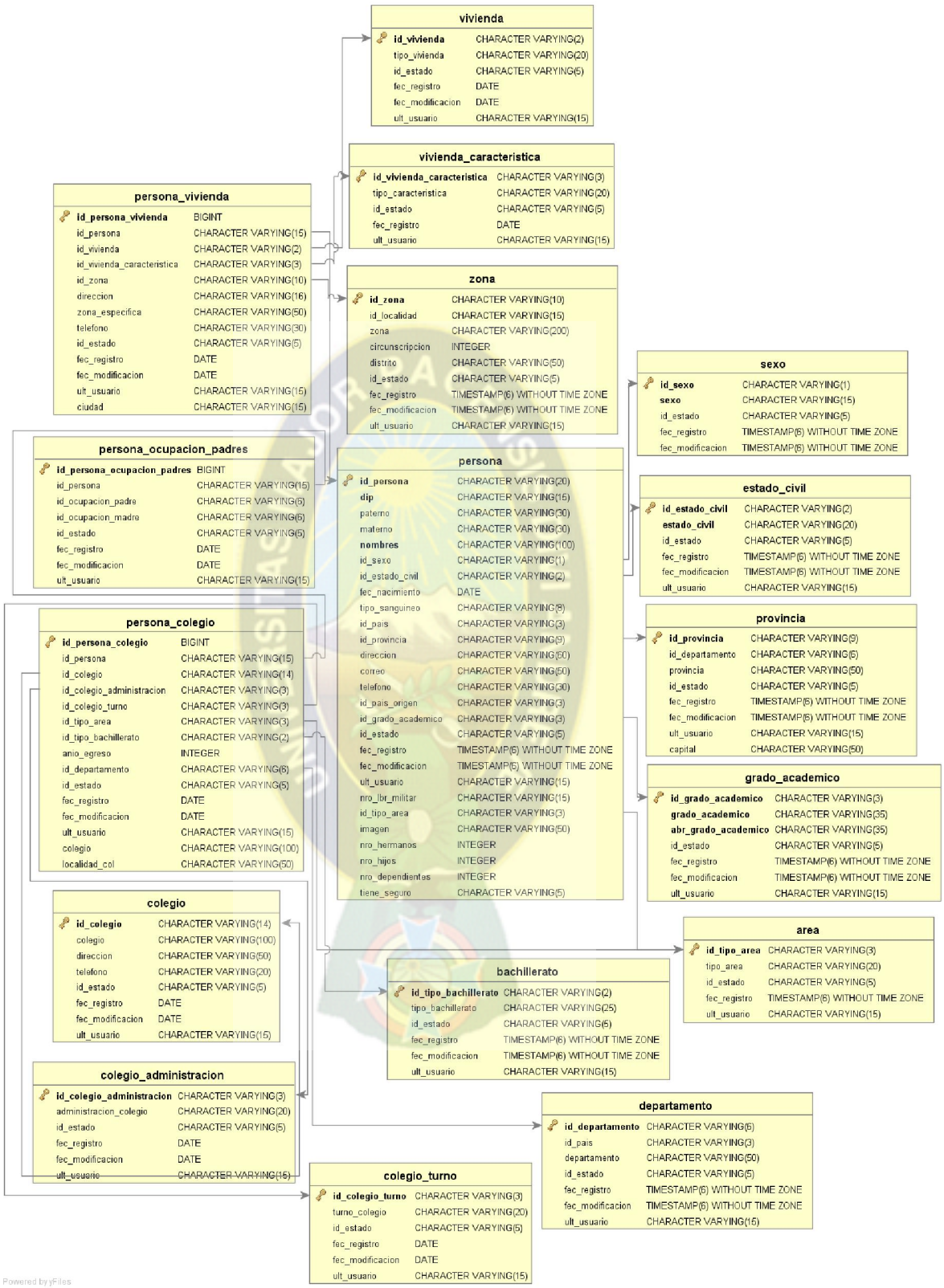


Figura 14: Modelo Relacional SRU, Persona
Fuente: BD Relacional Fuente de Origen Postgres

Fuente Origen de Datos: Aplicación Informática OK

Se analiza la estructura de la aplicación OK.

Aplicación Informática OK
Código de la variable género, estado civil
Código de la variable del departamento de nacimiento
Código de la variable de país nacionalidad
Código de las variables ocupación padre y madre
Código de la variable propiedad vivienda
Código de las variables vivienda características, vivienda zona
Código de la variable administración colegio
Código de las variable colegio área, turno
Código de la variable lugar de egreso de colegio
Código de la variable tipo bachillerato
Año de egreso del colegio
Código de la variable sector laboral
Código de la variable jornada laboral
Código de la variable trabaja como
nombre (paterno– materno -nombres)
carnet identidad, fecha de nacimiento
Dirección, Celular, Correo electrónico
Número de hermanos en la U.M.S.A.
Número de hijos y dependientes
numero registro universitario
Código de la variable facultad y carrera
Código de la variable modalidad de ingreso a la U.M.S.A.
Código de la variable de tipo de alumno
Código de la variable de nivel universitario
Código de la variable del tipo de ingreso del alumno.
Código de la variable categoria de matrícula universitaria
Código de la variable mención
Número del promes
Sede

Figura 15: Estructura de Datos en la Aplicación OK
Fuente: Elaboración Propia

4.3.3 Capa de Integración de Datos

En la presente capa se desarrolla el proceso de extracción, transformación y carga de datos, desde los orígenes de datos hasta el data warehouse, como se puede observar en la Figura 16 este proceso se aplicará mediante el mapeo de integración de datos el cual permitirá alimentar al almacén de datos.

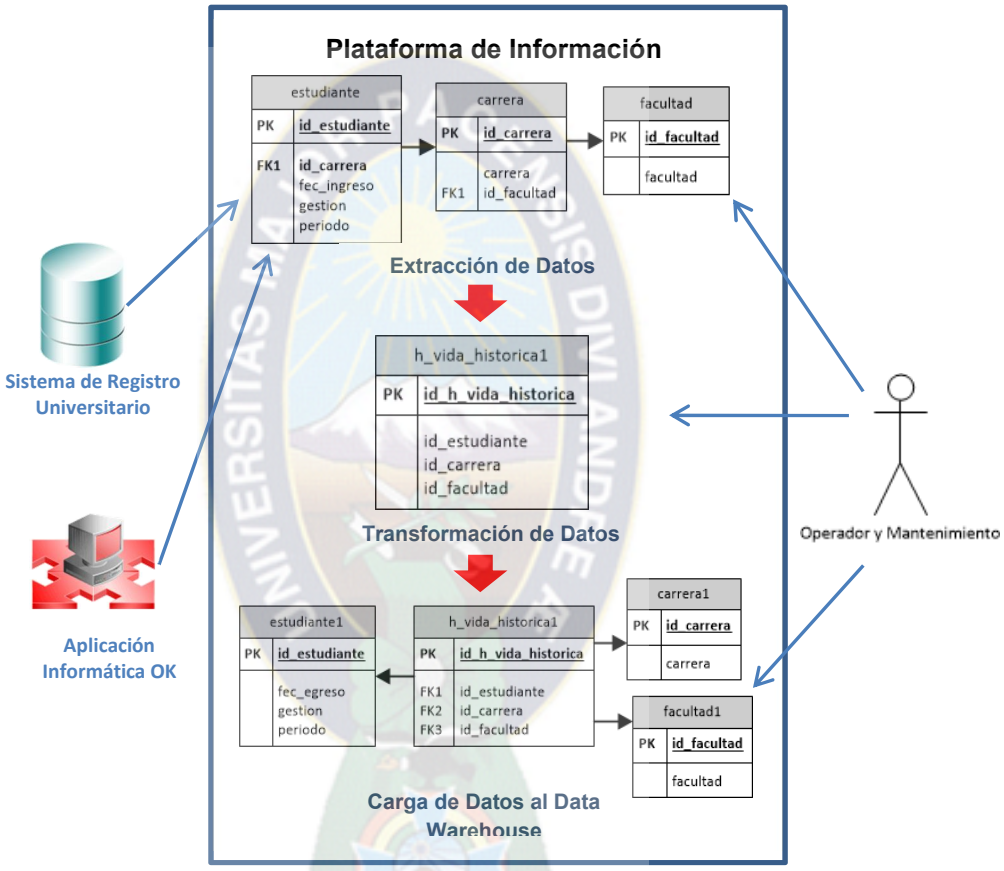


Figura 16: Proceso de Extracción, Transformación y Carga de datos (ETL)
Fuente: Elaboración Propia

En la Figura 17 se muestra el script del proceso de extracción, transformación y carga de datos, para realizar estas tres fases se creó tres espacios de trabajo en el repositorio de la herramienta Oracle Warehouse Builder las cuales son: Source, Stage y Star.

En el repositorio del Source es donde se almacena todos los datos extraídos desde las fuentes de datos del Sistema de Registro Universitario y Aplicación OK. En el espacio

de Stage es donde se realiza los mapeos de transformación de los datos, realizando operaciones de filtros, uniones y clasificadores de datos.

En el espacio de trabajo Star es donde se construye el modelo del esquema estrella para poder cargar los datos a la base de datos del Data Warehouse, con relación a la cantidad de registros que tiene que cargar a nuestro almacén de datos son: 232.068 registros relacionados con datos socioeconómicos de las gestiones 2010-2015, 1.617.245 registros relacionados con datos académicos de las gestiones 1990-2015, 55.959 registros relacionados con datos de seguro universitario y 33.110 registros con datos de postulantes a título, estas dos últimas de las gestiones 2008.

```
-- "JOINER_p" main
BEGIN
  IF NOT "H_DATOS_SOCIOECONOMICOS_PC1_St" THEN
       IF NOT "H_DATOS_SOCIOECONOMICOS_PC1_St" THEN
        "H_DATOS_SOCIOECONOMICOS_PC1_id":= (
          p_source=>get_source_name, p_source_uoid=>get_source_uoid,
          p_target=>'"H_DATOS_SOCIOECONOMICOS_PC1"',
          p_target_uoid=>'A6821BF97B3549118AB3830CEC088225');
        get_audit_detail_id := "H_DATOS_SOCIOECONOMICOS_PC1_id";
  get_audit_detail_type_id := OWBSYS.WB_RT_MAPAUDIT.register_audit_detail_type(
          p_parent_object_name=>'DIM_PERSONA', -- Operator DIM_PERSONA
          p_parent_object_type=>'Tabla', -- Tabla DIM_PERSONA
          p_object_name=>'DIM_PERSONA',
          p_object_uoid=>'310422C809F84F47ACB730E05AC52491', -- Tabla DIM_PERSONA
          p_object_type=>'Tabla', p_location_uoid=>'233A78AB9E2E4D52A17A282DBD1F795D' -- Location STAGE_LOCATION1);
/*Transformación de datos*/
get_target_name := '"H_DATOS_SOCIOECONOMICOS_PC1"';
        IF NOT "H_DATOS_SOCIOECONOMICOS_PC1_St" AND NOT get_buffer_done(get_buffer_done_index) THEN
          BEGIN
            get_rows_processed:= true; -- Set to indicate that some row data was processed (for debugger)
              error_stmt := SUBSTRB('"H_DATOS__0_ID_PERSO"("H_DATOS_SOCIOECONOMICOS_PC1_i") :=
                "JOINER_87_ID_PERSONA"("JOINER_i");',0,2000);
        error_column:= SUBSTRB ('"H_DATOS__0_ID_PERSO"',0,80);
        BEGIN error_value:= SUBSTRB ("JOINER_87_ID_PERSONA"("JOINER_i"),0,2000);
         EXCEPTION  WHEN OTHERS THEN error_value:= '*';
        END;
        IF NOT get_use_hc THEN
          "H_DATOS__0_ID_PERSO"("H_DATOS_SOCIOECONOMICOS_PC1_i") := "JOINER_87_ID_PERSONA"("JOINER_i");
        ELSIF get_row_status THEN   "SV_H_DATOS__0_ID_PERSO" :=  "JOINER_87_ID_PERSONA"("JOINER_i");
        END IF;
        error_stmt := SUBSTRB('"H_DATOS__1_ID_SEXO"("H_DATOS_SOCIOECONOMICOS_PC1_i") :=
--Bulk: types for collection variables
TYPE "T_JOINER_87_ID_PERSONA$1" IS TABLE OF VARCHAR2 (15) INDEX BY BINARY_INTEGER;
--Bulk: intermediate scalar variables
"SV_JOINER_87_ID_PERSONA$1" VARCHAR2 (15);
--Bulk: intermediate collection variables
"JOINER_87_ID_PERSONA$1" "T_JOINER_87_ID_PERSONA$1";
--Procedimiento de Transformación
PROCEDURE "JOINER_ES$1"(error_index IN NUMBER) IS
BEGIN
  OWBSYS.WB_RT_MAPAUDIT.error_source(
p_table=>SUBSTR('"DIM_PERSONA","DIM_SEXO","DIM_ESTADO_CIVIL","DIM_PAIS_NACIONALIDAD","DIM_PAIS_ORIGEN",
                "DIM_PROVINCIA","DIM_DEPARTAMENTO_NACIMIENTO","DIM_GRADO_ACADEMICO","DIM_OCUPACION_PADRE","DIM_OCUPACI
                ON_MADRE","DIM_VIVIENDA","DIM_VIVIENDA_CARACTERISTICA","DIM_VIVIENDA_ZONA","DIM_COLEGIO","DIM_COLEGIO_ADMIN
                ISTRACION","DIM_COLEGIO_TURNO","DIM_COLEGIO_AREA","DIM_COLEGIO_DEPARTAMENTO","DIM_COLEGIO_BACHILLERATO",
                "DIM_TRABAJO_TIEMPO","DIM_TRABAJO_TIPO","DIM_TRABAJO_OCUPACION","H_DATOS_SOCIOECONOMICOS"',0,80),
   p_column=>SUBSTR ('ID_PERSONA',0,80),  p_value=>SUBSTRB("JOINER_87_ID_PERSONA$1"(error_index),0,2000)  );
--Procedimiento de Carga de Datos desde el espacio Source
PROCEDURE "JOINER_RD$1" IS
BEGIN
  IF NOT get_read_success THEN get_abort:= TRUE; END IF;   IF get_abort OR get_abort_procedure THEN RETURN; END IF;
  BEGIN
    "JOINER_87_ID_PERSONA$1".DELETE; FETCH   "JOINER_c$1"
    BULK COLLECT INTO       "JOINER_87_ID_PERSONA$1"       LIMIT get_bulk_size;
    get_total_processed_rowcount:= get_total_processed_rowcount + "JOINER_87_ID_PERSONA$1".COUNT;
    IF "JOINER_c$1"%NOTFOUND AND "JOINER_87_ID_PERSONA$1".COUNT = 0 THEN      RETURN;    END IF;
```

Figura 17: Script del proceso de ETL, herramienta OWB
Fuente: Script extraído de la herramienta OWB

4.3.4 Capa de Almacén de Datos

Se realiza la construcción de los Data Marts en base a desnormalización de tablas entidad – relación y datos disponibles en las fuentes de información.

- **Diccionario de Datos de Dimensiones:** A continuación, se describirá las estructuras de las tablas hechos y dimensiones principales, las dimensiones restantes se describen en el Anexo C.

  La tabla Dim_Persona almacena los datos personales de todos los estudiantes matriculados en la U.M.S.A.

Dim_Persona		
Campo	**Tipo**	**Descripción**
id_persona	varchar(15)	Código que identifica a una persona.
carnet_identidad	varchar(15)	Número que identifica a una persona.
paterno	varchar(30)	Apellido paterno de una persona.
materno	varchar(30)	Apellido materno de una persona.
nombres	varchar(100)	Nombres de una persona.

Tabla 7: Diccionario de Dim_Persona
Fuente: Elaboración Propia

  La tabla Dim_Estudiante almacena información académica de los estudiantes matriculados de la U.M.S.A.

Dim_Estudiante		
Campo	**Tipo**	**Descripción**
id_estudiante	number	Código que identifica a un estudiante dentro de la UMSA.
fec_ingreso	date	Fecha de ingreso a la universidad de un estudiante.
gestión	number	Gestión o año de ingreso de un estudiante a la universidad.
periodo	number	Periodo del año en el que ingresa un estudiante a la universidad.
antiguedad	varchar(5)	Describe el estado de un estudiante si es nuevo o antiguo.

Tabla 8: Diccionario de Dim_Estudiante
Fuente: Elaboración Propia

- **Diccionario de Datos de Hechos:** La tabla H_Datos_Socioeconomicos almacena información social y económica de los estudiantes matriculados de la U.M.S.A.

H_Datos_Socioeconomicos		
Campo	**Tipo**	**Descripción**
id_h_datos_socioeconomicos	number	Código que identifica los datos socioeconómicos de un estudiante.
id_persona	varchar(15)	Relación con la dimensión Dim_Persona.
id_sexo	varchar(1)	Relación con dimensión Dim_Genero.
id_estado_civil	varchar(2)	Relación con dimensión Dim_Estado_Civil.
id_pais_nacionalidad	varchar(3)	Relación con dimensión Dim_Pais_Nacionalidad.
id_pais_origen	varchar(3)	Relación con dimensión Dim_Pais_Origen.
id_provincia	varchar(9)	Relación con dimensión Dim_Provincia.
id_dpto_nacimiento	varchar(6)	Relación con dimensión Dim_Dpto_Nacimiento.
id_grado_academico	varchar(3)	Relación con dimensión Dim_Grado_Academico.
id_ocup_padre	varchar(6)	Relación con dimensión Dim_Ocupacion_Padre.
id_ocup_madre	varchar(6)	Relación con dimensión Dim_Ocupacion_Madre.
id_vivienda	varchar(2)	Relación con dimensión Dim_Vivienda.
id_viv_caracteristica	varchar(3)	Relación con dimensión Dim_Viv_Caracteristica.
id_viv_zona	varchar(10)	Relación con dimensión Dim_Vivienda_Zona.
id_colegio	varchar(14)	Relación con dimensión Dim_Colegio.
id_col_administracion	varchar(3)	Relación con dimensión Dim_Col_Administracion.
id_col_turno	varchar(3)	Relación con dimensión Dim_Colegio_Turno.
id_col_area	varchar(3)	Relación con dimensión Dim_Colegio_Area.
id_col_departamento	varchar(6)	Relación con dimensión Dim_Col_Departamento.
id_col_bachillerato	varchar(2)	Relación con dimensión Dim_Col_Bachillerato.
id_trab_tiempo	varchar(3)	Relación con dimensión Dim_Trabajo_Tiempo.
id_trab_tipo	varchar(3)	Relación con dimensión Dim_Trabajo_Tipo.
id_trab_ocupacion	varchar(6)	Relación con dimensión Dim_Trab_Ocupacion.
per_fecha_nacimiento	date	Fecha de nacimiento del estudiante.
per_tipo_sanguineo	varchar(8)	Tipo sanguíneo del estudiante.
per_direccion	varchar(100)	Dirección del estudiante.
per_telefono	varchar(30)	Teléfono del estudiante.
per_correo	varchar(50)	Correo electrónico del estudiante.
per_nro_hermanos	number	Número de hermanos del estudiante.
per_nro_hijos	number	Número de hijos del estudiante.
per_nro_dependientes	number	Número de personas dependientes del estudiante.
viv_ciudad	varchar(15)	Ciudad de ubicación la vivienda del estudiante.
col_anio_egreso	number	Año de egreso del colegio del estudiante.
col_localidad	varchar(50)	Localidad del colegio.

Tabla 9: Diccionario de H_Datos_Socioeconomicos
Fuente: Elaboración Propia

La tabla H_Vida_Historica almacena información sobre las características académicas de los estudiantes dentro de la U.M.S.A.

H_Vida_Historica		
Campo	**Campo**	**Campo**
id_h_vida_historica	number	Código que identifica los datos académicos históricos del estudiante.
id_estudiante	number	Relación con la dimensión Dim_Estudiante.
id_persona	varchar(15)	Relación con la dimensión Dim_Persona.
id_programa	varchar(14)	Relación con dimensión Dim_Carrera_Programa.
id_facultad	number	Relación con dimensión Dim_Facultad.
id_ingreso_umsa	varchar(3)	Relación con dimensión Dim_Ingreso_UMSA.
id_estudiante_estado	varchar(10)	Relación con dimensión Dim_Estudiante_Estado.
id_estudiante_grado	varchar(3)	Relación con dimensión Dim_Estudiante_Grado.
id_estudiante_nivel	number	Relación con dimensión Dim_Estudiante_Nivel.
id_estudiante_tipo	varchar(3)	Relación con dimensión Dim_Estudiante_Tipo.
id_estudiante_costo_m	varchar(15)	Relación con dimensión Dim_Estudiante_Costo_M.
id_mencion	varchar(14)	Relación con dimensión Dim_Mencion.
id_sede_localidad	varchar(15)	Relación con dimensión Dim_Sede_Localidad.
num_matricula	number	Numero único de la matrícula universitaria por gestión.
gestión_matriula	number	Gestión de la matrícula universitaria.
periodo_matricula	number	Periodo de la matrícula en una gestión.
costo_matricula	number	Costo de la matricula según tipo de estudiante.
fec_reg_matricula	date	Fecha de registro de la matrícula.
id_estado_matricula	varchar(5)	Relación con dimensión Dim_Estado_Matricula.

Tabla 10: Diccionario de H_Vida_Historica
Fuente: Elaboración Propia

La tabla H_Seguro_Universitario almacena información de todos los estudiantes que cuentan con el seguro universitario PROMES.

H_Seguro_Universitario		
Campo	**Tipo**	**Descripción**
id_h_seguro_universitario	number	Código que identifica los datos de seguro universitario del estudiante.
id_persona	varchar(15)	Relación con la dimensión Dim_Persona.
id_estudiante	number	Relación con la dimensión Dim_Estudiante.
id_programa	varchar(14)	Relación con dimensión Dim_Carrera_Programa.
id_facultad	number	Relación con dimensión Dim_Facultad.
id_sede_localidad	varchar(15)	Relación con dimensión Dim_Sede_Localidad.
num_afiliacion	varchar(15)	Numero de afiliación del seguro.
fecha_afiliacion	date	Fecha de afiliación.
num_deposito	varchar(15)	Periodo de la matrícula en una gestión.
aux_nombres	varchar(100)	Costo de la matricula según tipo de estudiante.
aux_telefonos	varchar(30)	Fecha de registro de la matrícula.
aux_celular	varchar(20)	Relación con dimensión Dim_Estado_Matricula.
alergias	varchar(300)	Alergias del estudiante asegurado.
gestión_cedula	number	Gestión de la cedula del seguro universitario.
num_promes	varchar(15)	Numero único del estudiante asegurado.
fecha_registro_cedula	date	Fecha de registro de la cedula del seguro.

Tabla 11: Diccionario de H_Seguro_Universitario
Fuente: Elaboración Propia

La tabla H_Postulantes_Titulo almacena información de todos los estudiantes que postulan a la titulación.

H_Postulantes_Titulo		
Campo	**Tipo**	**Descripción**
id_h_postulantes_titulo	number	Código que identifica los datos de los estudiantes que postulan a un título.
id_persona	varchar(15)	Relación con dimensión Dim_Persona.
id_tipo_titulo	varchar(2)	Relación con dimensión Dim_Tipo_Titulo.
id_tipo_titulacion	varchar(6)	Relación con dimensión Dim_Tipo_Titulacion.
id_programa	varchar(14)	Relación con dimensión Dim_Carrera_Programa.
id_facultad	number	Relación con dimensión Dim_Facultad.
id_sede_localidad	varchar(15)	Relación con dimensión Dim_Sede_Localidad.
num_titulo	number	Número del título que se otorga a un estudiante.
gestión_titulo	number	Gestión del título emitido.
anio_egreso	number	Año de egreso del estudiante.
anio_titulacion	number	Año de titulación de un estudiante profesional.
fecha_expedicion	date	Fecha de expedición del título.

Tabla 12: Diccionario de H_Postulantes_Titulo
Fuente: Elaboración Propia

- **Modelo de las Tablas Hecho y Dimensiones aplicando el Esquema en Estrella**

Las características de usar un modelo en estrella para la construcción del Data Warehouse, es excluir las uniones entre las tablas cuando se realizan consultas, administrando así un mejor tiempo de respuesta, de igual manera posee un diseño fácilmente modificable y comprensible para el usuario que visualiza y manipula los datos, simplificando así el análisis, esta estructura también facilita la interacción con las herramientas de consulta y análisis.

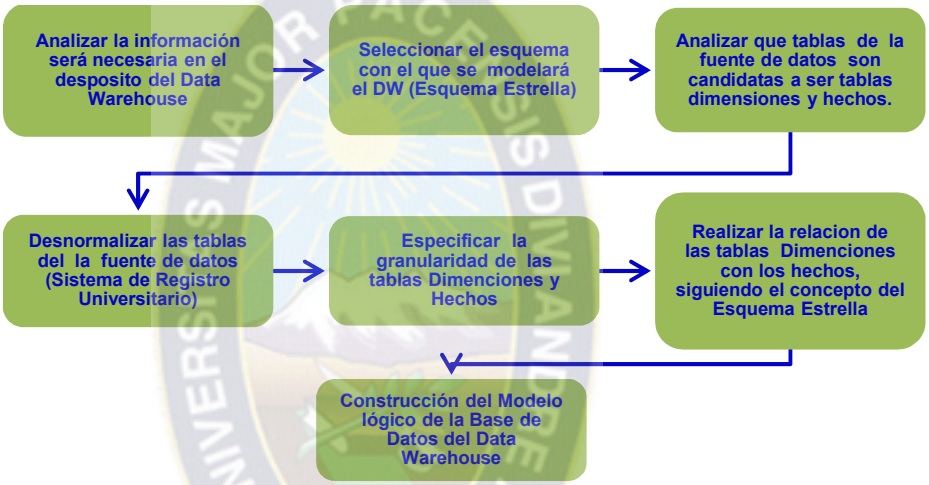


Figura 18: Proceso de Construcción del Esquema Estrella del Data Warehouse
Fuente: Elaboración Propia

Aplicando estos procesos al presente proyecto se pudo hallar 36 dimensiones y 4 tablas hecho según requerimientos, como resultado de esto se diseñó cuatro esquemas estrellas que modelan el Data Warehouse de la Plataforma Estadística.

A continuación se muestra el diseño lógico del Data Warehouse que está conformado por dimensiones y hechos que se encuentran relacionados, vale mencionar que las tablas hecho pueden compartir dimensiones para poder relacionar un esquema con otro, obteniendo así información más detallada según requerimientos analizados .

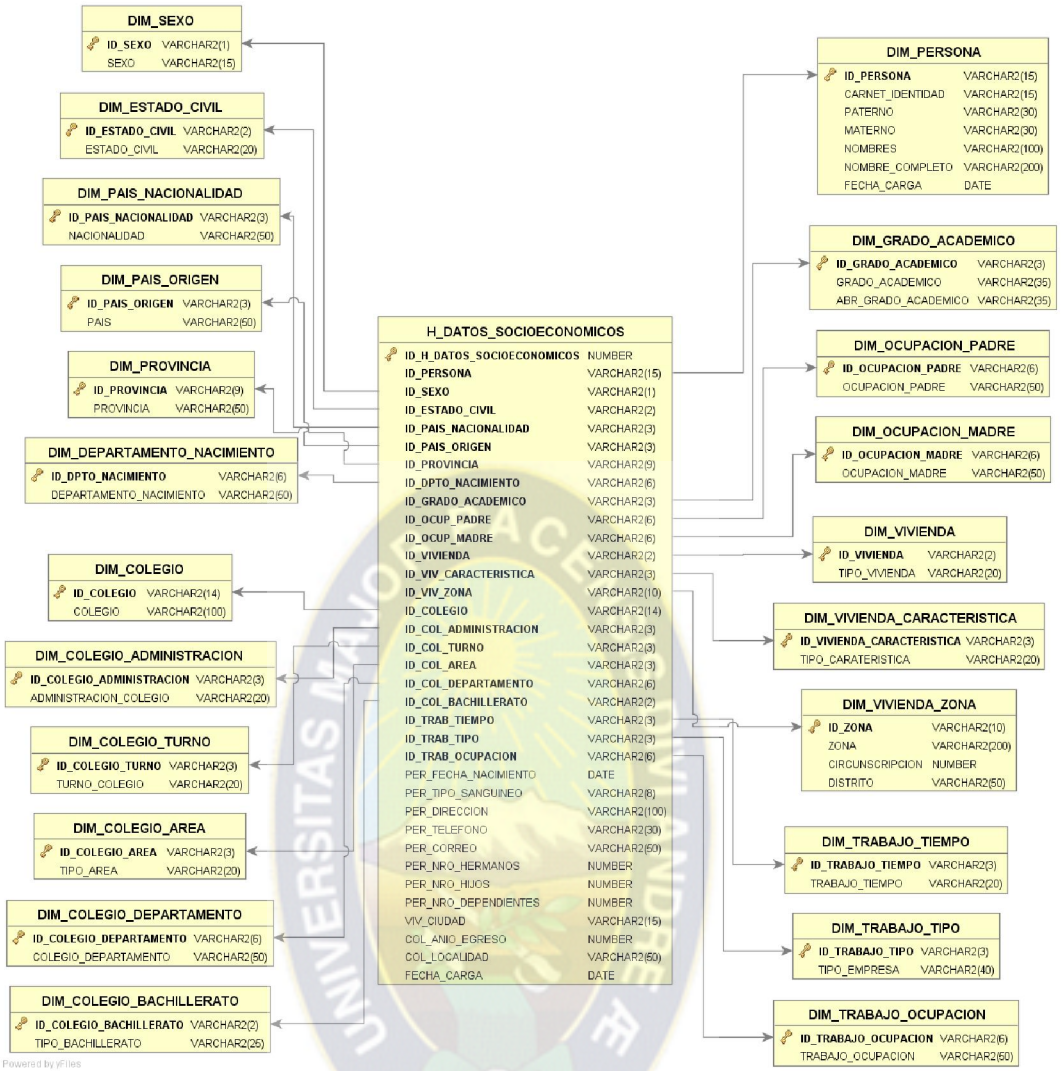


Figura 19: Modelo en Estrella H_Datos_Socioeconomicos
Fuente: BD del Data Warehouse en Oracle

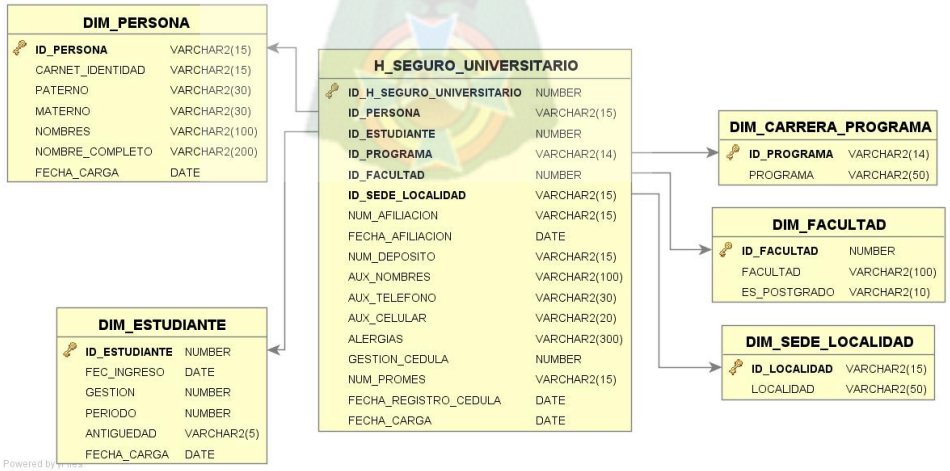


Figura 20: Modelo en Estrella H_Seguro_Universitario
Fuente: BD del Data Warehouse en Oracle

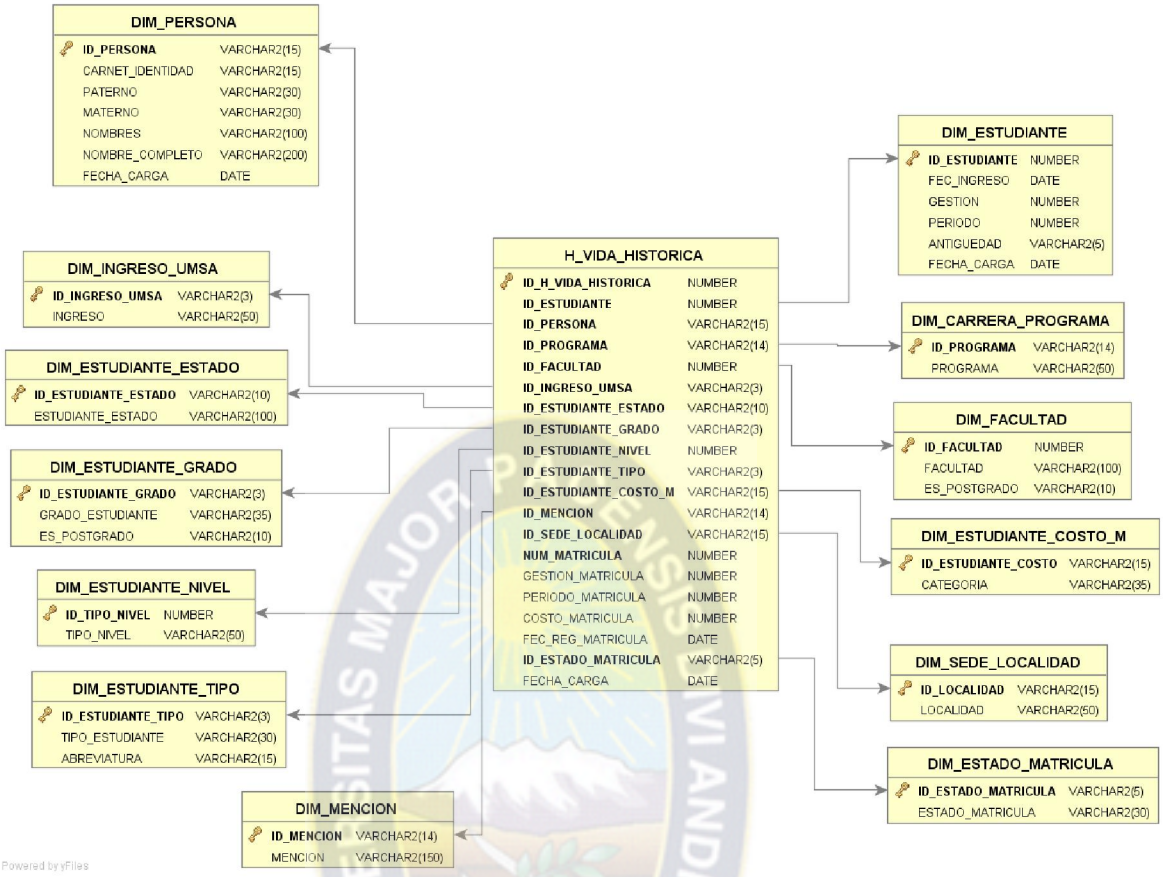


Figura 21: Modelo en Estrella H_Vida_Historica
Fuente: BD del Data Warehouse en Oracle

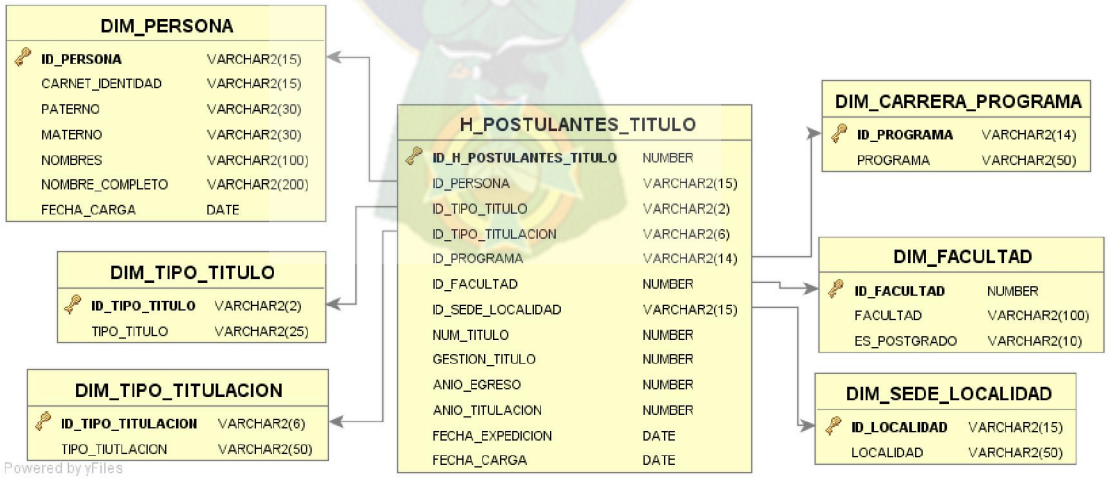


Figura 22: Modelo en Estrella H_Postulantes_Titulo
Fuente: BD del Data Warehouse en Oracle

4.3.5 Capa de Agregación de Datos

Fase de construcción de cubos OLAP que permite realizar la agregación de dato a nivel de requerimiento del usuario, a fin de brindar resultados con el nivel de detalle y agregado requerido de información.

La Figura 23 modela el proceso explotación de datos Socioeconomicos, en esta etapa se organiza y agrupa los datos según requerimientos de las Tablas 4 y Tabla 5, tiene como principal fuente de origen de información el Data Warehouse (Esquema estrella H_Datos_Socioeconomicos) de la Plataforma de Información Estadística.

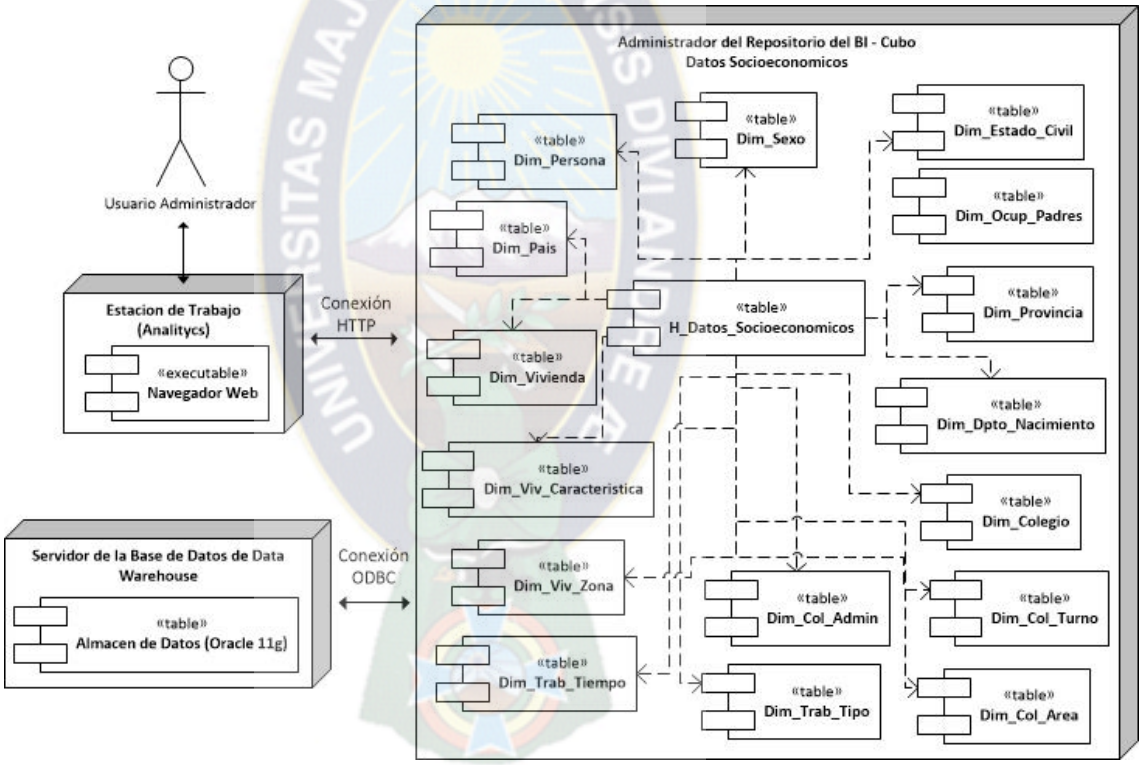


Figura 23: Hecho H_Datos_Socioeconomicos
Fuente: Elaboración Propia

La Figura 24 modela el proceso de explotación de datos académicos, en esta etapa se organiza y agrupa los datos según requerimiento de la Tabla 6, tiene como principal fuente de origen de información el Data Warehouse (Esquema estrella H_Vida_Historica) de la Plataforma de Información Estadística.

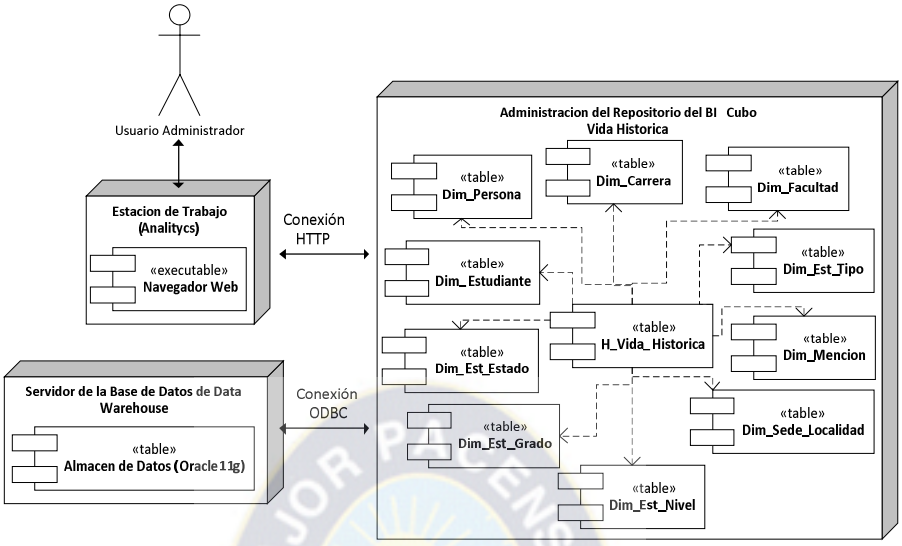


Figura 24: Hecho H_Vida_Historica
Fuente: Elaboración Propia

En la etapa de la explotación de datos desde el Data Warehouse se desarrolla la estructura de los Cubos OLAP, almacenando en estas consultas específicas de cada requerimiento.

La Figura 25 muestra la agregación de datos al Cubo OLAP H_Datos_Socioeconomicos, el cual almacena información socioeconómica de los estudiantes, este se relaciona con las dimensiones a través de las llaves primarias de cada dimensión, adquiriendo así más detalle de información.

Figura 25: Cubo OLAP de datos Socioeconomicos
Fuente: Elaboración Propia

La Figura 26 muestra la agregación de datos al Cubo OLAP H_Vida_Historica, el cual almacena información datos académicos de los estudiantes.

Figura 26: Cubo OLAP de datos Antecedentes Académicos
Fuente: Elaboración Propia

Como se puede observar en las Figuras 25 y 26, ambos cubos se relacionan a través del código de la dimensión Persona (ID_PERSONA), esto se realiza para cruzar datos socioeconómicos con antecedentes académicos, obteniendo así un gran detalle de información de cada estudiante.

4.3.6 Capa de Presentación de la Información

Figura 27: Interacción de los usuarios con la plataforma estadística
Fuente: Elaboración Propia

La Figura 27 nos muestra las características de modelaje de la interacción de los usuarios con la plataforma estadística, desarrollando de esta manera el proceso de

análisis del flujo de datos solicitados a los cubo OLAP, como resultado de este proceso se obtiene repostes estadísticos requeridos por el usuarios administrador de la información.

Diseño de la Infraestructura

Se definió la siguiente arquitectura física para el almacén de los datos de la Plataforma Estadística.

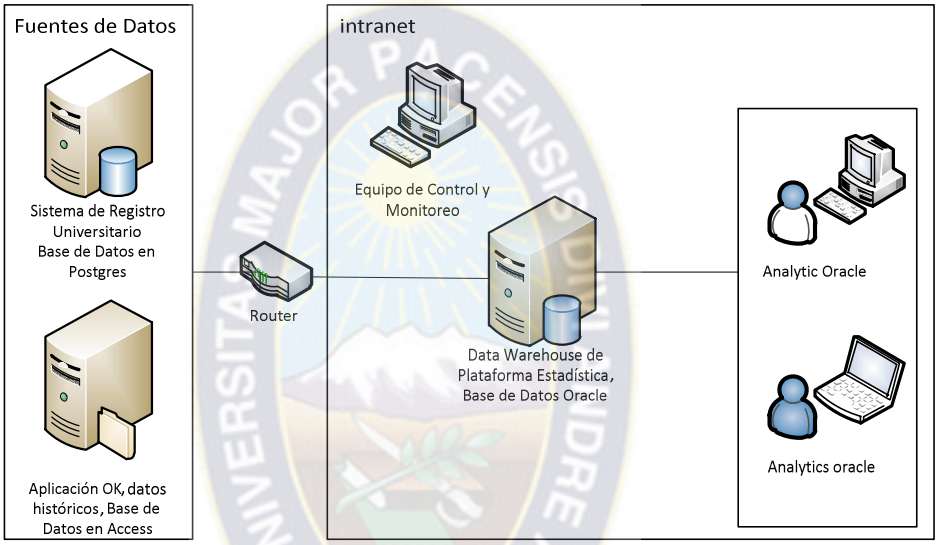


Figura 28: Arquitectura Física del Almacén de Datos
Fuente: Elaboración Propia

4.4 Implementación

En la fase de la implementación según la metodología se lleva a cabo los siguientes pasos: Extracción de los Datos de las fuentes de origen, transformación y carga de los datos, el proceso de explotación de datos desde el Data Warehouse.

Con la finalización de esta fase se obtendrá la Plataforma de Información Estadística disponible para su uso por parte de los usuarios finales.

4.4.1 Implementación Capa de Integración de Datos

Para realizar el proceso de ETL se utilizó la herramienta Oracle Warehouse Builder.

- **Extracción de Datos**

Para el proceso de extracción se realizó los siguientes pasos:

1. Análisis de Datos requeridos desde las fuentes de origen.
2. Sea ejecuto consultas sql para obtener los datos desde las fuentes de origen, como resultado de este paso se fueron construyendo las dimensiones y las tablas hecho.

La extracción de datos se realizó a través de consultas Sql, en la Figura 29 se detalla el query para armar la tabla de hechos de datos socioeconómicos.

```
select distinct 0 as id_h_datos_socioeconomicos, P.id_persona,
(case when (S.id_sexo like '' or S.id_sexo like ' ' or S.id_sexo is null) then 'N' else S.id_sexo end),
(case when (E.id_estado_civil like '' or E.id_estado_civil like ' ' or E.id_estado_civil is null) then 'NR' else E.id_estado_civil end),
(case when (PN.id_pais like '' or PN.id_pais like ' ' or PN.id_pais is null) then 'OT' else PN.id_pais end) as Pais_Nacionalidad,
(case when (PO.id_pais like '' or PO.id_pais like ' ' or PO.id_pais is null) then 'OT' else PO.id_pais end) as Pais_Origen,
(case when (PR.id_provincia like '' or PR.id_provincia like ' ' or PR.id_provincia is null) then 'OT' else PR.id_provincia end),
(case when (PR.id_departamento like '' or PR.id_departamento like ' ' or PR.id_departamento is null)
then 'OT' else PR.id_departamento end), t2.id_ocupacion_padre end,
(case when (t3.id_vivienda like '' or t3.id_vivienda like ' ' or t3.id_vivienda is null) then 'NR' else t3.id_vivienda end),
(case when (t3.id_vivienda_caracteristica like '' or t3.id_vivienda_caracteristica like ' ' or t3.id_vivienda_caracteristica is null)
then 'NR' else t3.id_vivienda_caracteristica end),
(case when (t3.id_zona like '' or t3.id_zona like ' ' or t3.id_zona is null) then 'NR' else t3.id_zona end),
(case when (t4.id_colegio like '' or t4.id_colegio like ' ' or t4.id_colegio is null) then 'OTR' else t4.id_colegio end),
(case when (t4.id_colegio_administracion like '' or t4.id_colegio_administracion like ' ' or t4.id_colegio_administracion is null)
then 'NR' else t4.id_colegio_administracion end),
(case when (t4.id_colegio_turno like '' or t4.id_colegio_turno like ' ' or t4.id_colegio_turno is null) then 'NR' else
t4.id_colegio_turno end),
(case when (t3.ciudad like '' or t3.ciudad like ' ' or t3.ciudad is null) then 'NR' else t3.ciudad end) as Vivienda_Ciudad,
(case when (t4.anio_egreso is null) then '0' else t4.anio_egreso end) as Colegio_Año_Egreso,
(case when (t4.localidad_col like '' or t4.localidad_col like ' ' or t4.localidad_col is null) then 'NR' else t4.localidad_col end)
as Colegio_Localidad, current_date as fecha_carga from persona P
LEFT JOIN sexo as S ON(S.id_sexo like P.id_sexo)
LEFT JOIN estado_civil E ON(E.id_estado_civil like P.id_estado_civil)
LEFT JOIN pais PN ON(PN.id_pais like P.id_pais)
LEFT JOIN pais PO ON(PO.id_pais like P.id_pais_origen)
LEFT JOIN provincia PR ON(PR.id_provincia like P.id_provincia)
LEFT JOIN ( select x1.id_persona,x1.id_ocupacion_padre,x1.id_ocupacion_madre
                from (select max(id_persona_ocupacion_padres) as max_id_persona_ocupacion_padres, id_persona, count
                        (id_persona) as nro
              from persona_ocupacion_padres
              group by id_persona, id_ocupacion_padre      order by nro) as t1, persona_ocupacion_padres x1
             where t1.max_id_persona_ocupacion_padres=x1.id_persona_ocupacion_padres and      t1.id_persona
                        like x1.id_persona
                          order by x1.id_persona) as t2  ON(t2.id_persona = P.id_persona)
LEFT JOIN (select x1.id_persona,x1.id_vivienda,x1.id_vivienda_caracteristica,x1.id_zona,x1.ciudad
              from (select max(id_persona_vivienda), id_persona, count (id_persona) as nro
                        from persona_vivienda
                        group by id_persona    order by nro) as t1, persona_vivienda as x1
              where t1.max = x1.id_persona_vivienda and t1.id_persona like x1.id_persona
              order by x1.id_persona) as t3
ON(t3.id_persona = P.id_persona)
LEFT JOIN (select x1.id_persona,
x1.id_colegio,x1.id_colegio_administracion,x1.id_colegio_turno,x1.id_tipo_area,x1.id_tipo_bachillerato,x1.id_departamento,
              from (select max(id_persona_colegio),id_persona,count(id_persona)as nro
                   from persona_colegio
                        group by id_persona    order by nro) as t1, persona_colegio x1
              where t1.max=x1.id_persona_colegio and t1.id_persona like x1.id_persona
              order by x1.id_persona) as t4
ON(t4.id_persona=P.id_persona)
where P.id_estado like 'A' order by p.id_persona
```

Figura 29: Consulta Sql para la extracción de datos
Fuente: Elaboración Propia

En la Figura 30 mostramos los resultados de las consultas de la Figura 29, obteniendo datos para nuestras tablas candidatas a ser dimensiones y hechos, estos datos son administrados desde el espacio de trabajo Source como se muestra en la Figura 30.

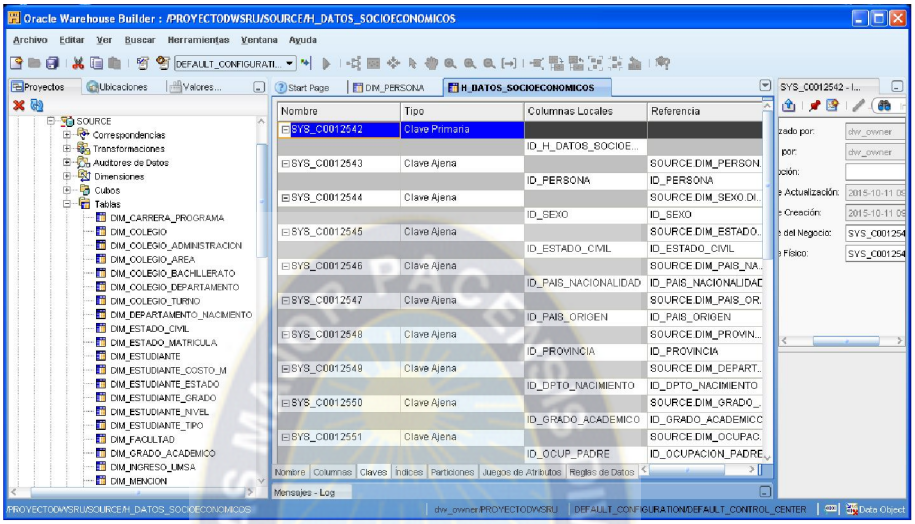


Figura 30: Tablas hecho y dimensiones construidas a partir de consultas
Fuente: Herramienta OWB

- **Transformación de Datos**

  En la Figura 31 muestra el desarrollo del mapeo realizando filtros y uniones de las tablas almacenadas en el espacio de trabajo Source, la figura muestra el mapeo de la tabla H_Datos_Socioeconomicos.

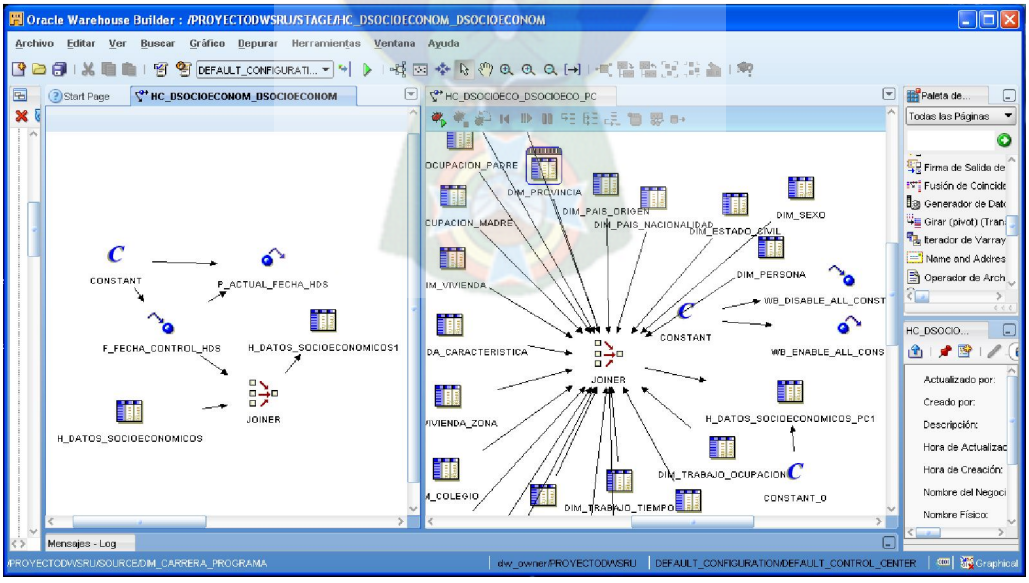


Figura 31: Mapeo de Transformación del Hecho Datos Socioeconomicos
Fuente: Herramienta OWB

En la Figura 32 podemos observar el código de transformación para obtener la tabla de datos socioeconómicos, se realiza el mapeado de los datos desde el espacio de trabajo Source al Stage, la herramienta que se utiliza es el joiner la cual nos permite incluir condiciones entre los datos de las dimensiones a través de sus llaves primarias.

```
create or replace PACKAGE BODY "HC_DSOCIOECO_DSOCIOECO_PC" AS
-- Define the package (for debugger)
-- "JOINER_c" Cursor declaration
CURSOR "JOINER_c" IS
 SELECT
/* JOINER.OUTGRP1 */
  "DIM_PERSONA"."ID_PERSONA" "ID_PERSONA",   "H_DATOS_SOCIOECONOMICOS"."FECHA_CARGA" "FECHA_CARGA_1$1"
 FROM
    "DIM_PERSONA"  "DIM_PERSONA"
   JOIN   "H_DATOS_SOCIOECONOMICOS"  "H_DATOS_SOCIOECONOMICOS" ON ( ( "DIM_PERSONA"."ID_PERSONA" =
"H_DATOS_SOCIOECONOMICOS"."ID_PERSONA" ) )
   a_table_to_analyze a_table_to_analyze_type;
FUNCTION "GET_CONSTANT_0_N_TABLA" RETURN VARCHAR2 IS
                BEGIN         RETURN "CONSTANT_0_N_TABLA";
END "GET_CONSTANT_0_N_TABLA";
FUNCTION "GET_CONSTANT_0_N_TABLA"(p_value VARCHAR2, p_write NUMBER) RETURN VARCHAR2 IS
                BEGIN         IF (p_write IS NOT NULL AND p_write = 1) THEN    "CONSTANT_0_N_TABLA" := p_value;
                END IF;
RETURN "CONSTANT_0_N_TABLA";
END "GET_CONSTANT_0_N_TABLA";
-- Access function for ATTRIBUTE CONSTANT_0.OUTGRP1.FECHA
FUNCTION "GET_CONSTANT_0_0_FECHA" RETURN DATE IS
BEGIN
  RETURN "CONSTANT_0_0_FECHA";
END "GET_CONSTANT_0_0_FECHA";
BEGIN
  IF get_abort THEN
    RETURN FALSE;
  END IF;
  get_abort_procedure := FALSE;
  IF NOT (get_audit_level = AUDIT_NONE) THEN
     batch_auditd_id := OWBSYS.WB_RT_MAPAUDIT.auditd_begin(  -- Template BatchAuditDetailBegin
      p_name=>'"H_DATOS_SOCIOECONOMICOS_P_Bat"',
      p_source=>'*',       p_source_uoid=>'*',
      p_target=>'"H_DATOS_SOCIOECONOMICOS_PC1"'     );
    get_audit_detail_type_id := OWBSYS.WB_RT_MAPAUDIT.register_audit_detail_type(
    p_rtd=>get_audit_detail_id,
    p_parent_operator_uoid=>'15B4F45A207745D5B1E3108DBCB2ADB2', -- Operator DIM_PERSONA
    p_parent_object_name=>'DIM_PERSONA',
    p_parent_object_uoid=>'310422C809F84F47ACB730E05AC52491', -- Tabla DIM_PERSONA
    p_parent_object_type=>'Tabla',    p_object_name=>'DIM_PERSONA',
    p_object_uoid=>'310422C809F84F47ACB730E05AC52491', -- Tabla DIM_PERSONA
    p_object_type=>'Tabla',    p_location_uoid=>'233A78AB9E2E4D52A17A282DBD1F795D' -- Location STAGE_LOCATION1
  );
    get_audit_detail_type_id := OWBSYS.WB_RT_MAPAUDIT.register_audit_detail_type(
    p_rtd=>get_audit_detail_id,
    p_parent_operator_uoid=>'237E2C8915704F63A5DC3C133A12F1E4', -- Operator H_DATOS_SOCIOECONOMICOS_PC
    p_parent_object_name=>'H_DATOS_SOCIOECONOMICOS_PC',
    p_parent_object_uoid=>'25DA89645CA2448797C869020237C692', -- Tabla H_DATOS_SOCIOECONOMICOS_PC
    p_parent_object_type=>'Tabla',    p_object_name=>'H_DATOS_SOCIOECONOMICOS_PC',
    p_object_uoid=>'25DA89645CA2448797C869020237C692', -- Tabla H_DATOS_SOCIOECONOMICOS_PC
    p_object_type=>'Tabla',    p_location_uoid=>'233A78AB9E2E4D52A17A282DBD1F795D' -- STAGE_LOCATION1
  );
  -- Bulk: types for collection variables
TYPE "T_JOINER_87_ID_PERSONA" IS TABLE OF VARCHAR2(15) INDEX BY BINARY_INTEGER;
TYPE "T_ROWKEY_JOINER" IS TABLE OF VARCHAR2(18) INDEX BY BINARY_INTEGER;
-- Bulk: intermediate scalar variables
"SV_JOINER_87_ID_PERSONA"  VARCHAR2(15);
"SV_ROWKEY_JOINER"  VARCHAR2(18);
-- Bulk: intermediate collection variables
"JOINER_87_ID_PERSONA" "T_JOINER_87_ID_PERSONA";
"ROWKEY_JOINER" "T_ROWKEY_JOINER";
"H_DATOS__0_ID_PERSO" "T_H_DATOS__0_ID_PERSO";
```

Figura 32: Código de Transformación
Fuente: Script extraído de la herramienta OWB

La Figura 33 muestra los procedimientos que se realiza para la transformación en el espacio de trabajo Stage, los datos que tiene que insertar a la tabla hecho de datos socioeconómicos.

```
--------------------------------------------------------------------------------
-- procedimiento que recorre las filas de una tabla obteniendo datos desde el espacio de trabajo del Source
--------------------------------------------------------------------------------
PROCEDURE "JOINER_ES"(error_index IN NUMBER) IS
BEGIN
  OWBSYS.WB_RT_MAPAUDIT.error_source (
    p_rta=>get_runtime_audit_id,    p_rowkey=>get_rowkey + error_index - 1,    p_seq=>1,    p_instance=>1,
p_table=>SUBSTR('"DIM_PERSONA","DIM_SEXO","DIM_ESTADO_CIVIL","DIM_PAIS_NACIONALIDAD","DIM_PAIS_ORIGEN","DIM_PROVINCIA",
"DIM_DEPARTAMENTO_NACIMIENTO","DIM_GRADO_ACADEMICO","DIM_OCUPACION_PADRE","DIM_OCUPACION_MADRE","DIM_VIVIENDA",
"DIM_VIVIENDA_CARACTERISTICA","DIM_VIVIENDA_ZONA","DIM_COLEGIO","DIM_COLEGIO_ADMINISTRACION","DIM_COLEGIO_TURNO",
"DIM_COLEGIO_AREA","DIM_COLEGIO_DEPARTAMENTO","DIM_COLEGIO_BACHILLERATO","DIM_TRABAJO_TIEMPO","DIM_TRABAJO_TIPO",
"DIM_TRABAJO_OCUPACION","H_DATOS_SOCIOECONOMICOS"',0,80),
  p_column=>SUBSTR ('ID_PERSONA',0,80),  p_value=>SUBSTRB("JOINER_87_ID_PERSONA"(error_index),0,2000),p_step=>get_step_number, p_role=>'S'
);    RETURN;
  END "JOINER_ES";
--------------------------------------------------------------------------------
-- Procedimiento del "JOINER_DML" para realizar las transformaciones en el starting, analizando casa registro de dato.
--------------------------------------------------------------------------------
PROCEDURE "JOINER_DML"(si NUMBER, firstround BOOLEAN) IS
 "H_DATOS_SOCIOECONOMICOS_P_ins0" NUMBER := "H_DATOS_SOCIOECONOMICOS_P_ins";
BEGIN
  IF NOT "H_DATOS_SOCIOECONOMICOS_PC1_St" AND (NOT get_use_hc OR get_row_status)
AND (NOT (get_use_hc AND "H_DATOS_SOCIOECONOMICOS_P_nul")) THEN
  -- Insert DML for "H_DATOS_SOCIOECONOMICOS_PC1"
      normal_action := 'INSERT';  error_action  := 'INSERT';
    get_target_name := '"H_DATOS_SOCIOECONOMICOS_PC1"';  get_audit_detail_id := "H_DATOS_SOCIOECONOMICOS_PC1_id";
  IF get_use_hc AND NOT firstround THEN     "H_DATOS_SOCIOECONOMICOS_PC1_si" := "H_DATOS_SOCIOECONOMICOS_PC1_ir"(si);
    IF "H_DATOS_SOCIOECONOMICOS_PC1_si" = 0 THEN       "H_DATOS_SOCIOECONOMICOS_PC1_i" := 0;
    ELSE
      "H_DATOS_SOCIOECONOMICOS_PC1_i" := "H_DATOS_SOCIOECONOMICOS_PC1_si" + 1;    END IF;
  ELSE
    "H_DATOS_SOCIOECONOMICOS_PC1_si" := 1;  END IF;
  LOOP
    BEGIN
      FORALL i IN "H_DATOS_SOCIOECONOMICOS_PC1_si".."H_DATOS_SOCIOECONOMICOS_PC1_i" - 1
        INSERT        /*+ APPEND PARALLEL("H_DATOS_SOCIOECONOMICOS_PC1") */
        INTO
          "H_DATOS_SOCIOECONOMICOS_PC"         "H_DATOS_SOCIOECONOMICOS_PC"."ID_PERSONA",
        VALUES
          "H_DATOS__0_ID_PERSO"(i),
          RETURNING ROWID BULK COLLECT INTO get_rowid;
      error_index := "H_DATOS_SOCIOECONOMICOS_PC1_si" + get_rowid.COUNT;
    EXCEPTION
Main_ES(get_step_number,error_rowkey.get_target_name,SUBSTRB('"H_DATOS_SOCIOECONOMICOS_PC"."ID_PERSONA"',0,80),SUBSTRB
("H_DATOS__0_ID_PERSO"(error_index),0,2000));
get_target_name := '"H_DATOS_SOCIOECONOMICOS_PC1"'; get_audit_detail_id := "H_DATOS_SOCIOECONOMICOS_PC1_id";
          IF NOT "H_DATOS_SOCIOECONOMICOS_PC1_St" AND NOT get_buffer_done(get_buffer_done_index) THEN
            BEGIN
              get_rows_processed := true; -- Set to indicate that some row data was processed (for debugger)
            error_stmt := SUBSTRB('"H_DATOS__0_ID_PERSO"("H_DATOS_SOCIOECONOMICOS_PC1_i") :=
"JOINER_87_ID_PERSONA"("JOINER_i");',0,2000);
          error_column := SUBSTRB('"H_DATOS__0_ID_PERSO"',0,80);
          BEGIN
            error_value := SUBSTRB("JOINER_87_ID_PERSONA"("JOINER_i"),0,2000);
          EXCEPTION
            WHEN OTHERS THEN
              error_value := '*';
          END;
          NOT get_use_hc THEN         "H_DATOS__0_ID_PERSO"("H_DATOS_SOCIOECONOMICOS_PC1_i") :="JOINER_87_ID_PERSONA"("JOINER_i");
          ELSIF get_row_status THEN
            "SV_H_DATOS__0_ID_PERSO" :="JOINER_87_ID_PERSONA"("JOINER_i");
          ELSE       NULL;
          END IF;
          IF get_use_hc AND get_row_status AND ("H_DATOS_SOCIOECONOMICOS_P_new") THEN             dml_bsize := dml_bsize + 1;
                IF "H_DATOS_SOCIOECONOMICOS_P_new"     AND (NOT "H_DATOS_SOCIOECONOMICOS_P_nul") THEN
                  "H_DATOS_SOCIOECONOMICOS_PC1_ir"(dml_bsize) := "H_DATOS_SOCIOECONOMICOS_PC1_i";
                "H_DATOS__0_ID_PERSO"("H_DATOS_SOCIOECONOMICOS_PC1_i") := "SV_H_DATOS__0_ID_PERSO";
                "H_DATOS_SOCIOECONOMICOS_P_srk"("H_DATOS_SOCIOECONOMICOS_PC1_i") := "SV_H_DATOS_SOCIOECONOMICOS_srk";
            "H_DATOS_SOCIOECONOMICOS_PC1_i" := "H_DATOS_SOCIOECONOMICOS_PC1_i" + 1;
          ELSE
            "H_DATOS_SOCIOECONOMICOS_PC1_ir"(dml_bsize) := 0;
          END IF;          END IF;
error_stmt := SUBSTRB('WB_DISABLE_ALL_CONSTRAINTS(("HC_DSOCIOECO_DSOCIOECO_PC"."CONSTANT_0_N_TABLA"/* ATTRIBUTE
CONSTANT.OUTGRP1.N_TABLA */))/* OPERATOR WB_DISABLE_ALL_CONSTRAINTS */;
END "HC_DSOCIOECO_DSOCIOECO_PC";
```

Figura 33: Código de procedimientos de transformación
Fuente: Script extraído de la herramienta OWB

- **Carga de Datos al Data Warehouse**

  Terminando la fase de transformación de datos desde el espacio de trabajo Source al Stage, se realizan mapeos para pasar del espacio de trabajo Stage al Star, este último es el modelado del esquema estrella el cual se cargara a la Base de Datos del Data Warehouse.

  En la Figura 34 muestra la estructura que tiene la tabla hecho Datos Socioeconomicos en el espacio de trabajo Star, siguiendo el concepto del esquema en estrella, en esta fase se cumplió con el total de registros (véase punto 4.3.3) a cargarse en el Data Warehouse.

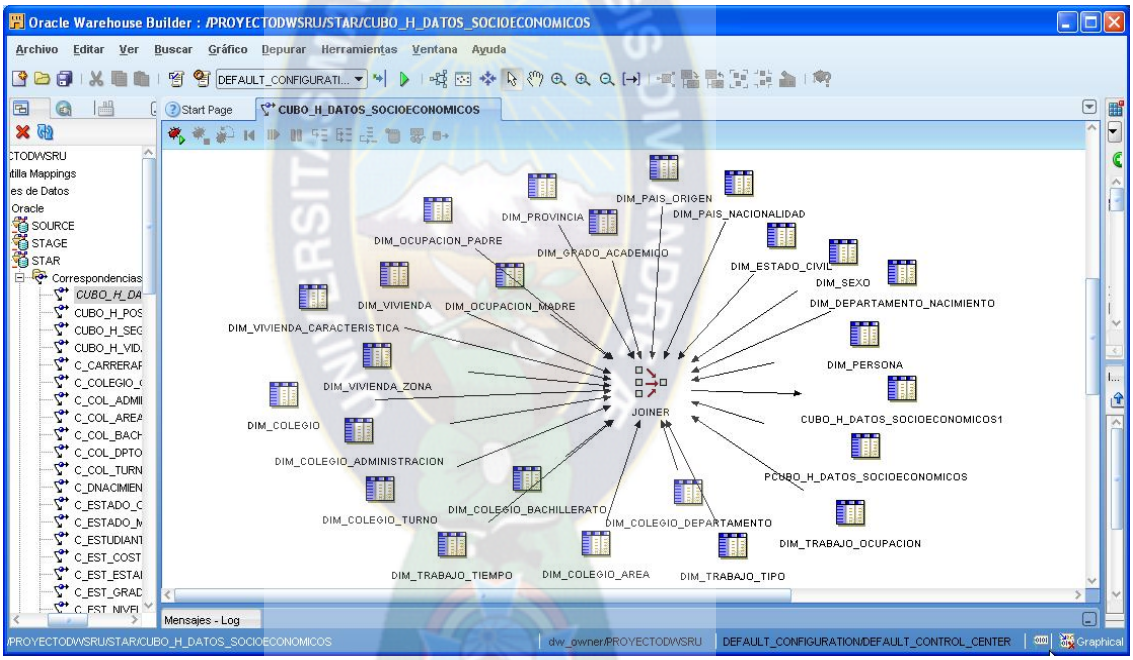


Figura 34: Mapeo para Carga de datos al hecho Datos Socioeconomicos
Fuente: Herramienta OWB

4.4.2 Implementación de Capa de Agregación de Datos

Para realizar el proceso de explotación de datos se utilizó las herramientas Oracle BI Administration Tool, Enterprise Manager Fusion Middleware. Esta fase de la implementación se divide en tres capas: Física, Modelo de Negocio y Presentación. En la capa física se realiza la conexión a la base de datos del Data Warehouse. En cuanto a la Capa de Modelo de Negocio nos permite realizar las agregaciones a las tablas hecho y

dimensiones como se muestra en la Figura 35, es aquí donde se mapea las agregaciones de cada dato según requerimientos de información, esta etapa de explotación de datos concluye en la capa de presentación donde se halla todos los datos ya configurados para su respectivo uso o administración por el usuario.

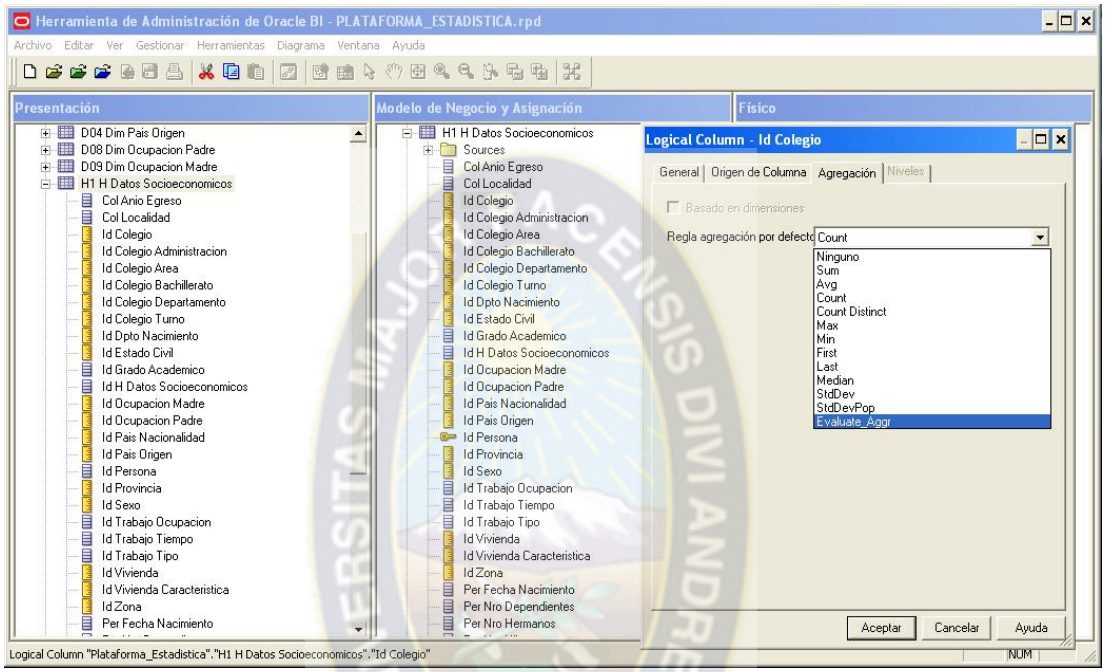


Figura 35: Agregación de medidas en las tablas hecho
Fuente: Herramienta BI Oracle

4.4.3 Implementación Capa de Presentación de Datos

En la presenten capa se especificara los permisos y accesos de Tipos de Usuarios, como también los resultados obtenidos de los procesos descritos anteriormente, estos resultados se presentan al usuario final empleando la herramienta de Analytics de Oracle.

Tipo de Usuario	Permisos de Acceso a la Base de Datos	
	Data Warehouse	Base de Datos OLAP
Usuario Operador y Mantenimiento	Acceso Total	Acceso Total
Administrador que trabaja con la información	Lectura	Lectura

Tabla 13: Permisos a Usuarios
Fuente: Elaboración Propia

La Figura 36 muestra la pantalla principal del Data Warehouse de la Plataforma Estadística.

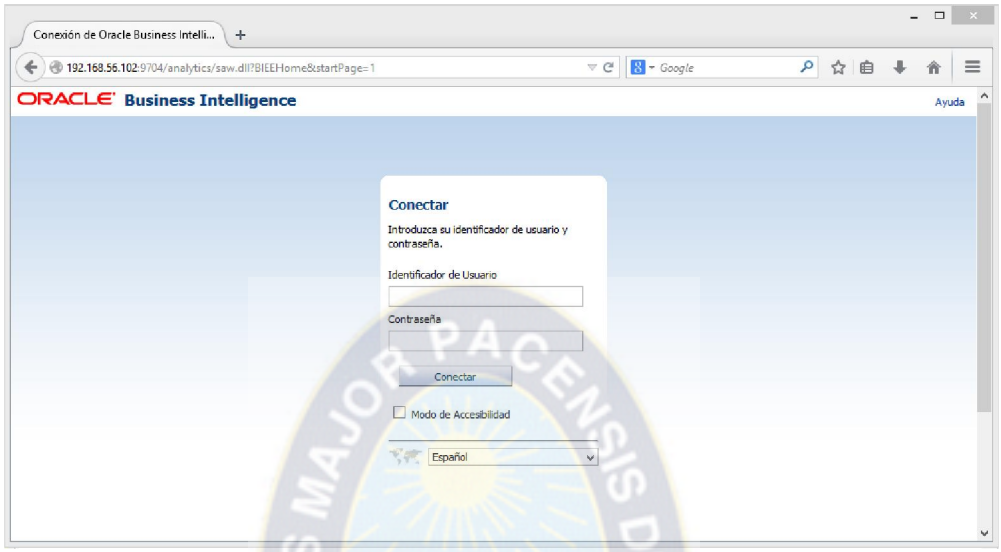
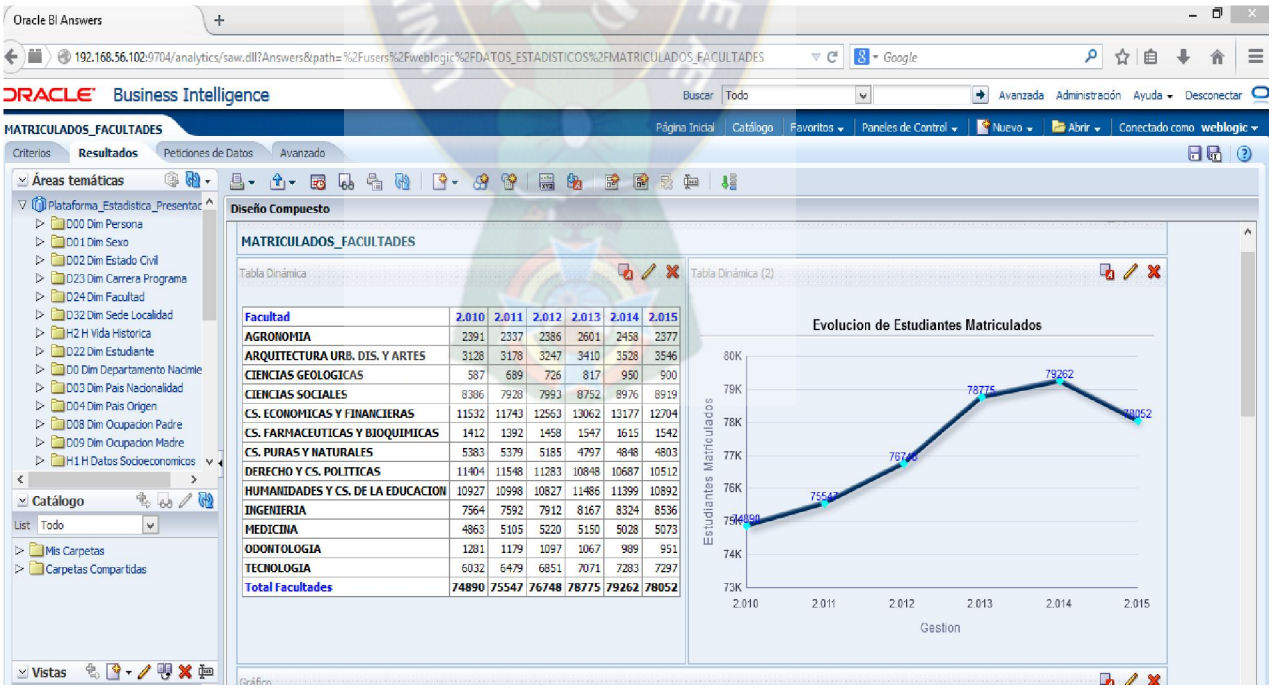


Figura 36: Página principal del Data Warehouse
Fuente: Pantalla principal de la herramienta Analytics

La Figura 37, los reportes que permiten la personalización de visualización mediante filtros, con relación al crecimiento poblacional estudiantil por gestión a nivel U.M.S.A.



Facultad	2.010	2.011	2.012	2.013	2.014	2.015
AGRONOMIA	2391	2337	2386	2601	2458	2377
ARQUITECTURA URB. DIS. Y ARTES	3128	3178	3247	3410	3528	3546
CIENCIAS GEOLOGICAS	587	689	726	817	950	900
CIENCIAS SOCIALES	8386	7928	7993	8752	8976	8919
CS. ECONOMICAS Y FINANCIERAS	11532	11743	12563	13062	13177	12704
CS. FARMACEUTICAS Y BIOQUIMICAS	1412	1392	1458	1547	1615	1542
CS. PURAS Y NATURALES	5383	5379	5185	4797	4848	4803
DERECHO Y CS. POLITICAS	11404	11548	11283	10848	10687	10512
HUMANIDADES Y CS. DE LA EDUCACION	10927	10998	10827	11486	11399	10892
INGENIERIA	7564	7592	7912	8167	8324	8536
MEDICINA	4863	5105	5220	5150	5028	5073
ODONTOLOGIA	1281	1179	1097	1067	989	951
TECNOLOGIA	6032	6479	6851	7071	7283	7297
Total Facultades	74890	75547	76748	78775	79262	78052

Figura 37: Indicador del crecimiento poblacional estudiantil
Fuente: Reportes de estudiantes matriculados

La Figura 38 nos muestra los reportes de la cantidad de estudiantes matriculados a nivel carrera de cada facultad de la U.M.S.A.

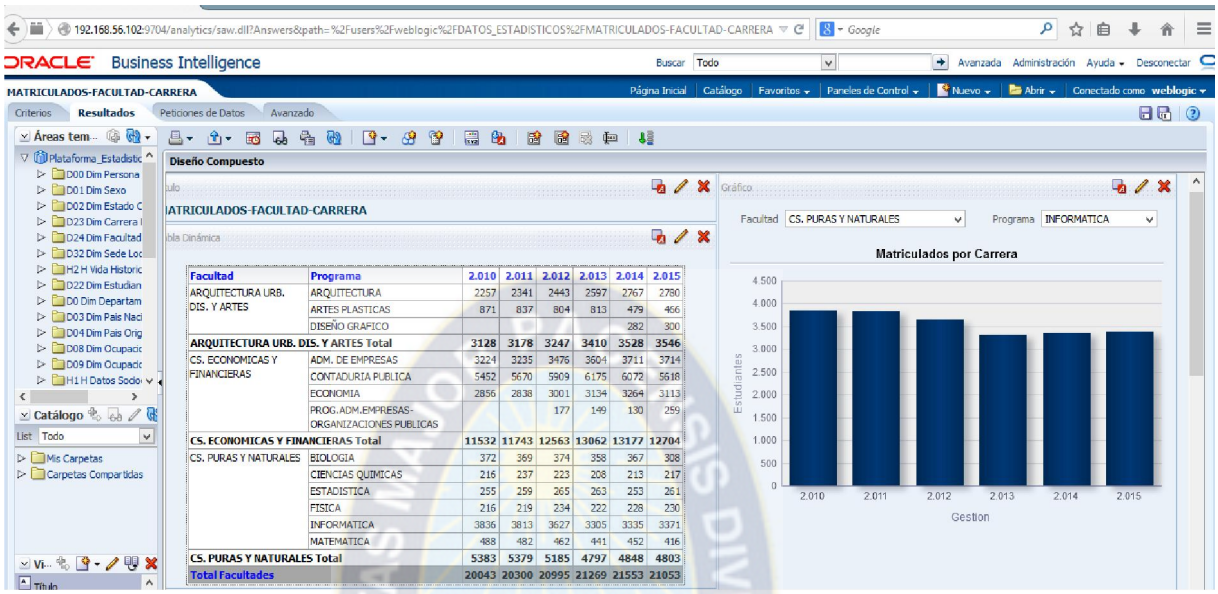


Figura 38: Indicadores de estudiantes matriculados por carrera
Fuente: Reportes de estudiantes matriculados

La Figura 39 detalla la cantidad de estudiantes matriculados por género a nivel carrera.

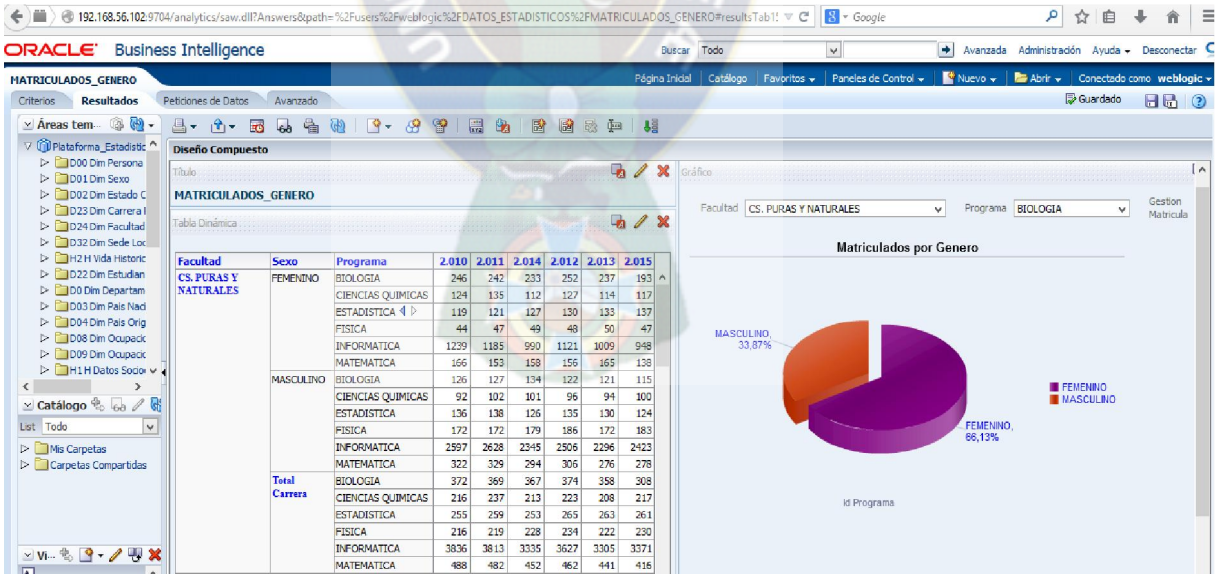
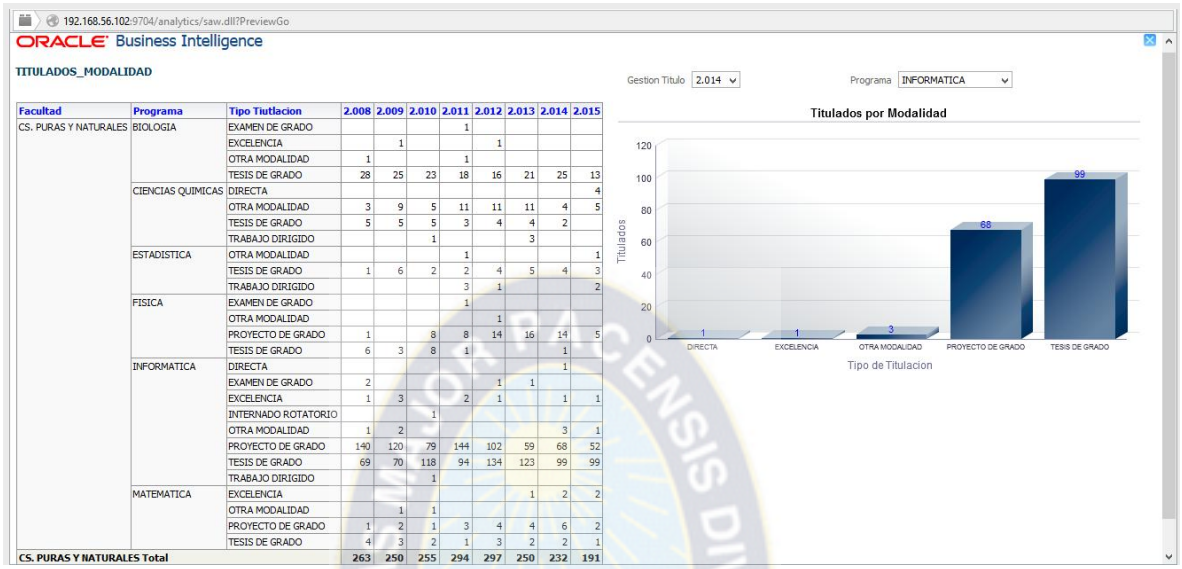


Figura 39: Cantidad de estudiantes matriculados por Género
Fuente: Reportes de estudiantes matriculados

La Figura 40 nos muestra el reporte de estudiantes titulados, el cual también permite la personalización de visualización mediante filtros.



Facultad	Programa	Tipo Tiutlacion	2.008	2.009	2.010	2.011	2.012	2.013	2.014	2.015
CS. PURAS Y NATURALES	BIOLOGIA	EXAMEN DE GRADO				1				
		EXCELENCIA		1			1			
		OTRA MODALIDAD	1			1				
		TESIS DE GRADO	28	25	23	18	16	21	25	13
	CIENCIAS QUIMICAS	DIRECTA								4
		OTRA MODALIDAD	3	9	5	11	11	11	4	5
		TESIS DE GRADO	5	5	5	3	4	4	2	
		TRABAJO DIRIGIDO			1			3		
	ESTADISTICA	OTRA MODALIDAD				1				1
		TESIS DE GRADO	1	6	2	2	4	5	4	3
		TRABAJO DIRIGIDO				3	1			2
	FISICA	EXAMEN DE GRADO				1				
		OTRA MODALIDAD					1			
		PROYECTO DE GRADO	1		8	8	14	16	14	5
		TESIS DE GRADO	6	3	8	1			1	
	INFORMATICA	DIRECTA							1	
		EXAMEN DE GRADO	2				1	1		
		EXCELENCIA	1	3		2	1		1	1
		INTERNADO ROTATORIO			1					
		OTRA MODALIDAD	1	2					3	1
		PROYECTO DE GRADO	140	120	79	144	102	59	68	52
		TESIS DE GRADO	69	70	118	94	134	123	99	99
		TRABAJO DIRIGIDO			1					
	MATEMATICA	EXCELENCIA						1	2	2
		OTRA MODALIDAD		1	1					
		PROYECTO DE GRADO	1	2	1	3	4	4	6	2
		TESIS DE GRADO	4	3	2	1	3	2	2	1
CS. PURAS Y NATURALES Total			263	250	255	294	297	250	232	191

Figura 40: Reporte de titulados por modalidad
Fuente: Reportes de estudiantes titulados

En la Figura 41 mostramos un reporte de análisis con relación a estudiantes y lugar o zona donde vive cada estudiante.

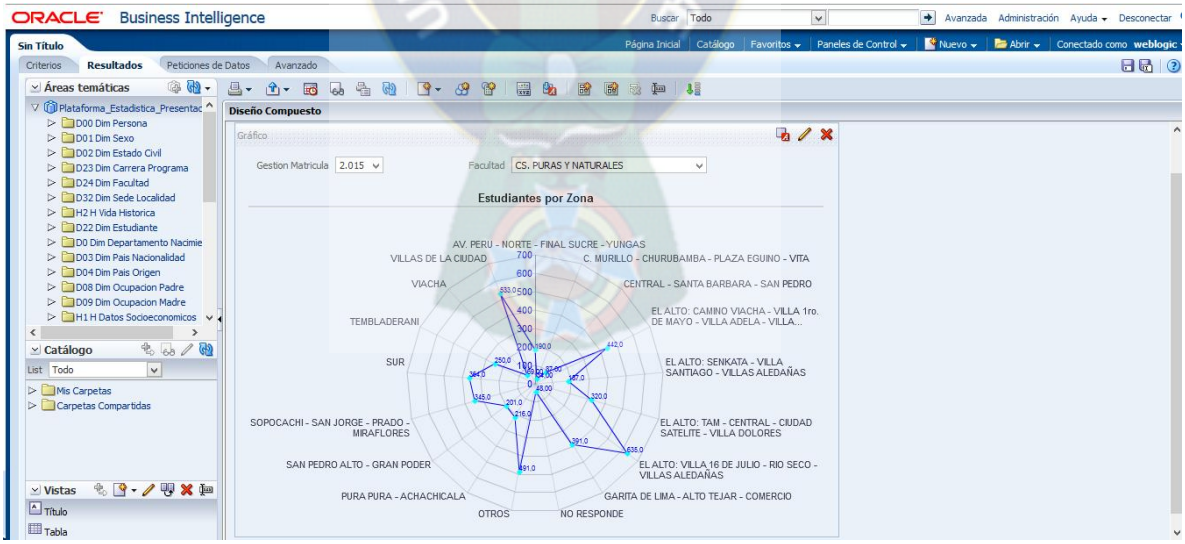


Figura 41: Reporte de titulados por modalidad
Fuente: Reportes de estudiantes matriculados

4.5 Revisión

Uno de los factores más importante a evaluar en un Data Warehouse es la calidad de datos que contiene, ya que esta se ha convertido en la materia prima para cumplir los requerimientos según necesidades de los usuarios tomadores de decisión, por ello es importante que los datos del almacén reflejen correctamente el mundo real, pero es también muy importante que los datos sean interpretados correctamente.

4.5.1 Análisis Calidad de Datos

Por lo mencionado anteriormente se aplicara métricas ISO/IEC 25012 para realizar el análisis de calidad de datos del Data Warehouse.

Las métricas que evaluaremos con relación a la calidad de datos del Data Warehouse son:

Completitud: Es el grado de datos que contiene cada tabla hecho, esto asociado con las tablas dimensiones que contiene los atributos esperados en un contexto especifico.

NT= Número Total de tablas del Data Warehouse

NT= 43 Tablas construidas en el Data Warehouse

Nombre tabla	Calculo de la Métrica NA(t) Numero de Atributos
H_Datos_Socioeconomicos	35
H_Vida_Historica	20
H_Seguro_Universitario	17
H_Postulantes_Titulo	13

Tabla 14: Calculo de métricas NA (t)
Fuente: Elaboración Propia

Consistencia: Representa a los registros o tuplas que contiene las tablas hechos, los cuales son coherentes con otros datos en un contexto especifico.

Nombre tabla	Calculo de la Métrica NR(t) Numero de registros(tuplas)
H_Datos_Socioeconomicos	232.068
H_Vida_Historica	1.617.245
H_Seguro_Universitario	55.959
H_Postulantes_Titulo	33110

Tabla 15: Calculo de métricas NR (t)
Fuente: Elaboración Propia

Exactitud: Las tablas hecho contienen atributos que correctamente representan el valor asignado en un contexto específico.

Nombre tabla	Calculo de la Métrica NVN(t) Numero de Valores Nulos o faltantes
H_Datos_Socioeconomicos	0
H_Vida_Historica	2836
H_Seguro_Universitario	237
H_Postulantes_Titulo	348

Tabla 16: Calculo de métricas NVN (t)
Fuente: Elaboración Propia

Credibilidad: Las atributos de las tablas hecho fueron analizadas en el proceso de extracción de datos de las fuentes origen, según contexto especifico.

Nombre tabla	Calculo de la Métrica NVE(t) Numero de Valores Erróneos
H_Datos_Socioeconomicos	0
H_Vida_Historica	0
H_Seguro_Universitario	0
H_Postulantes_Titulo	0

Tabla 17: Calculo de métricas NVE (t)
Fuente: Elaboración Propia

Para verificar la calidad de datos almacenados en el Data Warehouse se analiza la siguiente formula:

$$\textbf{Formula (1):}\ GCD\ (t) = 1 - \frac{NVE\ (t) + NVN\ (t)}{NR\ (t) * NA\ (t)}$$

GCD (t) es el grado de corrección de los datos de una determinada tabla (t), expresa la relación entre la cantidad de datos con valores erróneos, fuera de rangos o nulos/faltantes y la cantidad de datos totales de la tabla.

- Si GCD (T) es ALTO implica que los datos de la tabla son útiles para el proyecto de explotación de información.
- Si GCD (T) es BAJO implica que deben analizarse los datos y evaluar la relevancia de la tabla.

La unidad de medida de la métrica es % (porcentaje) y toma valores entre 0 y 1.

Aplicando la Formula (1) la Tabla H_Datos_Socioeconomicos:

$$GCD\ (t) = 1 - \frac{(0+0)}{232.068 * 35} = 1\text{-}0=1$$

Aplicando la Formula (1) la Tabla H_Vida_Historica:

$$GCD\ (t) = 1 - \frac{(0+2836)}{1.617.245 * 20} = 1 - 0.0000876 = 1$$

Aplicando la Formula (1) la Tabla H_Seguro_Universitario:

$$GCD\ (t) = 1 - \frac{(0+237)}{55.959 * 17} = 1\text{-}0.00025=1$$

Aplicando la Formula (1) la Tabla H_Postulantes_Titulo:

$$GCD\ (t) = 1 - \frac{(0+348)}{33.110 * 13} = 1\text{-}0.00080=1$$

Se aplicó la Formula 1, para realizar el análisis de la calidad de datos a las tablas Hechos del data warehouse, es así que para realizar la evaluación se tomó en cuenta las características más importantes de la norma como son: completitud, consistencia, exactitud y credibilidad de los datos de las tablas Hechos.

Como se muestra en las Tablas 14, 15, 16 y 17se evaluó cada característica obteniendo un resultado por cada Tabla Hecho, este resultado se aplicó a la Formula 1, de igual manera para cada tabla Hecho, el resultado de este proceso para cada tabla nos dio un valor 1 lo cual significa según la norma que los datos de las tablas Hecho son de calidad al 100% y se puede afirmar que se cuenta con datos confiables, consistentes y completo.

4.5.2 Procedimientos de Pruebas

La Tabla 13 detalla las funciones a probar con relación a las cantidad de registros que obtendremos de las fuentes de datos los cuales tienen que ser migrados en su totalidad al Data Warehouse.

Caso de Prueba					
Entradas Requeridas		Acceso a la base de datos de los orígenes y del Data Warehouse.			
Objeto/Descripción de la Prueba		Comprobar que las fuentes de Origen de Datos estén realizando correctamente la extracción de Datos al Data Warehouse.			
Prueba	**Función a Probar**	**Acción o Instrucción**	**Resultado Esperado**	**Resultado Obtenido**	**Éxito/ Fallo**
1	Cantidades de registros en el Sistema SRU.	Realizar consultas de conteo para comparar la información.	Mismo número de registros.	Número de registros igual a número esperado.	Éxito
2	Cantidad de registros en la Aplicación OK	Realizar consultas de conteo para comparar la información.	Mismo número de registros.	Número de registros igual a número esperado.	Éxito
3	Realización de los procesos de transferencia de datos	Correr los paquetes implementados	Datos migrados en su totalidad	Cantidad de registros migrados igual a la cantidad de registros esperados.	Éxito

Tabla 18: Caso de Prueba de ETL
Fuente: Elaboración Propia

La Tabla 14 describe las funciones a probar con relación a la explotación de los datos desde el Data Warehouse, verificando las cantidades exactas que debe mostrar en comparación a las fuentes de datos.

Caso de Prueba					
Entradas Requeridas	Acceso a la base de datos del Data Warehouse.				
Objeto/Descripción de la Prueba	Comprobar que los Cubos OLAP están realizando correctamente las agregaciones.				
Prueba	**Función a Probar**	**Acción o Instrucción**	**Resultado Esperado**	**Resultado Obtenido**	**Éxito /Fall**
1	Cantidad total de estudiantes matriculados relacionados con datos Socioeconomicos.	Operaciones de conteo en la base de datos del DW, agregando las variables socioeconómicas.	Mismo cantidad de registros, dependiendo del tipo de variable socioeconómica que se consulte.	Misma cantidad de registros dependiendo de la variable socioeconómica consultada.	Éxito
2	Cantidad total de estudiantes matriculados con relación a datos académicos.	Operaciones de conteo en la base de datos del DW, agregando las variables académicas.	Misa cantidad de registros dependiendo del tipo de variable académica que se consulte.	Misma cantidad de registros dependiendo de la variable académica consultada.	Éxito
3	Cantidad total de estudiantes con relación a cada variable relaciona con seguro universitario.	Operaciones de conteo en el DW, agregando las variables de seguro universitario.	Misa cantidad de registros dependiendo del tipo de variable de seguro universitario que se consulte.	Misma cantidad de registros dependiendo de la variable de seguro universitario consultada.	Éxito
4	Cantidad total de estudiantes con relación a cada variable relaciona con postulantes a título.	Operaciones de conteo en DW, agregando las variables postulantes a título.	Misa cantidad de registros dependiendo del tipo de variable de postulantes a título que se consulte.	Misma cantidad de registros dependiendo de la variable de postulantes a titulo consultada.	Éxito

Tabla 19: Caso de Prueba cubo OLAP
Fuente: Elaboración Propia

4.5.3 Cumplimiento de Requerimientos

La Tabla 15 detalla las funciones a probar con relación al cumplimiento de requerimientos de información.

Descripción 1 (D1): Perspectiva del cubo que muestra la información.

Descripción 2 (D2): Implementado.

Descripción 3 (D3): Requerimiento Verificado por los Usuarios.

Requerimiento	D1	D2	D3
Realizar un análisis estadístico sobre datos sociales y económicos de la población estudiantil. La información estadística debe ser analizada a nivel carrera, facultad y universidad.	SRU /Aplicación OK	SI	SI
Obtener información estadística socioeconómica más descriptiva con análisis más complejos, los datos deben ser analizados a nivel carrera y facultad.	SRU/ Aplicación OK	SI	SI
Conocer información para ejecutar proyectos de Guardería Infantil, etc.	SRU / OK	SI	SI
Conocer información sobre estudiantes que se benefician con el seguro universitario PROMES, analizar datos a nivel carrera, facultad y sedes.	SRU	SI	SI
Conocer información sobre el crecimiento poblacional, comparaciones poblacionales e implementar porcentajes de medición.	SRU	SI	SI
Gestionar el crecimiento poblacional estudiantil según emisión de matrícula, analizar a nivel carrera facultad y U.M.S.A.	SRU	SI	SI
Generar información sobre los años de estudio que lleva un universitario en una carrera, facultad y	SRU	SI	SI
Conocer información sobre los antecedentes universitarios, es muy significativo para realizar una mejor gestión académica, en consecuencia a este análisis poder ejecutar proyectos y políticas en beneficios a los estudiantes de la U.M.S.A.	SRU	SI	SI

Tabla 20: Cumplimiento de requerimientos de información
Fuente: Elaboración Propia

Con los datos expuestos tanto en la calidad de datos, Pruebas de procedimientos de ETL y explotación, como el cumplimiento con éxito a los requerimientos solicitados podemos determinar que la Plataforma de Información Estadística cumple con los estándares de calidad para su puesta en producción.

CAPITULO V

CONCLUSIONES Y RECOMENDACIONES

5.1 Conclusiones

Terminando el proceso de construcción de la Plataforma de Información Estadística, se concluye que:

- ✓ Se ha implementado una plataforma de información estadística orientada al soporte de decisiones centralizando, procesando y administrando datos históricos respecto a datos sociales, económicos y académicos de los estudiantes de la U.M.S.A.
- ✓ Se estableció los niveles de relación de dos fuentes de información, identificando en cada una de ellas los requerimientos de información y así unificando información similar.
- ✓ El proyecto se desarrolló utilizando una metodología propia de Data Warehouse denominada Rapid Warehousing Methodology.
- ✓ Se diseñó un Data Warehouse que permite el almacenamiento de la información histórica de las fuentes proveedoras de datos.
- ✓ Se implementó procesos de limpieza y cargado de datos desde las fuentes orígenes de datos a la base de datos del Data Warehouse.
- ✓ Se desarrolló soluciones OLAP que permite la agregación de información según requerimientos de los usuarios finales.
- ✓ Para la interacción de la plataforma estadística con el usuario se desarrolló vistas que permiten diseñar reportes personalizados a través de filtros representando cuadros e indicadores estadísticos.

- ✓ Se cumplió con cada uno de los objetivos propuestos en el Capítulo I, es así que los puntos anteriormente mencionados demuestran la realización a detalle de los objetivos del presente proyecto.

5.2 Recomendaciones

- ❖ La plataforma de información estadística a toma en cuenta datos del Sistema de Registro Universitario y la Aplicación OK, sin embargo al realizar todo el análisis de la problemática, se pudo observar que la plataforma puede ser bien completada relacionándola con el Sistema Académico Universitario para enriquecer la información estadística y tener conocimientos más representativos con relación a los estudiantes de la U.M.S.A.

- ❖ Incorporar datos referentes a variables sociales, previo análisis por expertos en el área de estadística, para contar con información más completa con relación a la situación socioeconómica de la población universitaria.

- ❖ Incluir un módulo de Data Mining para la implementación de modelos que permitan un análisis estadístico automatizado de la información, dado que la Universidad Mayor de San Andrés ya cuenta con proyectos implementados sobre Data Warehouse en el área de presupuestos y registro universitario.

BIBLIOGRAFÍA

[1] DYCHE, HILL.: **E-Data, Transformación de datos en Información con Data Warehousing;** Ed. 1ed.

[2] GILL, HARJINDERS: **Data Warehousing. La integración de información para la mejor Toma de Decisiones.**

[3] RALPH KIMPALL, MAGIE ROSS: **The Data Warehouse Toolkit, the complete guide to dimensional modeling**; second edition, Canada, Wesley, 419

[4] WILEY. MAURIZIO RAFANELLI.: **Multidimensional Databases – Problems and Solutions (OLAP);** Canada: Wesley, 1500

[5] W. H. INMON: **Building the Data Warehouse;** 3er Edición.

[6] PATON GONZALES, ELIANA BETZY: **Metodología de Modelamiento de Data Warehouse;** Universidad Mayor de San Andrés, 2002.

[7] SEVILLA BERRIOS, ERICKA GRACIELA: **Guía Metodológica para la Definición de Desarrollo de un Data Warehouse;** Universidad Americana, 2003.

[8] PERALTA, VERÓNICA: **Diseño Lógico de data Warehouse a partir de Esquemas Conceptuales Multidimensionales,** Universidad de la Republica Pedeciba Informática, 2001.

[9] PEÑA AYALA, ALEJANDRO: **Inteligencia de Negocios: Una Propuesta para su Desarrollo en las organizaciones,** Instituto Politécnico Nacional de México, Ed. Primera 2006.

[10] BERSON A. AND SMITH S.: **Data Warehousing, Data Mining and OLAP;** Me Graw Hill, 1997.

[11] ELIZABET TEJADA AVILA: **Data Warehousing con procesamiento de datos textuales;** Universidad de Granada, 2010.

[12] PAUL LANE: **Oracle Data Base, Data Warehousing Guide,** 2013.

[13] ROBERT WREMBEL Y CHRISTIAN KONCILIA: **Data Warehouses an OLAP Concepts, Architectures and Solutions**; IGI Global, 2006.

[14] PABLO MARTIN GUTIERREZ, **Data Warehouse: Marco de Calidad**; Universidad Carlos III, 2013.

[15] DIEGO MARTIN BASSO**, Propuesta de métricas para Proyectos de Explotación de Información;** Escuela de Postgrado UTN.BA, 2014.

[16] UNIVERSIDADES ANDALUZIA, **Data Warehouse, Metodología de Validación;** 2010.

REFERENCIAS WEB

[Consultas sobre Rapid Warehousing Methodology SAS Institute]

<https://books.google.com.bo/books?id=dGcCwEEJdjEC&pg=PA171&lpg=PA171&dq=Metodolog%C3%ADa+propuesta+por+SAS+Institute&source=bl&ots=UBOzK6YSao&sig=iA9e-0UPyd7HXPqooexUWtFEc5s&hl=es&sa=X&ei=LhErVdfPDbiKsQTX0oHgBw&ved=0CDYQ6AEwBA#v=onepage&q=Metodolog%C3%ADa%20propuesta%20por%20SAS%20Institute&f=false> [Consulta: Julio, 2015].

[Consultas sobre proceso de ETL]

<https://churriwifi.wordpress.com/2009/12/> [Consulta: Agosto, 2015].

[Oracle Data Warehouse Builder]

<http://www.oracle.com/technetwork/es/articles/datawarehouse/instalar-data-warehouse-builder-rac-425081-esa.html> [Consulta: Agosto, 2015].

[Consulta sobre conceptos de Data Warehouse]

<www.rkimball.com> [Consulta: Agosto, 2015].

[Consulta sobre conceptos de Data Warehouse]

<www.datawarehousing.com> [Consulta: Agosto, 2015].

[Oracle BI y Base de Datos]

<http://www.dba-oracle.com/> [Consulta: Octubre, 2015].

[Guía de Herramientas de Oracle]

<http://www.oracle.com/technology/products/discoverer/index.html> [Consulta: Octubre, 2015].

ANEXOS

Anexo A

Tabla de Variables Analizadas en el Sistema de Registro Universitario y Aplicación Informática OK

Información Socioeconómicos	Datos Socioeconómico
Datos Personales	Nombre, Apellidos y Fecha de
	Carnet de Identidad
	Estado Civil, Lugar de Nacimiento
	Número de hermanos en la UMSA
	Número de Hijos, Numero de
	Seguro Medico
Datos Ocupación Laboral	Tipo de Trabajo del Padre, Madre
Datos Vivienda	Tipo Vivienda
	Característica Vivienda
	Dirección Vivienda
	Teléfono
	Zona Aproximada Vivienda
Datos Colegio Egreso	Nombre Colegio
	Ubicación Colegio
	Tipo Administración Colegio
	Tipo Área Colegio
	Turno Colegio
	Tipo de Bachillerato
	Año de egreso colegio
Datos Ocupación Laboral Estudiante	Tipo de Trabajo
	Tiempo de trabajo

Datos Socioeconómicos de estudiantes en la U.M.S.A.

Información Antecedentes Universitario	Datos Antecedentes Universitarios
Datos Estudiantes de Pregrado y Postgrado	Modalidad de Ingreso a la U.M.S.A.
	Carrera
	Facultad
	Registro de Matricula
	Estudiante Tipo (Antiguo, Nuevo, etc.)
	Estado Estudiante (Activo, Inactivo, etc.)
	Estudiante Grado (Titulado, Egresado, etc.)
	Estudiante Nivel (Pregrado, Postgrado)
	Estudiante Mención
	Estudiante Tipo Titulación
	Estudiante Modalidad de Titulación

Datos de antecedente de estudiantes universitarios en U.M.S.A.

Anexo B

A) Formulario 01 - Registro Estudiantil

U. M. S. A. MATRICULACION R.

UNIVERSIDAD MAYOR DE SAN ANDRES DIVISION DE GESTIONES Y ADMISIONES DIVISION DE SISTEMAS DE INFORMACION	FORMULARIO 01 **REGISTRO ESTUDIANTIL**	NUMERO REGISTRO (No llenar)

* ESCRIBA CON LETRA DE IMPRENTA y/o MARQUE CON UNA x LA CASILLA QUE CORRESPONDA

DATOS PERSONALES

APELLIDO PATERNO

APELLIDO MATERNO

NOMBRE COMPLETO

C. IDENTIDAD

SEXO

MASCULINO	1		
FEMENINO	2		

ESTADO CIVIL

SOLTERO	1		DIVORCIADO	3	
CASADO	2		VIUDO	4	

CARRERA A LA QUE SE INSCRIBE

FECHA DE NACIMIENTO

DIA MES AÑO

LUGAR DE NACIMIENTO (NACIONALES)

LOCALIDAD

PROVINCIA

FACULTAD A LA QUE SE INSCRIBE

CIENCIAS PURAS	1		CS. GEOLOGICAS	7	
CS. SOCIALES	2		DERECHO	8	
HUMANIDADES	3		MEDICINA	9	
AGRONOMIA	4		CS. FARMACEUTICAS	10	
ARQUITECTURA	5		ODONTOLOGIA	11	
CS. ECONOMICAS	6		INGENIERIA	12	
			TECNICA	13	

DEPARTAMENTO DE NACIMIENTO (NACIONALES)

CHUQUISACA	1		LA PAZ	2		ORURO	3	
COCHABAMBA	4		SANTA CRUZ	5		POTOSI	6	
TARIJA	7		BENI	8		PANDO	9	

LUGAR DE NACIMIENTO (EXTRANJEROS)

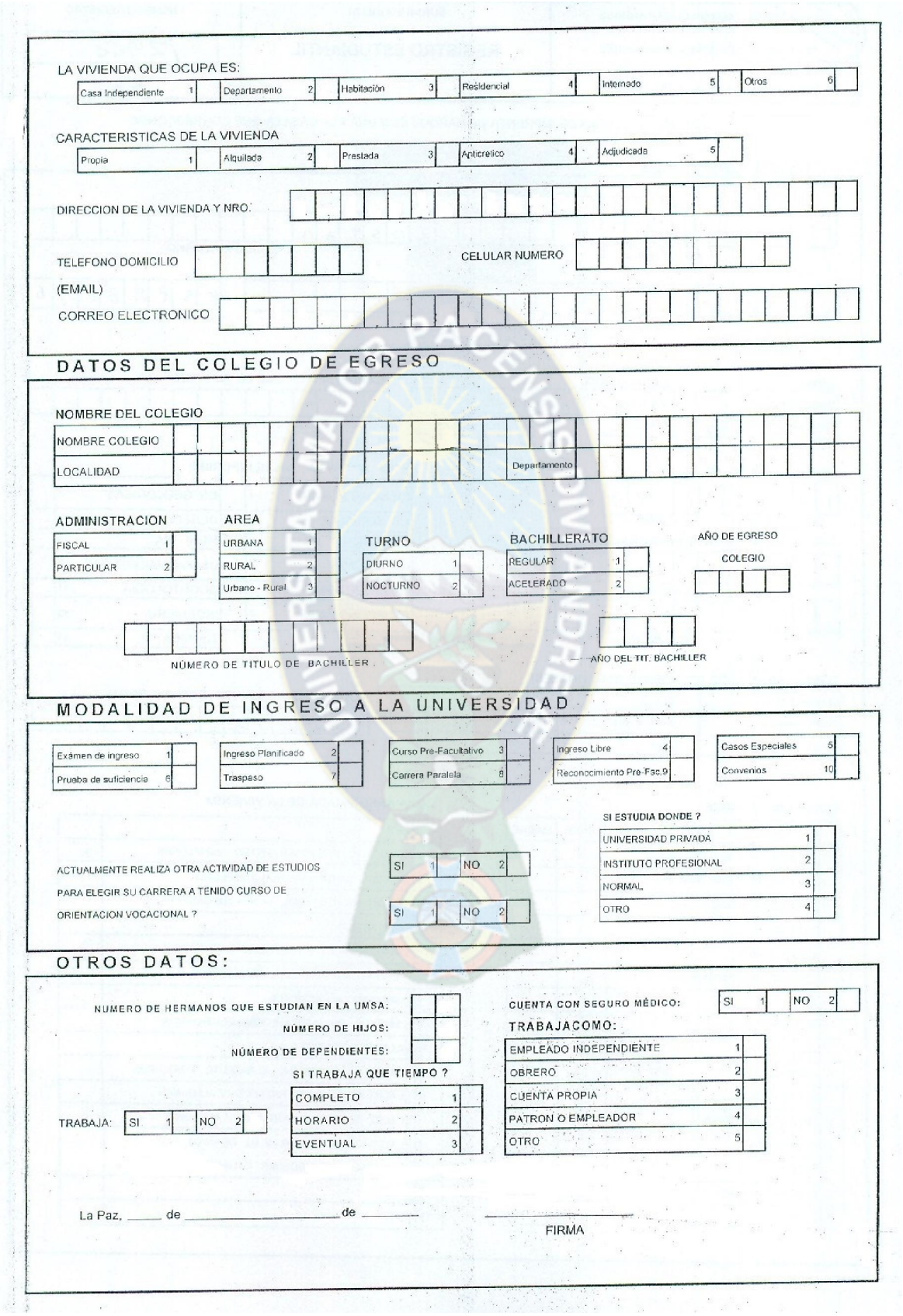
OCUPACION PADRES

		PADRE	MADRE
1	PROFESIONAL INDEPENDIENTE		
2	PROFESIONAL DEPENDIENTE		
3	EMPLEADO PUBLICO		
4	PROFESOR		
5	SECRETARIA		
6	CHOFER PROFESIONAL		
7	OBRERO		
8	ARTESANO		
9	AGRICULTOR		
10	COMERCIANTE		
11	EMPRESARIO		
12	JUBILADO RENTISTA		
13	DESOCUPADO		
14	FALLECIDO		
15	LABORES DE CASA		
16	OTRO		

ZONA APROXIMADA DE LA VIVIENDA

1	SUR	
2	SOPOCACHI - SAN JORGE - PRADO - MIRAFLORES	
3	CENTRAL - SANTA BÁRBARA - SAN PEDRO	
4	AV. PERÚ - NORTE - FINAL SUCRE - YUNGAS	
5	TEMBLADERANI	
6	SAN PEDRO ALTO - GRAN PODER	
7	GARITA DE LIMA - EL TEJAR - COMERCIO	
8	PURA PURA - ACHACHICALA	
9	c. MURILLO - CHURUBAMBA - Plaza EGUINO - VITA	
10	VILLAS DE LA CIUDAD	
11	EL ALTO: T.A.M. - CENTRAL - C. SATÉLITE - V. DOLORES	
12	EL ALTO: V. 16 de JULIO - RIO SECO - V. ALEDAÑAS	
13	EL ALTO: SENKATA - V. SANTIAGO - V. ALEDAÑAS	
14	EL ALTO: CAMINO VIACHA - V. 1ro. DE MAYO VILLA ADELA - VILLA BOLIVAR - CAMINO ORURO	
15	VIACHA	
16	OTROS	

B) Formulario 01 - Registro Estudiantil



LA VIVIENDA QUE OCUPA ES:

Casa Independiente 1	Departamento 2	Habitación 3	Residencial 4	Internado 5	Otros 6

CARACTERISTICAS DE LA VIVIENDA

Propia 1	Alquilada 2	Prestada 3	Anticretico 4	Adjudicada 5

DIRECCION DE LA VIVIENDA Y NRO.

TELEFONO DOMICILIO

CELULAR NUMERO

(EMAIL)

CORREO ELECTRONICO

DATOS DEL COLEGIO DE EGRESO

NOMBRE DEL COLEGIO

NOMBRE COLEGIO

LOCALIDAD

Departamento

ADMINISTRACION

FISCAL 1
PARTICULAR 2

AREA

URBANA 1
RURAL 2
Urbano - Rural 3

TURNO

DIURNO 1
NOCTURNO 2

BACHILLERATO

REGULAR 1
ACELERADO 2

AÑO DE EGRESO COLEGIO

NÚMERO DE TITULO DE BACHILLER

AÑO DEL TIT. BACHILLER

MODALIDAD DE INGRESO A LA UNIVERSIDAD

Exámen de ingreso 1	Ingreso Planificado 2	Curso Pre-Facultativo 3	Ingreso Libre 4	Casos Especiales 5
Prueba de suficiencia 6	Traspaso 7	Carrera Paralela 8	Reconocimiento Pre-Fac.9	Convenios 10

ACTUALMENTE REALIZA OTRA ACTIVIDAD DE ESTUDIOS

SI 1 NO 2

PARA ELEGIR SU CARRERA A TENIDO CURSO DE ORIENTACION VOCACIONAL ?

SI 1 NO 2

SI ESTUDIA DONDE ?

UNIVERSIDAD PRIVADA 1
INSTITUTO PROFESIONAL 2
NORMAL 3
OTRO 4

OTROS DATOS:

NUMERO DE HERMANOS QUE ESTUDIAN EN LA UMSA:

NÚMERO DE HIJOS:

NÚMERO DE DEPENDIENTES:

CUENTA CON SEGURO MÉDICO: SI 1 NO 2

TRABAJACOMO:

EMPLEADO INDEPENDIENTE 1
OBRERO 2
CUENTA PROPIA 3
PATRON O EMPLEADOR 4
OTRO 5

SI TRABAJA QUE TIEMPO ?

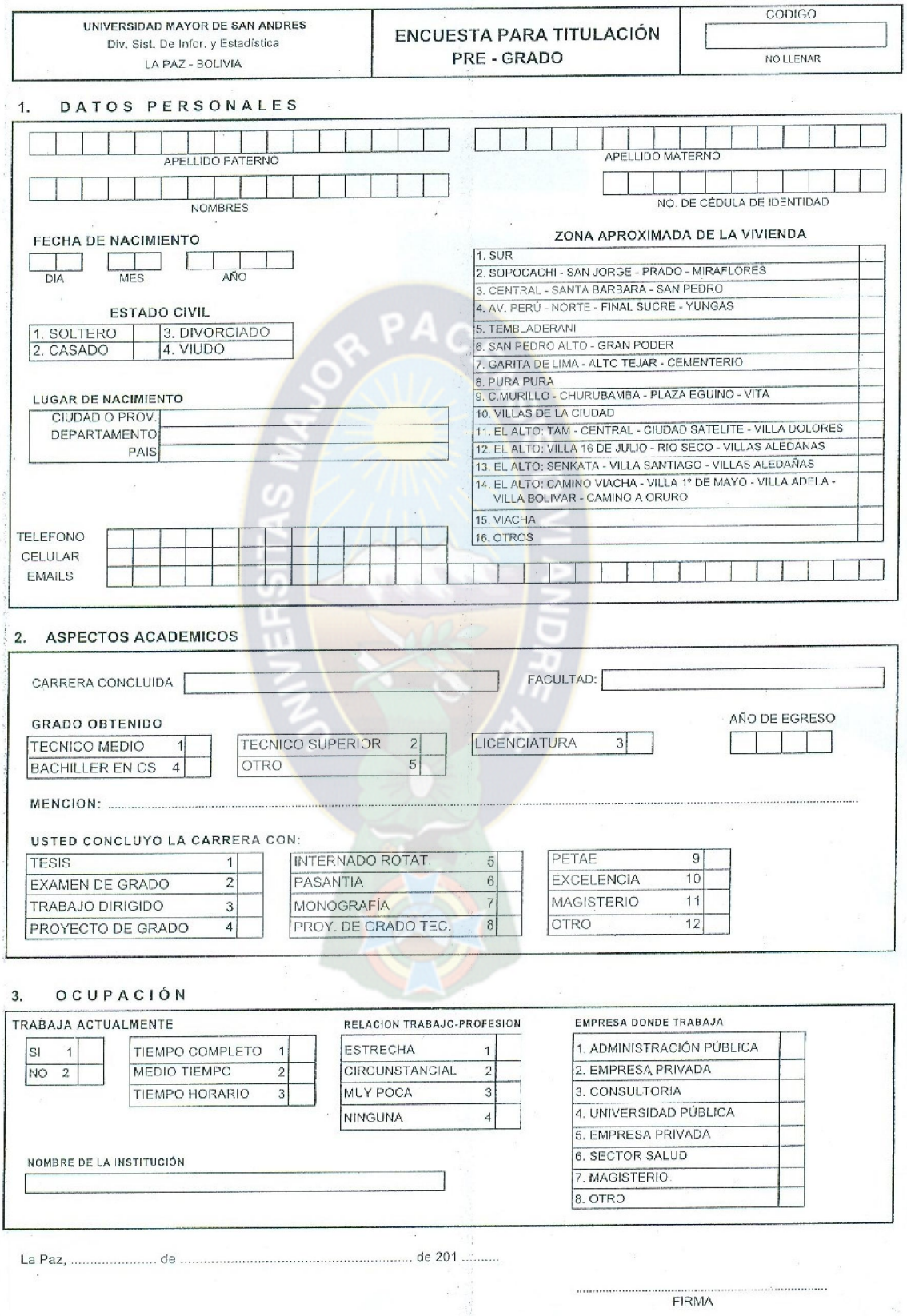
COMPLETO 1
HORARIO 2
EVENTUAL 3

TRABAJA: SI 1 NO 2

La Paz, ____ de _____ de _____

FIRMA

C) Formulario 03 – Registro de estudiantes postulantes a titulación



UNIVERSIDAD MAYOR DE SAN ANDRES
Div. Sist. De Infor. y Estadística
LA PAZ - BOLIVIA

ENCUESTA PARA TITULACIÓN PRE - GRADO

CODIGO
NO LLENAR

1. DATOS PERSONALES

APELLIDO PATERNO

APELLIDO MATERNO

NOMBRES

NO. DE CÉDULA DE IDENTIDAD

FECHA DE NACIMIENTO

DIA MES AÑO

ESTADO CIVIL

1. SOLTERO		3. DIVORCIADO	
2. CASADO		4. VIUDO	

LUGAR DE NACIMIENTO

CIUDAD O PROV.	
DEPARTAMENTO	
PAIS	

ZONA APROXIMADA DE LA VIVIENDA

1. SUR	
2. SOPOCACHI - SAN JORGE - PRADO - MIRAFLORES	
3. CENTRAL - SANTA BARBARA - SAN PEDRO	
4. AV. PERÚ - NORTE - FINAL SUCRE - YUNGAS	
5. TEMBLADERANI	
6. SAN PEDRO ALTO - GRAN PODER	
7. GARITA DE LIMA - ALTO TEJAR - CEMENTERIO	
8. PURA PURA	
9. C.MURILLO - CHURUBAMBA - PLAZA EGUINO - VITA	
10. VILLAS DE LA CIUDAD	
11. EL ALTO: TAM - CENTRAL - CIUDAD SATELITE - VILLA DOLORES	
12. EL ALTO: VILLA 16 DE JULIO - RIO SECO - VILLAS ALEDANAS	
13. EL ALTO: SENKATA - VILLA SANTIAGO - VILLAS ALEDAÑAS	
14. EL ALTO: CAMINO VIACHA - VILLA 1º DE MAYO - VILLA ADELA - VILLA BOLIVAR - CAMINO A ORURO	
15. VIACHA	
16. OTROS	

TELEFONO
CELULAR
EMAILS

2. ASPECTOS ACADEMICOS

CARRERA CONCLUIDA

FACULTAD:

GRADO OBTENIDO

TECNICO MEDIO	1	TECNICO SUPERIOR	2	LICENCIATURA	3
BACHILLER EN CS	4	OTRO	5		

AÑO DE EGRESO

MENCION: .....

USTED CONCLUYO LA CARRERA CON:

TESIS	1	INTERNADO ROTAT.	5	PETAE	9
EXAMEN DE GRADO	2	PASANTIA	6	EXCELENCIA	10
TRABAJO DIRIGIDO	3	MONOGRAFÍA	7	MAGISTERIO	11
PROYECTO DE GRADO	4	PROY. DE GRADO TEC.	8	OTRO	12

3. OCUPACIÓN

TRABAJA ACTUALMENTE

SI 1	
NO 2	

TIEMPO COMPLETO	1
MEDIO TIEMPO	2
TIEMPO HORARIO	3

RELACION TRABAJO-PROFESION

ESTRECHA	1
CIRCUNSTANCIAL	2
MUY POCA	3
NINGUNA	4

EMPRESA DONDE TRABAJA

1. ADMINISTRACIÓN PÚBLICA	
2. EMPRESA PRIVADA	
3. CONSULTORIA	
4. UNIVERSIDAD PÚBLICA	
5. EMPRESA PRIVADA	
6. SECTOR SALUD	
7. MAGISTERIO	
8. OTRO	

NOMBRE DE LA INSTITUCIÓN

La Paz, ... de ..... de 201 .....

.....
FIRMA

Anexo C

Dim_Genero		
Campo	**Tipo**	**Descripción**
id_sexo	varchar(1)	Código que identifica el género de un estudiante.
sexo	varchar(15)	Descripción del tipo de género que corresponde a estudiantes.

Cuadro 1: Diccionario de Dim_Genero

Dim_Estado_Civil		
Campo	**Tipo**	**Descripción**
id_estado_civil	varchar(2)	Código que identifica el estado civil de un estudiante.
estado_civil	varchar(20)	Descripción de tipo de estado civil de un estudiante.

Cuadro 2: Diccionario de Dim_Estado_Civil

Dim_Pais_Nacionalidad		
Campo	**Tipo**	**Descripción**
id_pais_nacionalidad	varchar(3)	Código que identifica la nacionalidad de un estudiante.
pais_nacionalidad	varchar(50)	Nombre del país de nacionalidad.

Cuadro 3: Diccionario de Dim_Pais_Nacionalidad

Dim_Pais_Origen		
Campo	**Tipo**	**Descripción**
id_pais_origen	varchar(3)	Código que identifica el país de origen de un estudiante.
pais_origen	varchar(50)	Nombre del país de origen.

Cuadro 4: Diccionario de Dim_Pais_Origen

Dim_Provincia		
Campo	**Tipo**	**Descripción**
id_provincia	varchar(9)	Código que identifica una provincia.
provincia	varchar(50)	Nombre de la provincia.

Cuadro 5: Diccionario de Dim_Provincia

Dim_Departamento_Nacimiento		
Campo	**Tipo**	**Descripción**
id_dpto_nacimiento	varchar(6)	Código que identifica el departamento de nacimiento de un estudiante.
departamento	varchar(50)	Nombre del departamento.

Cuadro 6: Diccionario de Dim_Departamento_Nacimiento

Dim_Grado_Academico		
Campo	**Tipo**	**Descripción**
id_grado_academico	varchar(3)	Código que identifica el grado académico de un estudiante.
grado_academico	varchar(35)	Nombre del grado académico.
abr_grado_academico	varchar(35)	Abreviatura del nombre de grado académico.

Cuadro 7: Diccionario de Dim_Grado_Academico

Dim_Ocupacion_Padre		
Campo	**Tipo**	**Descripción**
id_ocupacion_padre	varchar(6)	Código que identifica la ocupación laboral del padre.
ocupación_ padre	varchar(50)	Descripción de la ocupación laboral.

Cuadro 8: Diccionario de Dim_Ocupacion_Padre

Dim_Ocupacion_Madre		
Campo	**Tipo**	**Descripción**
id_ocupacion_madre	varchar(6)	Código que identifica la ocupación laboral de la madre.
ocupación_madre	varchar(50)	Descripción de la ocupación laboral.

Cuadro 9: Diccionario de Dim_Ocupacion_Madre

Dim_Vivienda		
Campo	**Tipo**	**Descripción**
id_vivienda	varchar(2)	Código que identifica el tipo de vivienda donde vive el estudiante.
tipo_vivienda	varchar(20)	Descripción del tipo de vivienda.

Cuadro 10: Diccionario de Dim_Vivienda

Dim_Vivienda_Caracteristica		
Campo	**Tipo**	**Descripción**
id_vivienda_caracteristica	varchar(3)	Código que identifica la característica de la vivienda del estudiante.
tipo_caracteristica	varchar(20)	Descripción del tipo de característica de vivienda.

Cuadro 11: Diccionario de Dim_Vivienda_Caracteristica

Dim_Vivienda_Zona		
Campo	**Tipo**	**Descripción**
id_zona	varchar(10)	Código que identifica a una zona específica.
zona	varchar(200)	Nombre de la zona.
circunscripción	number	Circunscripción a la que pertenece la zona.
distrito	varchar(50)	Distrito a la que pertenece la zona.

Cuadro 12: Diccionario de Dim_Vivienda_Zona

Dim_Colegio		
Campo	**Tipo**	**Descripción**
id_colegio	varchar(14)	Código que identifica a un colegio en específico.
colegio	varchar(100)	Nombre del colegio.

Cuadro 13: Diccionario de Dim_Colegio

Dim_Colegio_Administracion		
Campo	**Tipo**	**Descripción**
id_colegio_administracion	varchar(3)	Código que identifica el tipo de administración que tiene un colegio.
administración_c	varchar(20)	Descripción del tipo de administración del colegio.

Cuadro 14: Diccionario de Dim_Colegio_Administracion

Dim_Colegio_Turno		
Campo	**Tipo**	**Descripción**
id_colegio_turno	varchar(3)	Código que identifica el tipo de turno en que realizo estudios de colegio un estudiante.
turno_colegio	varchar(20)	Descripción de tipo de tuno.

Cuadro 15: Diccionario de Dim_Colegio_Turno

Dim_Colegio_Area		
Campo	**Tipo**	**Descripción**
id_colegio_area	varchar(3)	Código que identifica el área de ubicación de un colegio.
tipo_area	varchar(20)	Descripción del tipo de área.

Cuadro 16: Diccionario de Dim_Colegio_Area

Dim_Colegio_Departamento		
Campo	**Tipo**	**Descripción**
id_colegio_departamento	varchar(6)	Código que identifica el departamento donde se encuentra ubicado el colegio.
departamento_colegio	varchar(50)	Nombre del departamento.

Cuadro 17: Diccionario de Dim_Colegio_Departamento

Dim_Colegio_Bachillerato		
Campo	**Tipo**	**Descripción**
id_colegio_bachillerato	varchar(2)	Código que identifica el nivel de bachillerato adquirido de un estudiante dentro del colegio.
tipo_bachillerato	varchar(25)	Descripción del tipo de bachillerato.

Cuadro 18: Diccionario de Dim_Colegio_Bachillerato

Dim_Trabajo_Tiempo		
Campo	**Tipo**	**Descripción**
id_trabajo_tiempo	varchar(3)	Código que identifica el tipo de jornada laboral que tiene un estudiante.
trabajo_tiempo	varchar(20)	Descripción de tipo de jornada laboral.

Cuadro 19: Diccionario de Dim_Trabajo_Tiempo

Dim_Trabajo_Tipo		
Campo	**Tipo**	**Descripción**
id_trabajo_tipo	varchar(3)	Código que identifica el tipo del lugar de trabajo del estudiante.
tipo_empresa	varchar(40)	Descripción de lugar de trabajo.

Cuadro 20: Diccionario de Dim_Trabajo_Tipo

Dim_Trabajo_Ocupacion		
Campo	**Tipo**	**Descripción**
id_trabajo_ocupacion	varchar(6)	Código que identifica la ocupación laboral del estudiante.
trabajo_ocupacion	varchar(50)	Descripción de ocupación laboral.

Dim_Carrera_Programa		
Campo	**Tipo**	**Descripción**
id_programa	varchar(14)	Código que identifica una carrera de la UMSA.
programa	varchar(50)	Nombre de la carrera.

Cuadro 22: Diccionario de Dim_Carrera_Programa

Dim_Facultad		
Campo	**Tipo**	**Descripción**
id_facultad	number	Código que identifica una facultad de la UMSA.
facultad	varchar(100)	Nombre de la facultad.
es_postgrado	varchar(10)	Descripción de la facultad si es de nivel postgrado o pregrado.

Cuadro 23: Diccionario de Dim_Facultad

Dim_Ingreso_UMSA		
Campo	**Tipo**	**Descripción**
id_ingreso_umsa	varchar(3)	Código que identifica el tipo de ingreso a la UMSA con relación al estudiante.
ingreso	varchar(50)	Descripción del tipo de ingreso.

Cuadro 24: Diccionario de Dim_Ingreso_UMSA

Dim_Estudiante_Estado		
Campo	**Tipo**	**Descripción**
id_estudiante_estado	varchar(10)	Código de estado del estudiante dentro la UMSA.
estudiante_estado	varchar(100)	Descripción del tipo de estado del estudiante.

Cuadro 25: Diccionario de Dim_Estudiante_Estado

Dim_Estudiante_Grado		
Campo	**Tipo**	**Descripción**
id_estudiante_grado	varchar(3)	Código del grado académico que adquiere un estudiante al momento de cursar un curso de pregrado o postgrado.
grado_estudi	varchar(35)	Descripción del tipo de grado.
es_postgrad	varchar(10)	Nivel de grado académico si es postgrado o pregrado.

Cuadro 26: Diccionario de Dim_Estudiante_Grado

Dim_Estudiante_Nivel		
Campo	**Tipo**	**Descripción**
id_tipo_nivel	number	Código que identifica el nivel académico que tiene un estudiante.
tipo_nivel	varchar(50)	Descripción del tipo nivel.

Cuadro 27: Diccionario de Dim_Estudiante_Nivel

Dim_Estudiante_Tipo		
Campo	**Tipo**	**Descripción**
id_estudiante_tipo	varchar(3)	Código que identifica el tipo de estudiante si es nuevo o antiguo.
tipo_estudiante	varchar(30)	Descripción del Tipo de estudiante.
abreviatura	varchar(15)	Abreviatura del tipo de estudiante.

Cuadro 28: Diccionario de Dim_Estudiante_Tipo

Dim_Estudiante_Costo_M		
Campo	**Tipo**	**Descripción**
id_estudiante_costo	varchar(15)	Código que identifica el Costo de una matrícula.
categoria	varchar(35)	Categoría de la matricula estudiantil dependiendo del tipo de estudiante.

Cuadro 29: Diccionario de Dim_Estudiante_Costo_M

Dim_Mencion		
Campo	**Tipo**	**Descripción**
id_mencion	varchar(14)	Código que identifica un tipo de mención que tiene una carrera.
mención	varchar(150)	Nombre de mención.

Cuadro 30: Diccionario de Dim_Mencion

Dim_Sede_Localidad		
Campo	**Tipo**	**Descripción**
id_localidad	varchar(15)	Código que identifica la sede donde esa matriculado un estudiante.
localidad	varchar(50)	Nombre de la localidad o sede.

Cuadro 31: Diccionario de Dim_Sede_Localidad

Dim_Estado_Matricula		
Campo	**Tipo**	**Descripción**
id_estado_matricula	varchar(5)	Código que identifica el estado de una matrícula de un estudiante.
estado_matricula	varchar(30)	Descripción del estado.

Cuadro 32: Diccionario de Dim_Estado_Matricula

Dim_Tipo_Titulacion		
Campo	**Tipo**	**Descripción**
Id_tipo_titulacion	varchar(6)	Código que identifica el tipo de modalidad de defensa del para la obtención de un título.
Tipo_titulacion	varchar(50)	Descripción del tipo de modalidad.

Dim_Tipo_Titulo		
Campo	**Tipo**	**Descripción**
Id_tipo_titulo	varchar(2)	Código que identifica el tipo de título que se otorga a un estudiante.
tipo_titulo	varchar(25)	Descripción de tipo de título.

Cuadro 34: Diccionario de Dim_Tipo_Titulo