



UNIVERSIDAD TÉCNICA DEL NORTE

FACULTAD DE INGENIERÍA EN CIENCIAS APLICADAS

CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES

PROYECTO PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO EN SISTEMAS COMPUTACIONALES

TEMA

CONSTRUCCIÓN DE UN DATA MART PARA LA GESTIÓN ACADÉMICA DE LA UNIVERSIDAD TÉCNICA DEL NORTE

AUTOR

JOFFRE OMAR DÍAZ AYALA

DIRECTORA

DRA. CPA. MARÍA DE LA PORTILLA MBA.

Ibarra, abril del 2016



UNIVERSIDAD TÉCNICA DEL NORTE
CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE INVESTIGACIÓN
A FAVOR DE LA UNIVERSIDAD TÉCNICA DEL NORTE

Yo, JOFFRE OMAR DÍAZ AYALA, con cédula de identidad Nro. 1001650348, manifiesto mi voluntad de ceder a la Universidad Técnica del Norte los derechos patrimoniales consagrados en la ley de propiedad intelectual del Ecuador, artículo 4, 5 y 6, en calidad de autor del trabajo de grado denominado: “CONSTRUCCIÓN DE UN DATA MART APLICADO A LA GESTIÓN ACADÉMICA DE LA UNIVERSIDAD TÉCNICA DEL NORTE”, que ha sido desarrollado para optar por el título de Ingeniero en Sistemas Computacionales, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En mi condición de autor me reservo los derechos morales de la obra antes mencionada, aclarando que el trabajo aquí descrito es de mi autoría y que no ha sido previamente presentado para ningún grado o calificación profesional.

En concordancia suscribo este documento en el momento que hago entrega del trabajo final en formato impreso y digital a la biblioteca de la Universidad Técnica del Norte

Nombre: JOFFRE OMAR DÍAZ AYALA

Cédula: 1001650348

Ibarra a los 5 días del mes de abril del 2016



**UNIVERSIDAD TÉCNICA DEL NORTE
BIBLIOTECA UNIVERSITARIA**

**AUTORIZACIÓN DE USO Y PUBLICACIÓN A FAVOR DE LA
UNIVERSIDAD TÉCNICA DEL NORTE**

1. IDENTIFICACIÓN DE LA OBRA

La UNIVERSIDAD TÉCNICA DEL NORTE dentro del proyecto Repositorio Digital institucional determina la necesidad de disponer los textos completos de forma digital con la finalidad de apoyar los procesos de investigación, docencia y extensión de la universidad. Por medio del presente documento dejo sentada mi voluntad de participar en este proyecto, para lo cual ponemos a disposición la siguiente investigación:

DATOS DE CONTACTO	
CÉDULA DE IDENTIDAD	1001650348
APELLIDOS Y NOMBRES	JOFFRE OMAR DÍAZ AYALA
DIRECCIÓN	Calixto Miranda 875 y Tobías Mena
EMAIL	joffrediaza@gmail.com
TELÉFONO FIJO	062956256
TELÉFONO MOVIL	0993897359
DATOS DEL TRABAJO DE GRADO	
TÍTULO	“CONSTRUCCIÓN DE UN DATA MART APLICADO A LA GESTIÓN ACADÉMICA DE LA UNIVERSIDAD TÉCNICA DEL NORTE.”
AUTOR	JOFFRE OMAR DÍAZ AYALA
FECHA	5 de abril del 2016

2. AUTORIZACIÓN DE USO A FAVOR DE LA UNIVERSIDAD

Yo, JOFFRE OMAR DÍAZ AYALA, con cédula de identidad Nro. 1001650348, en calidad de autor y titular de los derechos patrimoniales del trabajo de grado descrito

anteriormente, hago entrega del ejemplar respectivo en forma digital y autorizo a la Universidad Técnica del Norte, la publicación de la obra en el Repositorio Digital Institucional y el uso del archivo digital en la biblioteca de la universidad con fines académicos, para ampliar la disponibilidad del material y como apoyo a la educación, investigación y extensión, en concordancia con la Ley de Educación Superior Artículo 144.

3. CONSTANCIAS

El autor manifiesta que la obra objeto de la presente autorización es original y se la desarrolló, sin violar derechos de autor de terceros, por lo tanto la obra es original y que es el titular de los derechos patrimoniales, por lo que asume la responsabilidad sobre el contenido de la misma y saldrá en defensa de la Universidad en caso de reclamación por parte de terceros.

EL AUTOR:



Nombre: JOFFRE OMAR DÍAZ AYALA

Cédula: 1001650348

Ibarra a los 5 días del mes de abril del 2016

CERTIFICACIÓN

Certifico que el proyecto de Trabajo de Grado “**CONSTRUCCIÓN DE UN DATA MART PARA LA GESTIÓN ACADÉMICA DE LA UNIVERSIDAD TÉCNICA DEL NORTE**”, ha sido realizado en su totalidad por el señor: Joffre Omar Díaz Ayala portador de la cédula de identidad número: 1001650348



Dra. Cpa. María De la Portilla Mba.
DIRECTORA

DEDICATORIA

El presente trabajo está dedicado a mis hijos Diana, Viviana, Cristian,
a mi esposa Luz Marina, quienes han sido fuentes de inspiración
y de la fortaleza que necesitaba durante todo este tiempo.

Dedico este trabajo de manera muy especial a
Ernesto Díaz y Fabiola Ayala, mis padres,
Por su apoyo incondicional en todo sentido.

AGRADECIMIENTOS

A dios, por haberme dado la oportunidad de llegar hasta aquí.

A la Universidad Técnica del Norte y su cuerpo docente por la formación recibida.

A la Dra. María De la Portilla por su paciencia como directora
del presente trabajo.

Al Ing. Pedro Granda, por las orientaciones y ayudas concedidas
durante el desarrollo de su asignatura.

Al personal del Departamento de Desarrollo Tecnológico e Informático de la
Universidad Técnica del Norte

ÍNDICE

1.	INTRODUCCIÓN	16
1.1.	ANTECEDENTES	17
1.2.	PROBLEMA.....	18
1.3.	OBJETIVOS	18
1.3.1.	OBJETIVO GENERAL.....	18
1.3.2.	OBJETIVOS ESPECÍFICOS.....	18
1.4.	ALCANCE.....	19
1.5.	JUSTIFICACIÓN	20
	CAPÍTULO II	22
	MARCO TEÓRICO.....	22
2.	SOPORTE A LA TOMA DE DECISIONES CON BI.....	22
2.1.	Toma de decisiones organizacionales.....	22
2.1.1.	El proceso de toma de decisiones.....	23
2.2.	Información y conocimiento.....	24
2.3.	Gestión del conocimiento	25
2.3.1.	Las tecnologías de información	27
2.3.2.	Sistema de Información	28
2.4.	Tipos básicos de sistemas de información.....	29
2.4.1.	Sistemas de información transaccionales.....	29
2.4.2.	Sistemas estratégicos.....	30
2.4.3.	Sistemas de apoyo de las decisiones.....	30
2.5.	Inteligencia de Negocios (BI)	32
2.5.1.	Necesidad de la Inteligencia de Negocios.....	32
2.5.2.	Importancia de la Inteligencia de Negocios.....	33
2.6.	Componentes de Inteligencia de Negocios.....	34
2.7.	Herramientas principales de inteligencia de negocios	35
2.7.1.	Almacén de datos (Data Warehouse).....	35
2.7.2.	Consulta y Reporte (Query & Report)	37
2.7.3.	Tienda de datos (Data mart).....	38
2.7.4.	Herramientas OLAP.....	39
2.7.5.	Minería de datos (Data Mining).....	40
2.8.	Introducción al modelado multidimensional.....	41
2.8.1.	Hechos	43
2.8.2.	Dimensiones.....	44

2.8.3. Medidas.....	46
2.9. Operaciones básicas OLAP.....	47
2.9.1. Drill Down y Roll up	47
2.9.2. Slice y Dice	47
2.9.3. Drill across	48
2.9.4. Pivoting.....	48
2.10. Arquitectura de los modelos analíticos	48
2.10.1. Modelo en estrella.....	49
2.10.2. Modelo copo de nieve.....	50
2.10.3. Modelo 3FN.....	52
2.11. Arquitectura de un Data warehouse	53
2.11.1. Arquitectura data warehouse básica	54
2.11.2. Arquitectura data warehouse con área de organización.....	55
2.11.3. Arquitectura Data Warehouse con Área de Organización y Data Marts	56
2.12. Oracle Business Intelligence.....	57
2.12.1. Oracle Business Intelligence Standard Edition (OBISE).....	57
2.12.2. Oracle Business Intelligence Standard Edition One (BISE1).....	58
2.12.3. Oracle Business Intelligence Enterprise Edition (BIEE)	61
2.12.4. Oracle Business Intelligence Server.	62
2.12.5. Oracle Business Intelligence Server Administrator.	64
2.12.6. Oracle Business Intelligence Answers.....	64
2.12.7. Oracle Business Intelligence Interactive Dashboards.....	64
2.12.8. Oracle Business Intelligence Publisher.....	65
2.12.9. Oracle Database Standard Edition One.....	65
2.12.10. Oracle Warehouse Builder.....	66
CAPÍTULO III.....	67
METODOLOGÍA	67
3. PRINCIPALES METODOLOGÍAS ACTUALES.....	67
3.1. Metodología propuesta por Bill Inmon.....	68
3.2. Rapid Warehousing Methodology	70
3.2.1. Definición de los objetivos	71
3.2.2. Definición de los requerimientos de información.....	71
3.2.3. Diseño y modelización.....	71
3.2.4. Implementación.....	72
3.2.5. Revisión	72
3.2.6. Gestión del Proyecto.....	73

3.3. Metodología propuesta por Ralph Kimball.....	73
3.3.1. Planificación del Proyecto	74
3.3.2. Definición de los Requerimientos.....	74
3.3.3. Línea Tecnológica.....	75
3.3.3.1. Diseño de la Arquitectura Técnica.....	75
3.3.3.2. Selección de Productos e Instalación.....	76
3.3.4. Línea de datos	76
3.3.4.1. Modelado Dimensional.....	76
3.3.4.2. Diseño Físico	77
3.3.4.3. Diseño del subsistema de procesos ETL.....	77
3.3.5. Línea de aplicaciones del BI	77
3.3.5.1. Diseño de aplicaciones del BI.....	78
3.3.5.2. Desarrollo de aplicaciones del BI	78
3.3.6. Despliegue.....	79
3.3.6.1. Implementación	79
3.3.6.2. Mantenimiento y Crecimiento	79
3.3.7. Gerenciamiento del Proyecto	80
3.4. Selección de la metodología	81
IMPLEMENTACIÓN Y DESARROLLO.....	82
4. Desarrollo de la metodología seleccionada.....	82
4.1. Planeación del proyecto	82
4.2. Definición de requerimientos.....	82
4.2.1. Matrícula.....	83
4.2.2. Promoción.....	83
4.2.3. Asignación de carga horaria a docentes.....	84
4.2.4. Graduación.....	84
4.2.5. Análisis de la fuente de datos.....	85
4.2.6. Calidad de datos	86
4.3. Línea Tecnológica.....	86
4.3.1. Arquitectura tecnológica.....	86
4.3.2. Selección de productos e instalación	87
4.4. Línea de Datos (Modelamiento Dimensional).....	88
4.4.1. Elección de la granularidad.....	88
4.4.2. Identificación de hechos	90
4.4.3. Elección y detalle de las dimensiones.....	91
4.4.3.1. Dimensión Estudiante.....	91

4.4.3.2. Dimensión Tiempo	92
4.4.3.3. Dimensión Matricula	93
4.4.3.4. Dimensión Nómina	93
4.4.3.5. Dimensión Materias	94
4.4.3.6. Dimensión Dependencia	94
4.4.3.7. Dimensión Docentes	94
4.4.3.8. Dimensión Ciclos Académicos	95
4.5. Modelo físico	96
4.5.1. Cubos Estudiantes y Docentes	99
4.6. Extracción, transformación y carga (ETL)	102
4.7. Línea de aplicación de Inteligencia de Negocios	103
4.7.1. Diseño del Business Intelligence (BI)	103
4.7.2. Desarrollo del BI	104
CAPÍTULO V	108
CONCLUSIONES Y RECOMENDACIONES	108
5.1. Conclusiones	108
5.2. Recomendaciones	109

ÍNDICE DE TABLAS

<i>Tabla 1: Levantamiento de reportes.....</i>	<i>20</i>
<i>Tabla 2: fuente de datos para los data mart.....</i>	<i>86</i>
<i>Tabla 3: Análisis de Estudiantes</i>	<i>89</i>
<i>Tabla 4: análisis de docentes</i>	<i>90</i>
<i>Tabla 5: Hechos o medidas a considerar en el proyecto</i>	<i>90</i>

INDICE DE FIGURAS

Figura 1: Toma de decisiones	22
Figura 2: El proceso de toma de decisiones	23
Figura 3: Construcción del conocimiento empresarial.....	25
Figura 4: Gestión del conocimiento	26
Figura 5: Tecnologías de la Información y Comunicación	28
Figura 6: Arquitectura de Inteligencia de Negocios.....	35
Figura 7: Data warehouse.....	36
Figura 8: Tablero de mando (Dashboard).....	38
Figura 9: Data mart como subconjunto de un Data warehouse	39
Figura 10: Cubo OLAP.....	40
Figura 11: Data Mining (Minería de Datos).....	41
Figura 12: Ejemplo de cubo sobre notas de un estudiante en una materia.	42
Figura 13: Tabla de hechos ventas.....	44
Figura 14: Tabla de la dimensión producto	45
Figura 15: Esquema en estrella de un Data warehouse.....	50
Figura 16: Esquema copo de nieve de un Data warehouse	51
Figura 17: Esquema 3FN	52
Figura 18: Estructura básica de un data warehouse.....	53
Figura 19: Arquitectura Data warehouse básica.....	54
Figura 20: Arquitectura Data warehouse con área de organización.....	55
Figura 21: Arquitectura Data warehouse con área de organización y Data marts	56
Figura 22: Arquitectura de OBISE One	59
Figura 23: Oracle BI Server	63
Figura 24: Building the Data warehouse	69
Figura 25: Metodología Rapid Warehouse Builder	70
Figura 26: Metodología de Ralph Kimball.....	74
Figura 27: Línea Tecnológica	75
Figura 28: Línea de datos	77
Figura 29: Línea de aplicación del BI	78
Figura 30: Despliegue	79
Figura 31: Administración del proyecto	80
Figura 32: Arquitectura Tecnológica OBI - UTN.....	87
Figura 33: Dimensión Estudiante	92
Figura 34: Dimensión Período.....	92
Figura 35: Dimensión Matrícula.....	93
Figura 36: Dimensión Período.....	93
Figura 37: Dimensión Asignatura	94
Figura 37: Dimensión Dependencia.....	94
Figura 39: Dimensión Dependencia.....	95
Figura 40: Dimensión Período.....	95
Figura 41: Creación de Ubicaciones en OWB 11g.....	96
Figura 42: Creación de módulos en OWB 11g	97
Figura 43: Importación de metadatos en OWB 11g.....	97
Figura 44: Creación de dimensiones.....	98
Figura 45: Árbol del proyecto.....	99
Figura 46: Esquema en estrella Aprobados	100
Figura 47: Esquema en estrella Graduados.....	100
Figura 48: Esquema en estrella Matriculados	101
Figura 49: Esquema en estrella Docentes.....	101
Figura 50: Procesos ETL.....	102
Figura 51: Flujo de Procesos.....	103

<i>Figura 52: Herramienta de administración de Oracle Business Intelligence</i>	<i>104</i>
<i>Figura 53: Ingreso al sistema.....</i>	<i>105</i>
<i>Figura 54: Estudiantes matriculados</i>	<i>106</i>
<i>Figura 55: Reporte de estudiantes matriculados</i>	<i>106</i>
<i>Figura 56: Estudiantes con 1ra, 2da y 3ra matrícula por género y etnia</i>	<i>107</i>
<i>Figura 57: Docentes según su relación laboral.....</i>	<i>107</i>

RESUMEN

En el presente trabajo se ha propuesto con el fin de apoyar la toma de decisiones que se realizan en el vicerrectorado académico de la Universidad Técnica del Norte, utilizando herramientas de inteligencia de negocios para el desarrollo del software necesario. Se ha optado por la mitología de Ralph Kimball para el diseño de la base de datos multidimensional.

El capítulo I presenta la idea que motivó el emprender este trabajo, en el capítulo II nos introduce en los conceptos de Inteligencia de Negocios. Las metodologías más populares para el desarrollo de un proyecto de data warehouse se detallan en el capítulo III, la implementación de los data mart se describe en el capítulo IV y finalmente en el capítulo V se redactan las conclusiones y recomendaciones.

ABSTRACT

This project has been set with the purpose of supporting the decision making ability by the Academic Vicerector Ship of the North Technical University; using the business intelligence tools for the development of the necessary software. Ralph Kimball methodology has been chosen the design of a multidimensional data base.

Chapter 1 encloses the main objective of this work, chapter 2 introduces business intelligence concepts, chapter 3 exposes the most popular methodologies for the development of a data warehouse projects. The implementation of the data mart is described in chapter 4 and finally, the conclusions and recommendations are summarized in chapter 5.

CAPÍTULO I

1. INTRODUCCIÓN

Durante mucho tiempo las empresas han contado con apoyo informático a través de aplicaciones operativas con bases de datos voluminosas, en este ambiente los datos fríos y simples no nos dan mayor información acerca de cómo marcha la empresa para entonces poder tomar decisiones acertadas. Durante las dos últimas décadas han surgido nuevas herramientas que posibilitan el desarrollo de aplicaciones de Inteligencia de Negocios (BI) que permiten a los ejecutivos de las empresas visualizar, entender y analizar sus datos, proporcionando una base de conocimientos confiable que refleja la trayectoria de la organización.

Tradicionalmente las aplicaciones de BI permiten a los responsables de la toma de decisiones la adquisición de conocimientos útiles sobre el rendimiento y los problemas de negocio, mediante información almacenada y por medio del uso de una serie de herramientas tecnológicas tales como almacenes de datos, mercado de datos, minería de datos, gestión del rendimiento empresarial, OLAP, OLTP e informes de gestión periódicos.

La investigación en estas áreas ha producido soluciones robustas, técnicamente desarrolladas y en base a diversas metodologías, con una variedad de productos comerciales disponibles que se basan en resultados obtenidos de las aplicaciones de inteligencia de negocios y la web.

Modelos, sistemas y tecnologías resumen los avances actuales de la investigación en inteligencia de negocios y la web, haciendo hincapié en soluciones que combinan ambas áreas en beneficio de la construcción de mejores aplicaciones de BI para dar soporte a los ejecutivos en la toma de decisiones y orientar el rumbo que deben seguir sus organizaciones.

Inmersas en este ámbito se encuentran también las instituciones educativas universitarias, y más aún cuando están en proceso de evaluación por los organismos de control y seguimiento de la educación superior del Estado Ecuatoriano, a quienes se les debe entregar indicadores que dan cuenta de la calidad de la educación que estas entidades brindan a sus estudiantes. Ante esta realidad se propone la realización del presente proyecto con el que se diseñarán los data mart propuestos para apoyar la gestión académica de la Universidad Técnica del Norte específicamente al Vice Rectorado Académico.

1.1. ANTECEDENTES

La Universidad Técnica del Norte desde que fue creada hace 27 años, durante este tiempo ha venido progresando en todo ámbito, y para ello ha sido necesario una planificación adecuada. La misma que en la parte académica se ha basado en estadísticas sobre los resultados académicos de sus estudiantes. Hace 14 años se implementa e instala un sistema informático integrado de administración académica.

Antes del año 2000 se realizaba esta tarea de forma manual y apoyada en hojas de cálculo desarrolladas en Excel, siendo la FICA la pionera en el desarrollo de un sistema de control de académico, en la actualidad ya se cuenta en toda la universidad con un sistema de control académico que funciona a través del portal de la institución, pero sin embargo los datos generados por los sistemas transaccionales no se ha utilizado como soporte a la toma de decisiones de la Universidad por la falta de un software especializado en este campo.

Durante estos últimos meses y ante las exigencias de los organismos de control del Ministerio de Educación y del Gobierno Nacional, se desarrollaron ciertos reportes sobre los resultados de la formación de los estudiantes utilizando un software que permitió presentar ciertos datos que requerían los evaluadores del CEAACES y SENESCYT.

Durante todo el tiempo de vida de la Universidad la gestión académica se la realizó en base a la experiencia de los docentes y directivos en otras instituciones y la experiencia adquirida día a día en el proceso educativo de la Universidad.

1.2. PROBLEMA

La gestión académica de la Universidad Técnica del Norte se basa en proceso de sistemático de análisis estadístico de datos realizado en forma casi manual, y en otros casos solicitando al departamento de informática reportes específicos para el seguimiento del proceso de aprendizaje de los estudiantes y la situación laboral de la planta docente.

1.3. OBJETIVOS

1.3.1. OBJETIVO GENERAL

Implementar una Solución de Inteligencia de Negocios mediante el diseño de una aplicación en Oracle Business Intelligence para mejorar la toma de decisiones de la Gestión Académica de la Universidad Técnica del Norte.

1.3.2. OBJETIVOS ESPECÍFICOS

- a) Determinar los requerimientos de reportes a nivel administrativo académico de la universidad, manteniendo reuniones con la Vicerrectora Académica de la Universidad para establecer el punto de partida en el desarrollo de la presente aplicación.
- b) Estudiar la herramienta de Inteligencia de Negocios de Oracle realizando prácticas sencillas llegando a un manejo eficiente y poder aplicarla en el desarrollo de la aplicación planteada.
- c) Determinar cuáles son los procesos involucrados en la gestión académica de la universidad, dialogando con las autoridades responsables y establecer de esta manera la prioridad de cada uno.

- d) Implementar una Solución de Inteligencia de Negocios, de apoyo a la toma de decisiones de la Gestión Académica de la Universidad Técnica del Norte, diseñando e implementando los data mart para cada proceso involucrado.

1.4. ALCANCE

El presente proyecto permitirá a las principales autoridades de la Universidad: Rectorado, Vicerrectorado Académico, Vicerrectorado Administrativo, Coordinador de Planeamiento Integral, Decanos, Sub Decanos y Coordinadores de Carrera, contar con información oportuna, real y competente que apoye la labor de una de las cuatro funciones principales de la universidad que es la académica, para el efecto contarán con reportes que se ajustan a los indicadores de las cuatro funciones como se muestran en la Tabla 1.

Con este fin se desarrollará el módulo de extracción, transformación y carga de datos, se diseñará e implementarán los Data mart y procesos necesarios para realizar análisis y generar información para los reportes solicitados en la Tabla 1.

ESTUDIANTES	
Estudiantes con 1ra, 2da y 3ra matrícula	
Cuántos se inscribieron	Por género Por etnia Por facultad Por carrera Por materia
Cuántos se graduaron tema de tesis	
Cuántos se matricularon	
Cuántos aprobaron	
Cuántos se retiraron	
Cuántos perdieron	
Cuántos anularon	
Cuáles tienen proyectos de Investigación.	
Cuáles tienen proyectos de vinculación.	
DOCENTES	
Todos los contratos de cada uno, cuando firmaron.	

Cuántos están a tiempo completo, medio y cuarto.
Cuanto tiempo dedican a formación.
Carga horaria por período académico.
Docentes que tienen horas de investigación.
Docentes que tienen horas de vinculación.
Cuál es la carga horaria en cada unidad académica.

Tabla 1: Levantamiento de reportes

Fuente: Elaboración propia

Las fuentes de las cuales se obtendrá los datos son el sistema académico, de recursos humanos y de nómina, que se encuentran en la base de datos del sistema integrado de la Universidad Técnica del Norte, específicamente se tomarán en cuenta los datos de las tablas que corresponden a las áreas mencionadas al inicio de este párrafo, ahí se encuentran almacenados los datos que permitirán cumplir con los requerimientos de la tabla 1.

1.5. JUSTIFICACIÓN

En la actualidad es de mucha importancia para las instituciones educativas de nivel superior el contar con herramientas informáticas apropiadas que les permita obtener información, para tomar decisiones estratégicas acertadas orientadas a mejorar la calidad de los procesos de enseñanza.

Tener a disposición esta información permitirá a los directivos obtener un panorama claro sobre el rendimiento de los estudiantes y desempeño docente, en forma individual y grupal, a fin de que se puedan ajustar los procesos educativos según las necesidades cada vez más exigentes de nuestra sociedad.

El Gobierno Nacional por medio de los organismos de control de la educación superior como son: CES, CEAACES, SEMPLADES y SENESCYT, entre otros, quienes exigen a las Instituciones de Educación Superior (IES) mucha información acerca de

los resultados que se obtienen en la formación de nuevos profesionales, y de esta manera hacer un seguimiento permanente que se resumen en valores cuantitativos y cualitativos en un proceso de evaluación y acreditación.

La Universidad cuenta actualmente con el software apropiado para el desarrollo de la aplicación con sus licencias respectivas, se trata de la Herramienta de Inteligencia de Negocios de Oracle versión 11g, tomando en cuenta que las bases de datos de la Universidad han sido diseñadas en Oracle.

CAPÍTULO II

MARCO TEÓRICO

2. SOPORTE A LA TOMA DE DECISIONES CON BI

2.1. Toma de decisiones organizacionales.

Peña (2006) afirma. “La toma de decisiones es una actividad intelectual que una persona realiza para escoger un curso de acción o elegir un objeto determinado de entre varias opciones, con el objeto de satisfacer una necesidad específica.” (p. 14).



Figura 1: Toma de decisiones

Fuente: Elaboración propia

El éxito o el fracaso de toda organización están directamente relacionado con las decisiones que tomen sus ejecutivos de nivel estratégico, táctico y operativo. En la actualidad la decisión que es tomada por un gerente o directivo de una empresa no puede basarse solo en la experiencia que tienen las personas que están al frente de la organización (Calzada y Abreu, 2009). Es claro que se debe contar con información confiable, certera y a tiempo, sobre el estado de la organización así como del medio

en el que realiza sus operaciones, disponer de información incorrecta sería fatal para la empresa y la conduciría al fracaso.

2.1.1. El proceso de toma de decisiones.

El proceso de toma de decisiones es una actividad intelectual que la realizamos todos en algún momento de nuestras vidas y en un periodo de tiempo, para poder ejecutar esta actividad necesitamos de un conocimiento previo del ámbito, características y circunstancias en la que nos encontramos; en la medida de la calidad de información que tengamos sobre estos aspectos se podrá generar conocimiento empresarial con el cual nuestras decisiones serán más acertadas. (Peña, 2006).



Figura 2: El proceso de toma de decisiones

Fuente:

http://www.ecopetrol.com.co/especiales/ReporteGESTIÓN2012/practicas_laborales_07

Si el ámbito en el que nos encontramos es una empresa, y en medio de una reunión para decidir sobre una nueva línea de productos y servicios, se debe tener a mano información fidedigna sobre las características del mercado en ese momento, las tendencias de los clientes, entre otros, para poder conducir al éxito los nuevos productos y servicios que pretende lanzar al mercado en la organización.

Ante esta realidad los ejecutivos de hoy en día necesitan de poderosas herramientas que les puedan apoyar en su difícil tarea de tomar las mejores y más acertadas decisiones, y es esta precisamente la gran ventaja que proporcionan las aplicaciones de Inteligencia de Negocios.

2.2. Información y conocimiento.

En el punto anterior se habla de la información y conocimiento el primero como un recurso imprescindible y el segundo como un producto del análisis de esta, para aplicarlos en el proceso de toma de decisiones empresariales, al respecto no se deben confundir datos, información y conocimiento. Cuando se hace referencia a un dato se trata de una unidad básica de información que por sí sola carece de significado como puede ser un color, la hora, la edad, una calificación y muchos otros. Con el advenimiento de la sociedad de la información las empresas comenzaron a almacenar los datos producto de sus operaciones en medios magnéticos, con el fin de poder analizarlos más tarde por medio de sus sistemas transaccionales a través de una variedad de reportes que estos ofrecen.

Peña (2006) sostiene que, para comprender el entorno de la administración del conocimiento, es necesario considerar tres conceptos claves que son: datos, información y conocimiento. Los datos son hechos aislados sin significado ni explicación y constituyen la materia prima para la creación de información.

Mientras que la Información es el resultado de la organización y tratamiento que se aplica a los datos para producir un significado adicional al que se les brinda de manera aislada, el conocimiento es el resultado de un proceso con un mayor grado de abstracción y síntesis de lo que significa esta información para la empresa al relacionarla con el medio en el que se desenvuelve. Ver figura 3.



Figura 3: Construcción del conocimiento empresarial

Fuente: Elaboración propia

Peña (2006) afirma “En cuanto al conocimiento, este representa un mayor grado de abstracción y síntesis del significado de la información al asociar el contexto en el que se inscribe” (p. 8).

El desarrollo tecnológico en la ingeniería del software hace posible el desarrollo de sistemas informáticos que permiten el tratamiento y organización de estos datos para producir información, ejemplos de estos son los llamados sistemas de información gerencial, sistemas de información empresarial, sistemas estratégicos, entre otros.

2.3. Gestión del conocimiento

Al hablar de gestión del conocimiento tenemos que tomar en cuenta los conceptos de administración y conocimiento, en el primer término las ciencias administrativas proveen de las herramientas necesarias para realizar un proceso administrativo y que al interior de cualquier organización es vital y se lo practica a diario. En cuanto al conocimiento ya se ha hablado sobre él en apartados anteriores.

Sin embargo al tratar de definir lo que es la administración del conocimiento o al tratar de establecer los pasos necesarios para realizarlo, vienen las dificultades en razón de que son temas de reciente aparición y al que algunos investigadores en este campo se plantean preguntas como: ¿de quién es el conocimiento?, ¿dónde está el conocimiento?, preguntas que tal vez no tengan respuesta. Lo que sí es seguro es que

el conocimiento carece de validez si no es transmitido a quién lo necesita para aplicarlo y de esta manera obtener ventajas en las actividades que se desempeñan.

Muchas empresas en la actualidad realizan procesos de administración sobre los datos que disponen o son generados al interior de la empresa, para su posterior almacenamiento en sus bases de datos por medio de sus sistemas de información. Estos datos no siempre están disponibles a tiempo y de una manera eficaz para que los ejecutivos puedan apoyar sus decisiones basándose en la realidad presente en esos datos almacenados, aun cuando disponemos en la actualidad de la tecnología apropiada para hacerlo.

Según Davenport y Prusak (como se citó en Calzada y Abreu, 2009) la administración del conocimiento es una disciplina que articula personas, procesos, contenido y tecnología. Martínez (1999) afirma “engendrar conocimiento, reunirlo, compartirlo, distribuirlo y aplicarlo para la gestión de la organización, con acciones que sirvan para crear valor añadido, diferenciador y que genere una mayor eficacia en todas sus áreas es lo que se denomina gestión del conocimiento” (p.49).



Figura 4: Gestión del conocimiento

Fuente: <http://recursosdeeconomiadelaeducacion.blogspot.com/>

En muchas organizaciones se confunde información con conocimiento y confían en sus sistemas de información transaccionales para obtenerlo, los directivos no disponen del conocimiento requerido sobre la realidad en la que se encuentran sus empresas y toman decisiones a ciegas en algunos casos, y en otros confiados en lo poco que pueden obtener de sus sistemas tradicionales, decisiones que no siempre representan los mejores aciertos en la gestión.

Es evidente que cuando el conocimiento no es aprovechado en beneficio de los demás carece de valor, este es el producido por el cerebro de los seres humanos, quienes forman parte de alguna empresa que a su vez, en medio de la interacción diaria con todos los miembros de la organización generan conocimiento para la empresa, el mismo que está oculto y acumulado en patentes, planos, software, entre otros. Para poder explotar esta base de conocimiento empresarial es necesario extraerlo, comunicarlo y aplicarlo, es aquí donde intervienen los procesos y la tecnología.

Entonces se deben integrar la infraestructura tecnológica con los recursos humanos y la información para generar conocimiento, juntarlo y publicarlo para que pueda ser aprovechado a todo nivel al interior de una organización. Entonces le permitirán a la empresa posicionarse en un lugar privilegiado frente a sus similares debido a que sus ejecutivos tienen argumentos firmes en los que apoyan las decisiones que toman.

2.3.1. Las tecnologías de información

La Asociación de la Tecnología de Información de América (ITAA) define a la tecnología de la información como: “el estudio, diseño, desarrollo, implementación, soporte o dirección de los sistemas de información computarizados, en particular de software de aplicación y hardware de computadoras”. Se puede definir entonces a la tecnología de la información como el conjunto de herramientas y recursos informáticos que se ocupan del tratamiento y administración de la información de forma automática, para que pueda ser aprovechada en beneficio de quienes la necesitan.



Figura 5: Tecnologías de la Información y Comunicación

Fuente: <http://www.edukanda.es>

La humanidad ha tenido que pasar por varias eras durante su evolución y desarrollo, y en la actualidad la humanidad es protagonista del gran avance que está teniendo la llamada era de la información o sociedad de la información. Las empresas de hoy en día le dan bastante importancia a la información, quienes para su aprovechamiento han invertido mucho en tecnología informática y de comunicaciones para lograr su tratamiento adecuado y eficaz.

Parte de esta tecnología informática la constituyen los sistemas de información que en un principio estaban orientados a la obtención, tratamiento y almacenamientos de datos provenientes de las operaciones diarias de la empresa.

2.3.2. Sistema de Información

Un sistema de información es el conjunto de procesos interrelacionados que se encargan de la captura, almacenamiento, procesamiento y presentación de datos, para viabilizar de manera más eficaz las actividades diarias de una organización.

En la actualidad los sistemas de información cumplen tres objetivos básicos en las organizaciones: automatización de procesos operativos, apoyo al proceso de toma de decisiones y logro de ventajas competitivas.

2.4. Tipos básicos de sistemas de información.

2.4.1. Sistemas de información transaccionales.

Desde su aparición su objetivo principal estaba encaminado a la automatización de los procesos operativos dentro de una empresa como: ingreso y salida de mercadería, ventas, depósitos, retiros, entre otros, son llamados sistemas transaccionales. (Peralta, 2000) Los usuarios típicos de esta clase de sistemas están a nivel operativo y sus principales características son:

- a) Estos son los sistemas que primero se adquieren o se desarrollan en una organización que empieza a introducir tecnología en sus operaciones.
- b) Esta clase de sistemas proveen operatividad a tareas comunes y que son realizadas todos los días de forma repetitiva y que influye en la economía de la empresa debido a que puede prescindir de cierta mano de obra.
- c) Captan muchos datos de entrada, realizan cálculos sencillos y poco sofisticados y así mismo producen mucha información de salida por medio de sus reportes.
- d) Acumulan grandes cantidades de información en sus bases de datos que son diseñadas generalmente con un modelo relacional, haciendo necesaria la inversión en potentes sistemas manejadores de bases de datos y hardware con gran capacidad de procesamiento.
- e) Se puede justificar la inversión en este tipo de sistemas sin ninguna complicación debido a que los beneficios son visibles casi inmediatamente, pues su inventario es actualizado inmediatamente, los pedidos se realizan en instante en que el stock es el mínimo, la contabilidad lleva sus procesos de manera automática.

2.4.2. Sistemas estratégicos.

Este tipo de sistema se desarrolla al interior de la organización por el departamento de sistemas, con el fin de lograr ventajas competitivas a través del uso de la tecnología de información, se los llama sistemas estratégicos. (Peralta, 2000)

“Los sistemas estratégicos están orientados a soportar la toma de decisiones, facilitando las tareas de dirección y proporcionando a esta un soporte básico” (Heredero, López, Romo y Medina, 2004 p. 41).

Sus principales características son:

- a) Lograr que la organización esté un paso adelante que sus similares en la lucha por captar la mayor parte del mercado.
- b) Es un software propietario, ya que es diseñado e implementado por los programadores del departamento de sistemas de las empresas, razón por la que la integración con otros paquetes implica mayores dificultades.
- c) Su desarrollo es modular, es decir primero empiezan con un área a la que quieren atender con su solución y luego continúan creciendo hacia toda la empresa desarrollando los módulos que hacen falta.

2.4.3. Sistemas de apoyo de las decisiones.

Tal y como lo sugiere el nombre esta clase de sistemas fueron creados para ofrecer criterios en los que se puedan apoyar las decisiones tomadas por los ejecutivos de una empresa. Esta clase de sistemas se implantan posteriormente a la implementación e instalación de los transaccionales, tienen como objetivo apoyar el proceso de toma de decisiones, ha estos se los conoce como sistemas de soporte a la toma de decisiones. (Peralta, 2000)

Son sistemas informáticos que constituyen un apoyo eficaz para los procesos de toma de decisiones de un tipo determinado, puesto que facilita el acceso y preparación de datos relevantes, ofrece capacidad de cálculo y tratamiento mediante el uso de modelos ad-hoc, de investigación operativa, estadísticos, financieros y logísticos, además facilita el aprovechamiento de todos estos componentes mediante el uso de herramientas gráficas de visualización de información. (Pastor, 2002, p. 14)

Las principales características de estos son:

- a) Suelen introducirse después de haber implantado los sistemas transaccionales más relevantes de la empresa, ya que estos últimos constituyen su plataforma de información.
- b) La información que generan sirve de apoyo a los mandos intermedios y a la alta administración en el proceso de toma de decisiones.
- c) Suelen ser intensivos en cálculos y escasos en entradas y salidas de información.
- d) No suelen ahorrar mano de obra. Debido a ello, la justificación económica para el desarrollo de estos sistemas es difícil, ya que no se conocen los ingresos del proyecto de inversión.
- e) Suelen ser Sistemas de Información interactivos y amigables, con altos estándares de diseño gráfico y visual, ya que están dirigidos al usuario final.
- f) Apoyan la toma de decisiones que, por su misma naturaleza son repetitivos y de decisiones no estructuradas que no suelen repetirse.

Se han citado tres tipos de sistemas de información que se considera son los más necesarios para describir parte de la evolución de los sistemas informáticos que se han desarrollado como soporte a las diferentes tareas cotidianas que se realizan al interior de las empresas. Esto como preámbulo para incursionar en el ambiente de inteligencia de negocios.

2.5. Inteligencia de Negocios (BI)

La inteligencia de negocios apareció por primera vez en los años 60, su desarrollo ha sido tan importante que en la actualidad se está convirtiendo en un conjunto de herramientas imprescindibles para la supervivencia de las empresas.

“Se entiende por Business Intelligence al conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización.” (Curto, 2010, p.19).

La inteligencia de negocios es el conjunto de productos y servicios de software que permiten a los ejecutivos de todos los niveles tener acceso a la información de una manera rápida y sencilla, con la finalidad de analizarla y generar de esta forma conocimiento empresarial con el que se mejorará la toma de decisiones de los gerentes y directivos de la empresa.

En el mercado existen muchas herramientas y empresas que ofrecen soluciones en este campo, tanto en el sector del software propietario como en el del software libre, aunque en este último la oferta es un tanto limitada. Entre las principales y más potentes herramientas de Inteligencia de Negocios tenemos la suite Business Objects de SAP Company, OBI de Oracle, Microstrategy, COGNOS, Microsoft Business Intelligence, Petaho.

2.5.1. Necesidad de la Inteligencia de Negocios

- a) La decisión de cambiar al paradigma empresarial de las herramientas de inteligencia de negocios les da a las organizaciones la posibilidad de evidenciar que las inversiones realizadas en sistemas transaccionales como: ERP, CRM, no han sido infructuosas al extraer la información de dichos sistemas para

convertirla en conocimiento organizacional que servirá para elegir el rumbo de la empresa, alcanzando y manteniendo ventajas competitivas con relación a sus principales competidores, tomando las mejores decisiones en todos los niveles del proceso del negocio.

- b) Es evidente que la información de calidad y a tiempo para los ejecutivos de la empresa es imprescindible a la hora de tomar decisiones en cualquier momento y lugar. Las soluciones de Inteligencia de Negocios lo hacen posible transformando los reportes fríos, compuestos por cantidades de datos operacionales producto de las transacciones diarias de la empresa, en información que nos dicen cómo está la organización en cada una de sus áreas y de esta forma identificar los procesos a mejorar, hacer proyecciones a futuro, proponer nuevos productos o servicios.
- c) La información que utilizan los ejecutivos de una organización debe poder mostrar una visión de la empresa en términos cuantificables, la inteligencia de negocios lo hace a través de métricas perfectamente definidas y que guardan estrecha relación con los objetivos que la empresa persigue expresadas en metas claras.
- d) Muchas compañías no han pensado en implementar una estrategia de Inteligencia de Negocios que responda a las metas y objetivos organizacionales, esto otorga, a las compañías que ya lo están haciendo ventajas competitivas a nivel local e incluso nacional.

2.5.2. Importancia de la Inteligencia de Negocios

La importancia de negocios radica en los beneficios potenciales que podemos obtener al implementar soluciones en este campo, entre estos beneficios tenemos:

- a) Reducción de Costos.
- b) Información útil y verídica.

- c) Identificación de Riesgos y Oportunidades.
- d) Mayor Efectividad y Productividad.
- e) Facilita la toma de decisiones.
- f) Mejora el tiempo de respuesta.
- g) Ayuda a identificar nuevas oportunidades de negocio.

2.6. Componentes de Inteligencia de Negocios.

Los principales componentes de Inteligencia de Negocios se resumen a continuación y se representan en la figura 6:

- a) Fuentes de información para alimentar el almacén de datos empresarial (Data warehouse) o departamental (Data mart).
- b) Software con el que se realizarán los procesos de extracción, transformación y carga de datos al Data warehouse.
- c) Almacén de datos empresarial (Data warehouse) o departamental (Data mart) con el diccionario de datos (Metadata), que provee facilidad de acceso y administración.
- d) Motor OLAP que proporciona la capacidad de cálculo, consultas, funciones de planeamiento, pronósticos y análisis de escenarios en la gran cantidad de datos almacenados en el data warehouse.
- e) Software de minería de datos con el que se pueden elaborar predicciones, tendencias, sobre el comportamiento del proceso de negocios.
- f) Software de presentación de reportes con el que se presentarán los requerimientos que el usuario final ha solicitado a través de un explorador de internet o con soluciones a la medida diseñadas en algún lenguaje de programación.

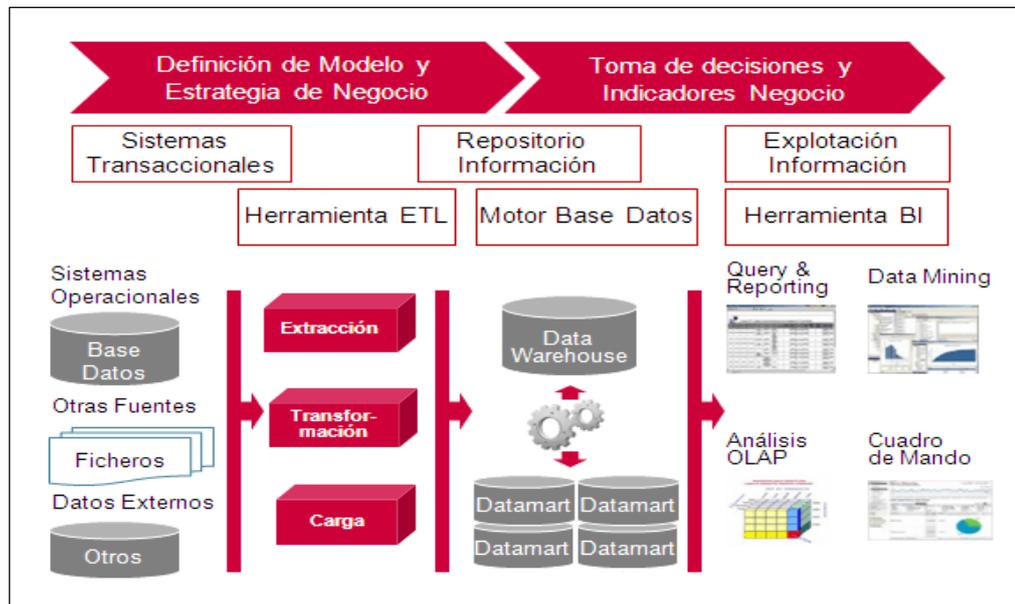


Figura 6: Arquitectura de Inteligencia de Negocios

Fuente: <http://rboacaranda.blogspot.com/2013/03/arquitectura-business-intelligence.html>

2.7. Herramientas principales de inteligencia de negocios

Una solución de inteligencia de negocios incluye una serie de herramientas y técnicas orientadas a gestionar, explotar, distribuir y estructurar información. Entre ellas destacan las siguientes:

2.7.1. Almacén de datos (Data Warehouse)

Un data warehouse o almacén de datos es una colección de datos de gran volumen que es extraída desde diferentes fuentes tanto internas como externas, bases de datos, archivos planos, hojas de cálculo, entre otras, con la finalidad de permitir la realización de consultas y análisis, el data warehouse es el núcleo de un proyecto de inteligencia de negocios. Sus principales características son:

- a) Es orientado a temas.
- b) Integrado.
- c) No es volátil.
- d) Variante en el tiempo.

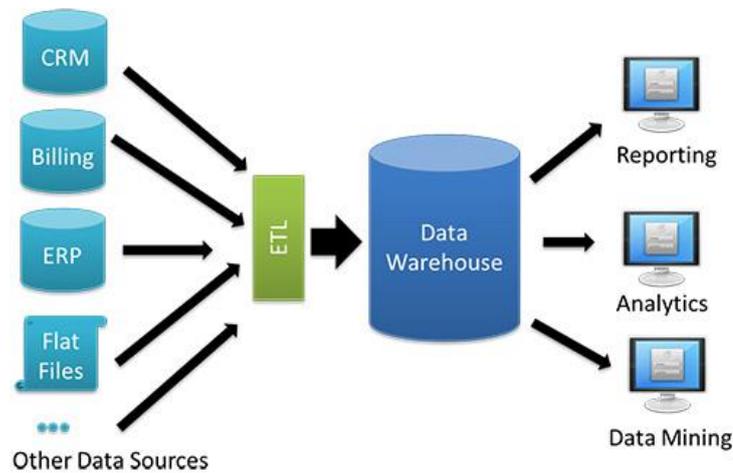


Figura 7: Data warehouse

Fuente: <http://savis.vn/solution/data-warehouse-solution/>

Es orientado a temas. En los sistemas transaccionales las bases de datos se optimizan para realizar tareas operacionales como agregar, eliminar, modificar, mover, copiar datos; en sistema comercial se registrará el ingreso y salida de mercadería de su inventario y se presentarán al final tablas con la cantidad de ingresos y egresos. Al contrario de un sistema dimensional en donde las bases de datos se optimizan para el análisis, orientando el diseño a los temas de interés de la empresa, que vienen a ser los procesos de negocio de los que se obtendrán indicadores con los que se podrán realizar tareas de gestión.

Integrado. Esta característica hace posible que los datos que formarán parte del almacén tengan el mismo formato para evitar inconsistencias, debido a que existen muchas fuentes de donde se los extraerá para llenar el data warehouse, y cada una de ellas con sus propios sistemas transaccionales en los que se trata de diferente forma a los datos entonces es necesario estandarizarlos. Un ejemplo podría ser una fecha que en algunos casos puede estar almacenada como: (1998-12-01) en otro (01/12/1998) o también como (1 de dic de 1998).

No volátil. En sistemas multidimensionales las bases de datos se optimizan para consultas complejas y de acuerdo a las necesidades del tomador de decisiones,

facilitando procesos de análisis que se hacen primordialmente sobre la base de datos históricas, por lo tanto no se admiten cambios solo se pueden hacer operaciones de lectura. Existen procesos de carga cada cierto periodo de tiempo pero la información no se altera, solo se incorporan los nuevos datos. En sistemas operacionales la información es susceptible a cambios y es precisamente una de las funciones necesarias en este tipo de ambientes.

De tiempo variante. La información guardada en un almacén de datos es precisa y requerida en un instante de tiempo determinado, se almacena como una larga serie de instantáneas que permanecen por largos intervalos de tiempo de entre 5 a 10 años durante el cual no se puede modificar, esto se debe a que los analistas requieren gran cantidad de datos para descubrir la información oculta. En contraste con las bases de datos tradicionales en donde los datos son requeridos en el momento del acceso, no se requiere realizar análisis por lo tanto no es necesaria la historia de los datos.

Para el caso del presente proyecto en el sistema operacional es importante el número de matriculados en el período actual, en el sistema multidimensional es importante el mismo número pero de los períodos precedentes también, con estos datos se podría planear el futuro en la oferta académica de la institución haciendo una proyección sobre el número de estudiantes que requieran matricularse.

2.7.2. Consulta y Reporte (Query & Report)

Las herramientas de consulta y reporte son muy importantes e imprescindibles para el desarrollo de aplicaciones de inteligencia de negocios, estos programas de consulta y reporte muestran los resultados del análisis realizado a la información almacenada en los data mart o data warehouse, de una manera sencilla, entendible y clara, este es el caso del tablero de mando que se muestra en la Figura 8, así como este podemos tener una variedad de reportes que dependerán del software de reportes que estemos usando y de la necesidad de los tomadores de decisiones que usarán la aplicación.

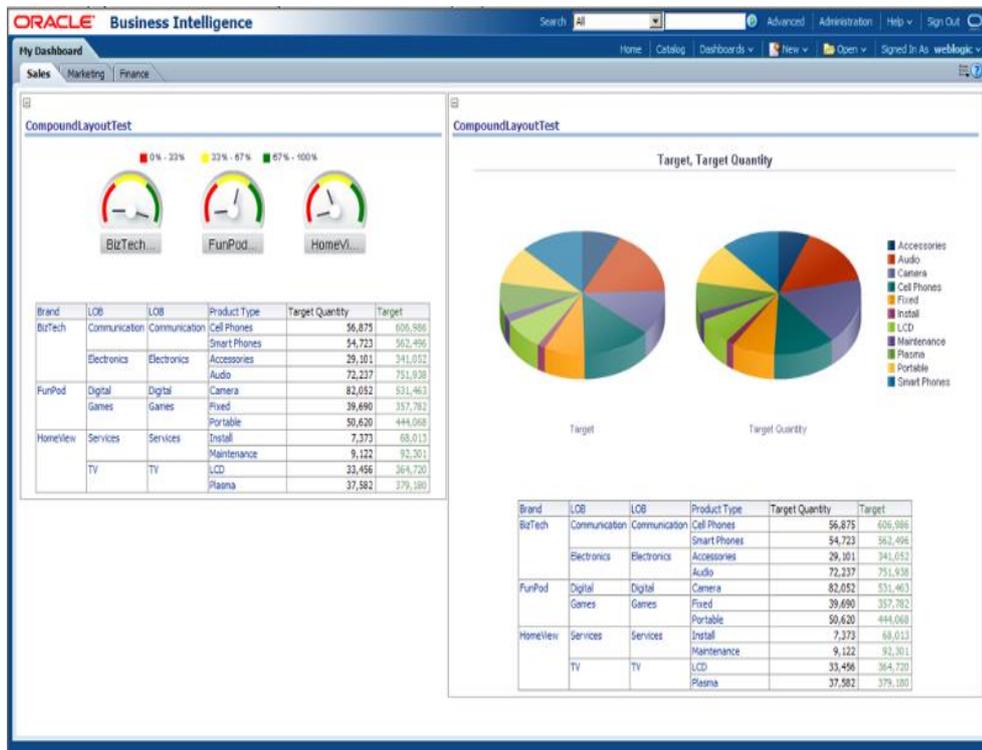


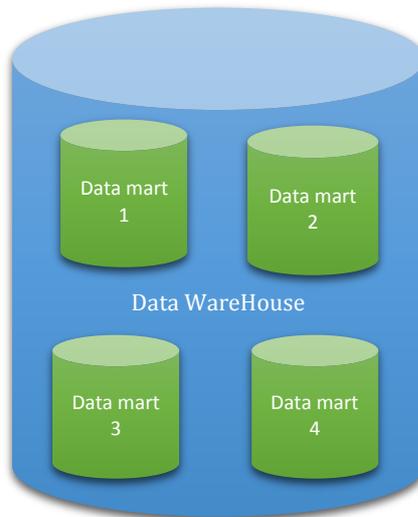
Figura 8: Tablero de mando (Dashboard)

Fuente: <http://http://www.praktics.com/soluciones/oracle-bi?lang=es>

2.7.3. Tienda de datos (Data mart)

Un data mart tiene las mismas características que un data warehouse, se puede entender como un subconjunto de este último, que se orienta a un área específica de la organización, como el departamento financiero, ventas, producción, por ejemplo, vendría a ser un almacén de datos más pequeño en volumen de almacenamiento de la información y al campo que abarca en la organización.

Se trata entonces de un repositorio que requiere menores costos y tiempo para construirse respecto de un data warehouse, es por esta razón que algunos autores lo llaman como tienda de datos, y en vista de que poseen las mismas características se explotan con el uso de herramientas OLAP para proporcionar todos los beneficios de un almacén de datos.



*Figura 9: Data mart como subconjunto de un Data warehouse
Fuente: Elaboración propia*

2.7.4. Herramientas OLAP

Procesamiento analítico en línea OLAP (On Line Analytical Processing), esta herramienta está compuesta por un motor OLAP que da respuesta a las solicitudes de información realizadas por los usuarios en forma eficiente, reflejando los requerimientos empresariales en lo que respecta al conocimiento sobre el estado y futura proyección de la organización.

A diferencia de los sistemas de procesamiento de transacciones en línea OLPT (On Line Transaction Processing), que se basa en un modelo relacional de base de datos, las herramientas OLAP se basan en el diseño multidimensional que da la posibilidad de analizar la información desde diferentes puntos de vista. Este análisis se lo puede realizar sobre los datos actuales como sobre los que se disponen en sus bases históricas, lo que proporciona una visión de cuál ha sido el desarrollo de la empresa en el tiempo.

La técnica OLAP posee entre sus capacidades más importantes, la de identificar patrones de comportamiento de los procesos del negocio y como estos han variado con el tiempo, permitiendo apoyar la toma decisiones en el análisis de diferentes variables,

haciendo posible dar respuesta a consultas complejas que reflejan las necesidades o requerimientos del negocio. El modelo multidimensional es representado por un cubo, en el que cada uno de los ejes constituye cada uno de los aspectos del análisis y cada elemento individual el resultado como se muestra en la figura 10.

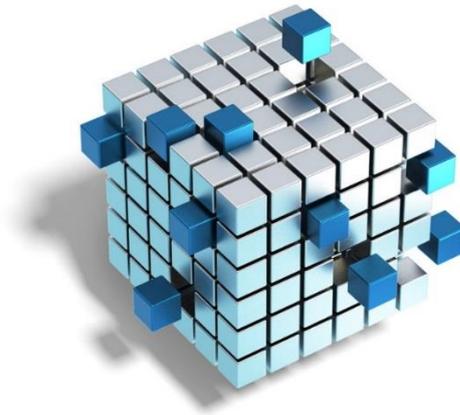


Figura 10: Cubo OLAP

Fuente: <http://www.dca.it>

2.7.5. Minería de datos (Data Mining)

Los datos que almacenan las empresas en sus bases de datos a simple vista no dicen nada, pero existe información oculta en esos grandes volúmenes de datos que crecen con el tiempo, para extraer esta información se utilizan técnicas de inteligencia artificial o de sistemas basados en el conocimiento, a todo este proceso se lo conoce como data mining o minería de datos. La minería de datos es un proceso conocido como descubrimiento de conocimiento en bases de datos o KDD (Knowledge Discovery in Data Bases) que consta de tres fases: preparación de los datos, extracción de información e interpretación de los resultados.

Normalmente la minería de datos se utiliza para detectar patrones de comportamiento de los datos con los que se puede hacer predicciones acerca de las tendencias de los clientes y de esta manera tomar decisiones que se adelanten al futuro. Otra de las áreas en donde se aplican esta clase herramienta es en la detección de fraudes en casinos, bancos, entre otros. En el presente proyecto solo se hace una descripción de esta

herramienta en vista de que no se realizará ningún proceso de minería de datos por que ese no es el tema del proyecto.



Figura 11: Data Mining (Minería de Datos)

Fuente: <http://www.stratebi.com>

2.8. Introducción al modelado multidimensional

El modelado multidimensional es una técnica para diseñar bases de datos que permitan estructurar la información de tal manera que se pueda realizar análisis desde diferentes puntos de vista, estas estructuras se basan en tablas de hechos y tablas de dimensiones. Con estas técnicas se pueden representar a las bases de datos como cubos de datos.

En las bases de datos multidimensionales, la información se almacena en forma dimensional y no relacional. Las dimensiones determinan la estructura de la información almacenada y definen adicionalmente caminos de consolidación. La información que se reúne se muestra como variables que, a la vez, se determinan por una o más dimensiones. De esta manera, y a partir de la intersección de esas dimensiones, se almacena el valor. (Reinosa, Maldonado, 2012, p.164).

Los cubos de datos de los que se habla en los modelos dimensionales tienen en sus aristas dimensiones, que serán elegidas de acuerdo al tipo de análisis que se requiera, en nuestro caso aplicando las técnicas del modelado dimensional al entorno universitario y específicamente a la matrícula de un estudiante, en un nivel de una

facultad, en una carrera específica, en un instante de un período de tiempo determinado, y como esto sucede con varios estudiantes se puede representar con la estructura que se muestra en la figura 12.

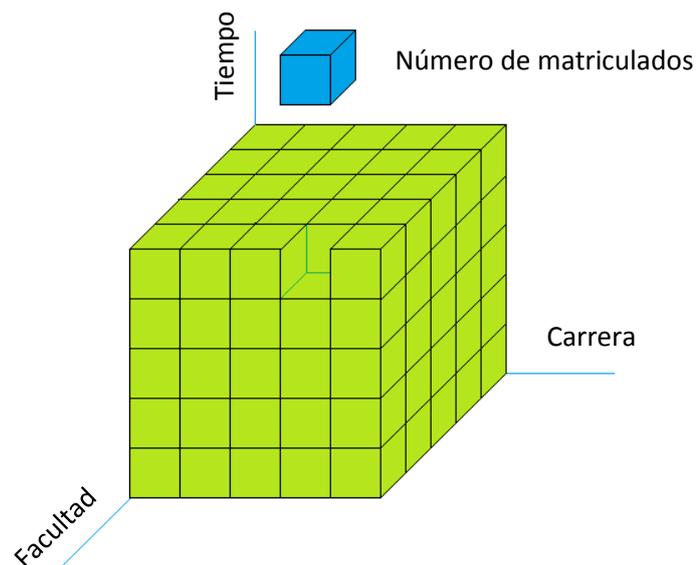


Figura 12: Ejemplo de cubo sobre notas de un estudiante en una materia.

Fuente: Elaboración propia

Esta estructura representada en la figura 12 nos ofrece la capacidad de realizar un análisis profundo haciendo variaciones en la combinación de las dimensiones y sus jerarquías asociadas, obteniendo como un posible resultado el número de matriculados producto de la intersección de las dimensiones.

La diferencia entre los modelos dimensional y transaccional radica en su diseño, el primero se basa en la des normalización de las tablas permitiendo realizar tareas de análisis, y el segundo tiene una alta normalización de tablas lo que facilita las tareas operacionales. Según Kimball, para que se pueda aprovechar los beneficios de un data warehouse debe ser construido desde el punto de vista dimensional en lugar del complejo enfoque transaccional.

La estructura de un modelo dimensional se basa en un conjunto de tablas, que contienen eventos en unas y características de estos en otras, se las conoce como tablas de hechos en el primer caso y dimensiones en el segundo.

2.8.1. Hechos

Se definen como hechos a ciertos sucesos o eventos que presentan en una organización, como por ejemplo personas que se inscriben y se matriculan, estudiantes que rinden exámenes, profesores que registran las notas, estudiantes que se gradúan. En las tablas de hechos no se almacenan valores textuales sino numéricos que son los datos con los que se pueden analizar, con hechos textuales sería casi imposible realizar análisis, por esto se dice que los hechos más útiles son los numéricos y aditivos, como el total de ventas de una factura, la nota obtenida por un estudiante en una asignatura.

Las tablas de hechos ocupan una gran parte de la base de datos dimensional y se caracterizan por tener un número reducido de columnas pero una amplísima cantidad de filas. Este espacio que ocupan las tablas de hechos está directamente relacionado con el nivel de detalle que se requiera, a mayor grado de detalle mayor será el tamaño necesario para guardar la información en la tabla. Según Kimball, todos los hechos que posiblemente sean parte del diseño deben estar en concordancia con la granularidad que se ha definido. (Trujillo, 2013) *La granularidad está relacionada con la lista de dimensiones que definen el alcance y el grado de detalle con los que se almacenan las medidas en la tabla de hechos* (p.35).

Para determinar qué información contendrá la tabla de hechos se debe recurrir a la granularidad y regirse estrictamente a esta. Volviendo al caso del presente proyecto se puede obtener los resultados de estudiantes aprobados, retirados, perdidos, entre otros. Pudiendo analizar esto desde diferentes puntos de vista que serán determinados por las dimensiones que se tomen en cuenta en el modelo. Se debe tener en cuenta que la información almacenada en la tabla de hechos no cambia en el futuro ya que es el resultado de eventos pasados.

Tabla de hechos ventas	
PK	<u>FK Id Clientes</u>
PK	<u>FK Id Fecha</u>
PK	<u>FK Id Productos</u>
PK	<u>FK Id Almnacen</u>
	Cantidad vendida Ventas en dólares

Figura 13: Tabla de hechos ventas

Fuente: Elaboración propia

Las tablas de hechos contienen las claves primarias de las tablas de dimensiones que son parte del modelo, pasando a ser claves foráneas y en conjunto todas ellas se convierten en la clave primaria de la tabla de hechos, como se puede observar en la figura 15. Las tablas de dimensión tienen una relación de muchos a muchos con la tabla de hechos y esta una relación de uno a muchos con cada dimensión en particular. Sin embargo en ocasiones podemos encontrar relaciones de muchos a muchos entre los hechos y las dimensiones.

2.8.2. Dimensiones

En el modelo multidimensional las dimensiones son tablas que contienen descripciones textuales de las características de los procesos de negocio que son objeto de análisis, hacen referencia a líneas o áreas específicas de la empresa, estas representan el contexto de análisis de los hechos. En el ámbito universitario, y sobre el análisis de comportamiento académico de los alumnos, se podrían considerar a las dimensiones: dependencia, estudiante, ingreso, entre otras, pudiendo encontrar características como el nombre de la carrera, el nombre de la escuela y el de la facultad a la que pertenece el estudiante.

Los nombres de los atributos en las de dimensión deben ser palabras completas con las que se pueda tener una idea clara de la característica que describe, por lo general deberían tener como mínimo de 10 a 15 caracteres y como máximo entre 30 y 50.

Según (Kimball, 2002) “Los mejores atributos son textuales y discretos. Los atributos deben consistir en palabras reales en lugar de abreviaturas crípticas.” (p.20).

Las tablas de dimensiones se caracterizan por tener muchas columnas o atributos. Se pueden encontrar diseños con tablas de entre 50 y 100 atributos. Estos atributos describen las filas de la tabla de dimensiones como se muestra en la figura 16. Según (Trujillo, 2013) “Una dimensión se compone de una serie de atributos organizados jerárquicamente. Estos atributos permiten analizar las medidas de los hechos a diferente nivel de detalle según se agreguen o desagreguen los datos” (p.30).

Dimensión Productos	
PK	<u>Id Productos</u>
	Descripción Código SKU Clase de empaque Nombre de la categoría Nombre de la subcategoría Fecha de elaboración Fecha de vencimiento Número de lote Contenido neto Contenido de grasa Contenido de sal Contenido de azúcar y más

Figura 14: Tabla de la dimensión producto

Fuente: Elaboración propia

Los atributos de las dimensiones se pueden ordenar lógicamente de forma jerárquica, de tal forma que en el primer lugar del orden está el menor nivel de detalle y conforme se avanza en este ordenamiento este irá aumentando, realizando tareas de agrupación en donde el criterio para hacerlo lo dan los atributos de la dimensión asociada.

A estos ordenamientos lógicos se los conoce como jerarquías y se deben especificar el momento del diseño de las dimensiones y de acuerdo a los requerimientos del negocio. Se puede establecer más de una jerarquía para una misma dimensión, pero siempre

podrá identificarse a una principal que se caracteriza por tener el mayor número de elementos.

Se puede asegurar la calidad del almacén de datos asegurando la calidad de sus dimensiones ya que como lo afirma (Kimball, 2002) *“El poder del almacén de datos es directamente proporcional a la calidad y profundidad de los atributos de sus dimensiones.”* (p.19-20). Por lo tanto mientras mayor nivel de detalle de las dimensiones mejor es el diseño del modelo multidimensional.

2.8.3. Medidas

Las medidas o métricas son indicadores cuantificables que dan una idea sobre el estado del proceso de negocio que es analizado, se usan campos numéricos para hacer referencia a las medidas de un proceso de negocio. Al respecto (Curto, 2010) afirma *“Aquellos conceptos cuantificables que permiten medir nuestro proceso de negocio”* (p.34).

Las medidas pueden ser atómicas, aquellas que están contenidas en el hecho mismo o derivadas aquellas que necesitan de hacer algún cálculo para obtenerlas, son el resultado de la aplicación una fórmula, un ejemplo de medida derivada podría ser el valor total en ventas que resultaría de multiplicar la cantidad vendida por el precio.

Además las medidas pueden tener diferentes niveles de granularidad y ser definidas en diferentes hechos. Por ejemplo podemos analizar la cantidad vendida del producto A en Imbabura, y se puede ser necesario ir más abajo a las ciudades, los cantones, las parroquias, los barrios, alcanzando en este caso 5 niveles de granularidad.

Las medidas son aditivas cuando se les puede aplicar la operación de suma sobre todas sus dimensiones, semiaditivas en el caso de que esta operación se puede aplicar sobre una parte de sus dimensiones y no aditiva si no se puede realizar esta operación sobre ninguna de sus dimensiones.

2.9. Operaciones básicas OLAP

La base fundamental de la inteligencia de negocios no son las consultas que se puedan realizar, sino las operaciones que se pueden realizar sobre esas consultas, operaciones que viabilizan los diferentes tipos de análisis que se puede hacer sobre la información cargada, en los almacenes de datos departamentales u organizacionales. Estas operaciones se describen a continuación.

2.9.1. Drill Down y Roll up

Estas son las operaciones OLAP básicas con las que se responde de acuerdo al grado de detalle con el que se quiera elaborar un reporte, en el caso de drill down, o profundizar en español, podemos pasar de un mayor grado de detalle a uno menor, y lo contrario pasar del nivel más bajo de detalle al más alto utilizaremos roll up, arrollar en español, siempre teniendo en cuenta la granularidad que se ha definido para cada proyecto en particular. Además hay que recordar que la granularidad es directamente proporcional al grado de detalle con el que se requiere analizar la información.

2.9.2. Slice y Dice

Las operaciones Slice (trocear) y dice (dado) se refieren al proceso de manipular las dimensiones para tener una vista particular que puede abarcar a toda una área del cubo, en el caso de slice, o a un sector específico para dice, se agregan restricciones para obtener estas vistas. Estas operaciones separan los datos y realizan múltiples combinaciones con ellos, esto nos permite realizar diferentes tipos de análisis con diferentes perspectivas, finalidad que se logra al agrupar o reagrupar los datos del cubo.

2.9.3. Drill across

Esta operación permite navegar de un hecho a otro mediante el uso de dimensiones comunes. Gracias a esta operación, se pueden consultar medidas de varios hechos en el mismo cubo.

2.9.4. Pivoting

El pivoting, o pivoteo en español, es la operación que proporciona la capacidad de reorientar la vista multidimensional de los datos al intercambiar la distribución de filas por columnas. Se podría considerar también el intercambio entre medidas y hechos a lo que ciertos autores también lo llaman como pivoteo.

2.10. Arquitectura de los modelos analíticos

En apartados anteriores se describen brevemente las diferencias entre los sistemas OLAP y OLTP. Estas diferencias hacen notar que las técnicas clásicas de diseño y el modelado del ciclo de vida utilizado para desarrollar un sistema operacional no resulten apropiados para el entorno analítico. A continuación se explican las técnicas y metodologías principales para garantizar el éxito de un proyecto de Data Warehousing.

Un modelo es un conjunto de conceptos, reglas y convenciones que permiten describir y manipular los datos que queremos almacenar en una base de datos. En ellas se pueden usar distintos modelos de datos para describir la información con que operan.

Un Modelo de Datos permite describir:

- a) La estructura de datos, el tipo de los datos que hay en la base y la forma en que se relacionan.

- b) Las restricciones de integridad son un conjunto de condiciones que deben cumplir los datos para reflejar correctamente la realidad deseada.

En entornos de data warehouse, los modelos más extendidos para representar los datos se basan en el hecho de que no todas las entidades tienen igual número de ocurrencias ni presentan igual curva de crecimiento en el volumen de dichos datos.

Los data warehouse están más orientados a la consulta para apoyar la toma de decisiones, que los sistemas operacionales que están más dedicados a trabajar con las actividades operativas diarias de la empresa. Estas diferentes visiones del problema tienen una gran importancia en la manera de modelar, en un sistema transaccional los datos se encuentran normalizados para evitar la redundancia de información y las tablas contendrán gran cantidad de filas, no existe la posibilidad de análisis multidimensional, al contrario un data warehouse se basa en estructuras de datos que facilitan este tipo de análisis. A continuación se expondrán las características de 3 modelos: El modelo en estrella, el modelo copo de nieve y el 3FN.

2.10.1. Modelo en estrella

“El esquema Star o Estrella es una técnica de modelado de datos que se utiliza para hacer corresponder un modelo multidimensional sobre una base de datos relacional. Se lo denomina de esta manera porque su forma se parece a una estrella” (Reinosa, 2012, p.179).

Es el más sencillo de todos los modelos de almacenamiento de datos, En el centro se encuentra la tabla de hechos y a su alrededor las tablas de dimensiones que se encuentran relacionadas con la primera a través de claves foráneas, como se muestra en la figura 15. Se puede implementar un modelo multidimensional usando la clásica base de datos relacional, es simple de manejar y su velocidad es excelente.

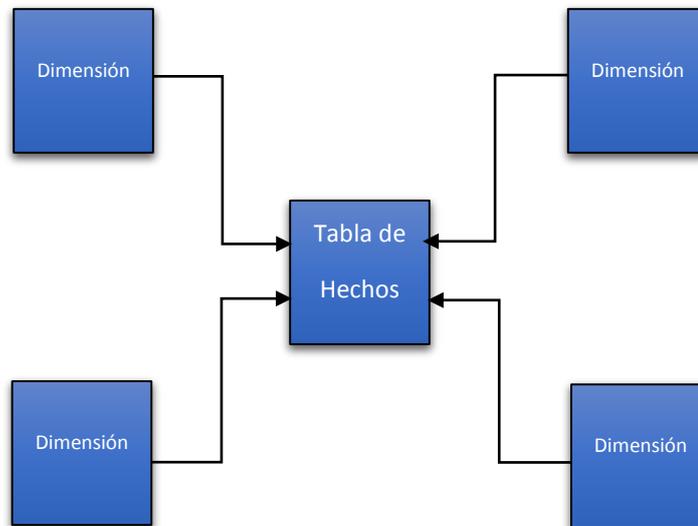


Figura 15: Esquema en estrella de un Data warehouse

Fuente: Elaboración propia

El esquema en estrella consiste en estructurar la información en procesos, vistas y métricas a modo de estrella. En la tabla de hechos encontramos los atributos destinados al hecho que constituye el proceso de negocio a medir, es decir, sus métricas.

Mientras, en las tablas de dimensión, los atributos se destinan a elementos de nivel (que representan los distintos niveles de las jerarquías de dimensión) y a atributos de dimensión (encargados de la descripción de estos elementos de nivel). En el esquema en estrella la tabla de hechos es la única tabla que tiene múltiples joins que la conectan con otras tablas. El resto de tablas del esquema (tablas de dimensión) únicamente hacen join con esta tabla de hechos. Las tablas de dimensión se encuentran además totalmente des normalizadas, es decir, toda la información referente a una dimensión se almacena en la misma tabla. La figura 6 muestra un ejemplo de diagrama utilizando el modelo en estrella.

2.10.2. Modelo copo de nieve

El esquema copo de nieve es una representación derivada del esquema en estrella, en el que las tablas de dimensión se normalizan en múltiples tablas. Por esta razón, la tabla de hechos deja de ser la única tabla del esquema que se relaciona con otras tablas,

y aparecen nuevas uniones entre tablas gracias a que las dimensiones de análisis se representan ahora en tablas de dimensión normalizadas. En la estructura dimensional normalizada, la tabla que representa el nivel base de la dimensión es la que hace la unión directamente con la tabla de hechos.

La diferencia entre ambos esquemas, estrella y copo de nieve, reside entonces en la estructura de las tablas de dimensión. Para conseguir un esquema en copo de nieve se ha de tomar un esquema en estrella y conservar la tabla de hechos, centrándose únicamente en el modelado de las tablas de dimensión, que si bien en el esquema en estrella se encontraban totalmente des normalizadas, ahora se dividen en sub tablas tras un proceso de normalización.

Es posible distinguir dos tipos de esquemas en copo de nieve, uno completo, en el que todas las tablas de dimensión en el esquema en estrella aparecen normalizadas y otro parcial, sólo se lleva a cabo la normalización de algunas de ellas. En la figura 15 se observa el modelo en copo de nieve.

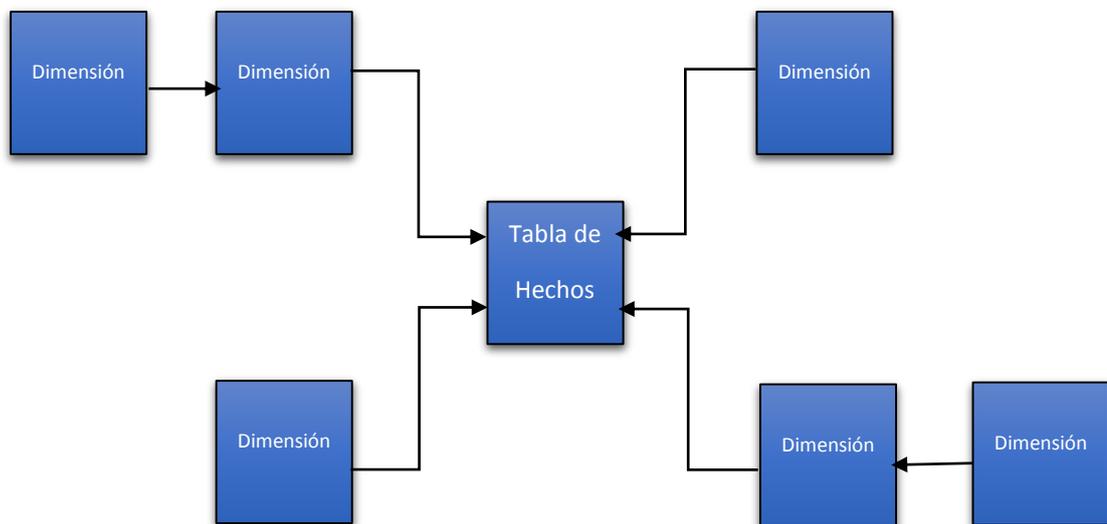


Figura 16: Esquema copo de nieve de un Data warehouse

Fuente: Elaboración propia

2.10.3. Modelo 3FN.

Aunque en los entornos de BI los esquemas de estrella son los más empleados, también se puede utilizar la tercera forma normal para su utilización en sistemas de almacenamiento de datos. El modelado de la Tercera Forma Normal (3FN) es una técnica clásica de las bases de datos relacionales que minimizan la redundancia de datos a través de la normalización de los datos.

Cuando se compara con un esquema en estrella, un esquema 3FN tiene normalmente un mayor número de tablas debido al proceso de normalización. Los esquemas 3FN se utilizan en los almacenes de datos grandes, especialmente en entornos con importantes requisitos de carga de datos que se utilizan para alimentar DM y ejecutar consultas de larga ejecución. En la figura 11 se puede ver un diseño de un esquema 3FN. Las principales ventajas de los esquemas 3FN son las siguientes:

- a) Proporcionan un diseño de esquema neutral e independiente de cualquier aplicación.
- b) Puede requerir menos transformación de datos que otros esquemas como los esquemas en estrella.

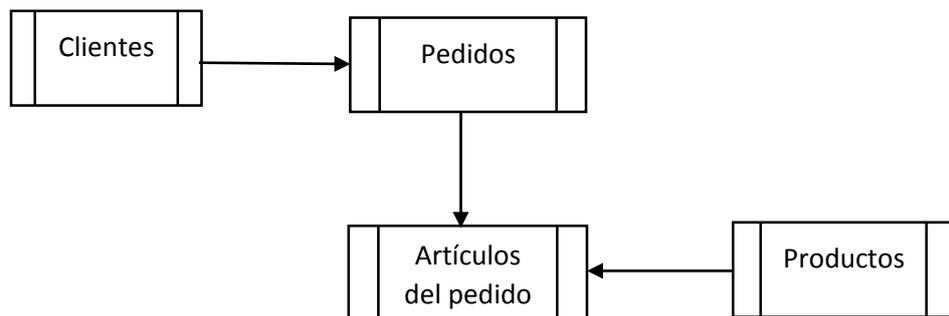


Figura 17: Esquema 3FN

Fuente: Elaboración propia

2.11. Arquitectura de un Data warehouse

Es necesario determinar cuál es la arquitectura en la que basaremos la construcción de un almacén de datos, de esta manera se puede comprender como se relacionan cada uno de los elementos. En la figura 18 se pueden observar las tres principales capas de la arquitectura de un data warehouse, es decir, el nivel de organización, el nivel de directorio y por último el nivel de gestión de procesos desde abajo hacia arriba.

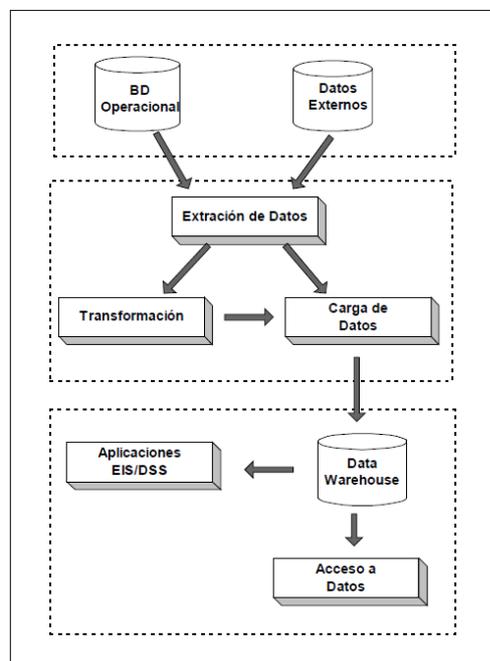


Figura 18: Estructura básica de un data warehouse

Fuente: <http://www2.itba.edu.ar/archivos/secciones/nader-tesisdemagister.pdf/>

La forma en la cual se estructura el almacenamiento de datos en el data warehouse genera una clasificación respecto a la forma de implementar su arquitectura. La estructura adoptada se debe realizar de la manera que mejor satisfaga las necesidades empresariales, siendo entonces dicha elección un factor clave en su efectividad.

Una vez establecidos los conceptos básicos sobre los diferentes niveles existentes que conforman los data warehouse, nos encontramos con varios modelos de arquitectura, apareciendo tres como los más habituales según el manual Data Warehousing Guide. (Oracle, 2013)

- a) Arquitectura data warehouse básica.
- b) Arquitectura data warehouse con área de organización.
- c) Arquitectura data warehouse con área de organización y data marts.

2.11.1. Arquitectura data warehouse básica

Como se puede observar en la figura 19, los usuarios finales pueden acceder directamente a los datos que se encuentran almacenados en bases de datos de sistemas operacionales o transaccionales a través del data warehouse.

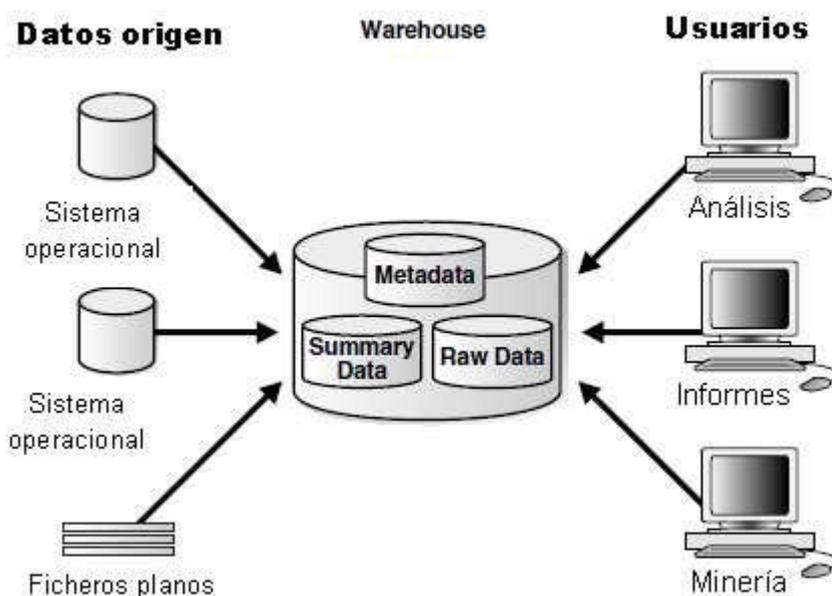


Figura 19: Arquitectura Data warehouse básica

Fuente:

https://docs.oracle.com/cd/E11882_01/server.112/e25554/concept.htm#DWHSG8070

Se pueden ver representados los metadatos, los datos en bruto o sea sin procesar que se originan en sistemas operacionales, además se puede observar también los datos de resumen que son muy importantes ya que ahorran tiempo de procesamiento, esto debido a que son el producto de cálculos complejos que se realizan previamente. Es lógico pensar que un resumen de datos es más fácil de manejar en contraste al gran volumen que se necesita para realizar análisis.

2.11.2. Arquitectura data warehouse con área de organización

Este modelo es similar al anterior con la única diferencia de que se agrega un área intermedia llamada área de organización, este nuevo elemento de esta arquitectura ha sido pensado para realizar una refinación de los datos antes de que pasen al data warehouse. De esta manera se mejora la administración del data warehouse además de que se facilita el proceso de resumir los datos.

El área de organización es el espacio en el que se colocan los datos en tránsito, por lo general procedentes de la capa de procesamiento ETL previa. En su forma más simple, son datos que han sido resumidos, de manera que la información detallada se ha sintetizado hasta el menor detalle. (Rodríguez, 2010, p.44)

Los administradores de bases de datos tienen la responsabilidad de decidir el modelo físico con el que se almacenarán los datos, esta arquitectura se observa en la figura 20.

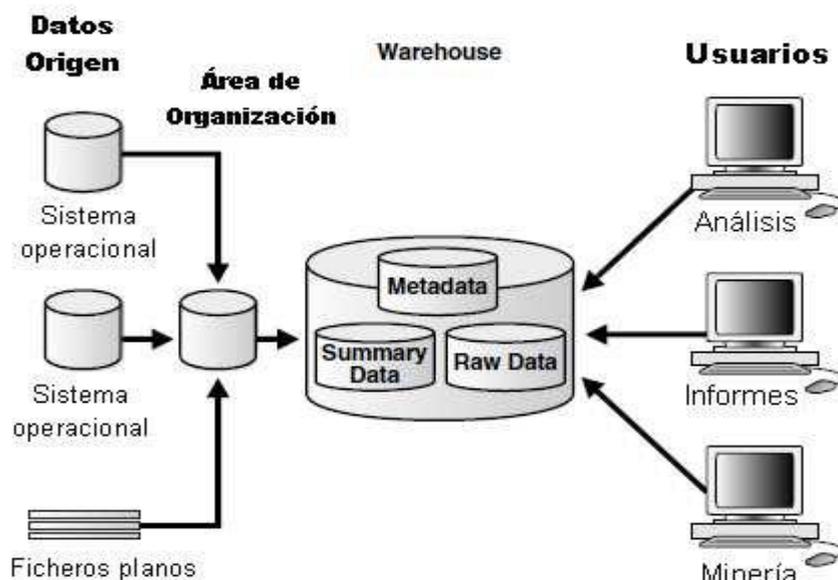


Figura 20: Arquitectura Data warehouse con área de organización

Fuente:

https://docs.oracle.com/cd/E11882_01/server.112/e25554/concept.htm#DWHSG8070

2.11.3. Arquitectura Data Warehouse con Área de Organización y Data Marts

A esta arquitectura se le ha incorporado otro elemento, los data marts, de forma que se implementa el almacén empresarial a partir de los departamentales. Se dispone de una base de datos, generalmente de detalle o de información común a todos los usuarios y además cada departamento, área o línea de negocio dispone de su propia base de datos.

Esta arquitectura se puede ajustar a las necesidades de la empresa según sean las diferentes áreas de negocio que existan en la organización, cada datamart se encuentra relacionado con una determinada línea de negocio como se observa en la figura 21.

“Los datos existentes en este contexto pueden ser agrupados, explorados y propagados de múltiples formas para que diversos grupos de usuarios realicen la explotación de los mismos de la forma más conveniente según sus necesidades” (Rodríguez, 2010, p.46).

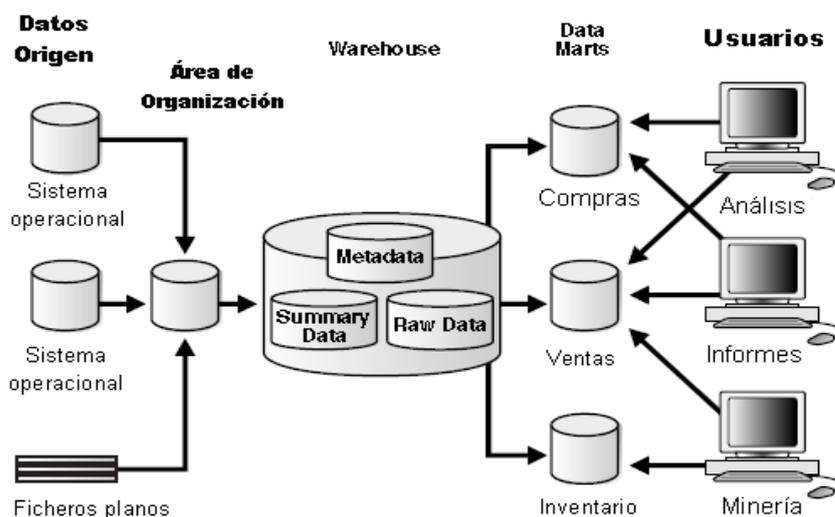


Figura 21: Arquitectura Data warehouse con área de organización y Data marts

Fuente:

https://docs.oracle.com/cd/E11882_01/server.112/e25554/concept.htm#DWHSG8070

Estos modelos se han optimizado para realizar consultas complejas, en donde la carga de los datos hacia el almacén se la realiza ejecutando procesos por lotes con una

frecuencia predeterminada. En la figura 21, se observa como los usuarios usan diferentes herramientas OLAP para acceder a la información, proporcionando una visión multidimensional de esta.

2.12. Oracle Business Intelligence

Oracle cuenta con una de las bases de datos más potentes del mercado y que además cuenta con una amplia gama de productos de software para su manejo, lo que hace de esta empresa una de las más competitivas en el mercado a nivel mundial.

En este conjunto de herramientas podemos encontrar la suite de inteligencia de negocios, que cuenta con tres clases de productos que dependiendo del tamaño de empresa y sus necesidades se puede decidir entre Oracle Business Intelligence Standard Edition (OBISE), Oracle Business Intelligence Standard Edition One (OBISE1) y Oracle Business Intelligence Enterprise Edition (OBIEE).

2.12.1. Oracle Business Intelligence Standard Edition (OBISE)

OBISE es una suite de herramientas de software que dispone de una serie de productos con los que se puede emprender el desarrollo de aplicaciones de inteligencia de negocios, es el resultado del constante progreso de la compañía Oracle en el área de software de soporte a las operaciones empresariales y en especial de las clásicas herramientas de presentación de reportes como Discoverer.

De todas las opciones esta es la más económica si tomamos en cuenta que el programa de licenciamiento por usuario de la suite, no cuenta con la base de datos. El acceso a los cuadros de mando y los informes vía web se lo realiza a través de IAS. En ambientes en donde se usan productos Oracle este conjunto de herramientas de software se acopla muy bien sin tener que recurrir a grandes inversiones económicas a la hora de emprender el desarrollo de aplicaciones de inteligencia de negocios. A continuación hacemos una breve referencia de cada una de los productos que vienen en la suite.

- *Oracle BI Discoverer*: Es un conjunto de programas que facilitan la creación de consultas ad hoc, pueden tener acceso a los datos a bases de datos relacionales como multidimensionales, se integra fácilmente con Oracle Web Center y Weblogic server mejorando la escalabilidad y rendimiento.
- *Oracle BI Spreadsheet Add-in*: Esta herramienta combina la potencia de Oracle OLAP con la funcionalidad de las hojas de cálculo en Excel facilitando el acceso a los datos desde una hoja electrónica común y corriente.
- *Oracle BI Beans*: Herramienta de desarrollo para construir aplicaciones de inteligencia de negocios a la medida, aprovechando la funcionalidad OLAP de la base de datos. Se integra perfectamente con JDeveloper formando un entorno de desarrollo más productivo en donde podemos personalizar nuestras aplicaciones.
- *Oracle Reports Services*: Es el módulo de presentación de informes de Oracle Fusion Middleware, con el que se publican y distribuyen informes de alta calidad y con un alto nivel de detalle tanto en entornos basados en la web y otros.

2.12.2. Oracle Business Intelligence Standard Edition One (BISE1)

Esta es la suite de inteligencia de negocios completa y accesible por su precio atractivo y por sus características diseñadas para la pequeña y mediana empresa, incluye todo lo necesario para desarrollar aplicaciones de Inteligencia de Negocios en poco tiempo. Su plataforma se basa exactamente en la misma tecnología que la suite Oracle Business Intelligence Enterprise Edition, por lo que es altamente escalable según crezcan las necesidades de la organización.

Ofrece los componentes básicos para gestionar y construir una solución de BI departamental, además de funcionalidades de inteligencia de negocios globales cuadros de mando interactivos, informes con formato, análisis ad hoc, de integración

de datos y administración del servidor, todo entregado en una moderna arquitectura web común. (Oracle, 2010)

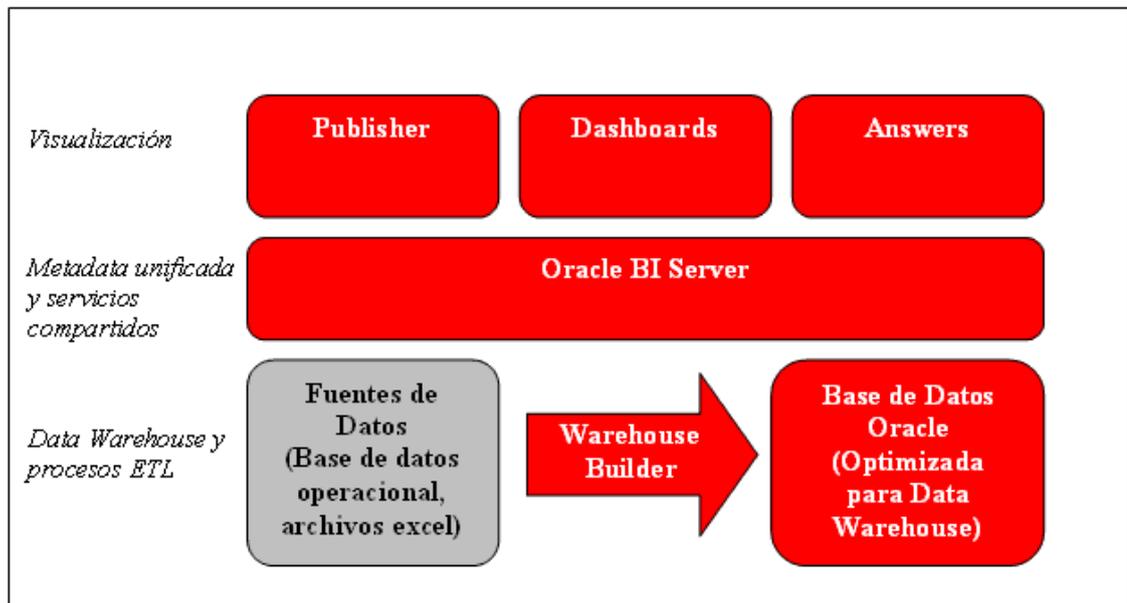


Figura 22: Arquitectura de OBISE One

Fuente: <http://www.neuronet.cl/productos/oracle-business-intelligence/>

Este conjunto de herramientas de inteligencia de negocios ha sido creado con la intención de que sea más fácil tomar mejores decisiones. La licencia no permite utilizar más de dos CPU's del servidor y sólo admite otra fuente de datos directa aparte de la base de datos que está incluida en el paquete. El licenciamiento es por usuario nominal, y se pueden licenciar entre 5 y 50 usuarios.

El paquete de instalación viene con todos elementos necesarios para el desarrollo de soluciones de BI de la versión Standard, además se incluye la herramienta de ETL Oracle Warehouse Builder en la versión básica el servidor de inteligencia de negocios y el servidor de base de datos por supuesto. Todo este software se instala en un solo sencillo proceso.

El servidor de BI permite acceder por web a la herramienta de informes analíticos y de cuadros de mando tanto para diseño como para explotación. Provee un entorno de

desarrollo muy completo y fácil de utilizar. Las herramientas que incluye la suite son las siguientes (Fernández, 2009).

- *Oracle BI Server*: Es el núcleo de la plataforma de BI ya que posee un potente motor de cálculo esencial para los procesos de análisis, el acceso y control de la información se la realiza desde el servidor.
- *Oracle BI Server Administrator*: Software de inteligencia de negocios con el que se puede trabajar en un proyecto completo desde su diseño de la interfaz dividido en tres secciones una para la fuente de datos y modelo físico, otra para el modelo multidimensional y por último otra para el diseño de reportes. Dispone de la facilidad de creación de los metadatos y niveles de abstracción.
- *Oracle BI Answers*: Elemento que permite a los usuarios finales crear fácilmente sus propios reportes con solo arrastrar y soltar además posibilita navegar entre los datos con capacidades de drill up y drill down.
- *Oracle BI Interactive Dashboards*: Cuadros de mando interactivos para entornos de análisis.
- *Oracle BI Publisher*: Software de creación de reportes empresariales y de distribución de informes con gran nivel de detalle, se integra muy fácilmente en ambientes donde ya se tiene instalado software de Oracle. Proporciona la facilidad de crear reportes de una manera rápida ahorrando costos a las empresas.
- *Oracle Data Base Standard Edition One*: Proporciona el servidor de la base de datos oracle con características que le permiten desempeñarse de mejor forma en ambientes BI proporcionando el rendimiento, escalabilidad y manejabilidad.

- *Oracle Warehouse Builder*: Herramienta completa de diseño de procesos de extracción, transformación y carga (ETL), con la facilidad de conectarse a diferentes fuentes de datos.

2.12.3. Oracle Business Intelligence Enterprise Edition (BIEE)

Esta es la suite está diseñada para la gran empresa y para trabajar con múltiples orígenes de datos. Está compuesta por todas las herramientas de Oracle BISE1, exceptuando la base de datos, más algunas complementarias aplicables al mismo entorno, y que amplían la funcionalidad de la misma con utilidades de CPM, monitorización y alertas o para poder utilizar las funciones analíticas en modo desconectado.

Si la suite se licencia por usuarios el mínimo es 50 y también se puede licenciar por procesador. No está limitada en cuanto a número de orígenes de datos, usuarios o CPU's más que por el presupuesto disponible. Como contiene todas las aplicaciones de la versión BISE1, especialmente el servidor de BI, el paso de la versión más modesta a la Enterprise es inmediato. Las aplicaciones que la componen son:

- *Oracle BI Delivers*: Software para crear alertas para monitorear las actividades del negocio.
- *Oracle BI Disconnected Analytics*: Funcionalidad analítica completa en modo desconectado para profesionales móviles. Utiliza técnicas de replicación.
- *Oracle BI Briefing Books*: Instantáneas de páginas de cuadros de mando para visualizar y compartir en modo desconectado.
- *Hyperion Interactive Reporting*: Reporting ad-hoc intuitivo e interactivo.

- *Hyperion SQR Production Reporting*: Generación de informes con formatos de presentación de alta calidad, preparado para trabajar con grandes volúmenes.
- *Hyperion Financial Reporting*: Reportes financieros y de gestión con formato en calidad de libro.
- *Hyperion Web Analysis*: Análisis, presentación y reporting OLAP basado en web.

Para el presente trabajo se ha escogido Oracle Business Intelligence Standard Edition One, esta decisión se la hizo en base a que la Universidad Técnica del Norte cuenta con las licencias de este software en un número de 5, una para cada usuario. A continuación se describen cada uno de los productos con los que cuenta esta versión del software para inteligencia de negocios de Oracle.

2.12.4. Oracle Business Intelligence Server.

Acceso centralizado a los datos y provee un motor de cálculo que se apoya en un modelo lógico de información empresarial. Oracle BI Server es un servidor de consultas, informes y análisis de alto impacto que presta servicios a los demás componentes de la Suite de Inteligencia Business como: respuestas, cuadros de mando, la minería de datos, informes y aplicaciones analíticas. El Oracle BI Server utiliza ODBC 2.0 y JDBC, los servicios del servidor de BI tienen dos funciones principales. En primer lugar, se compilan solicitudes de consulta entrantes en un código ejecutable.

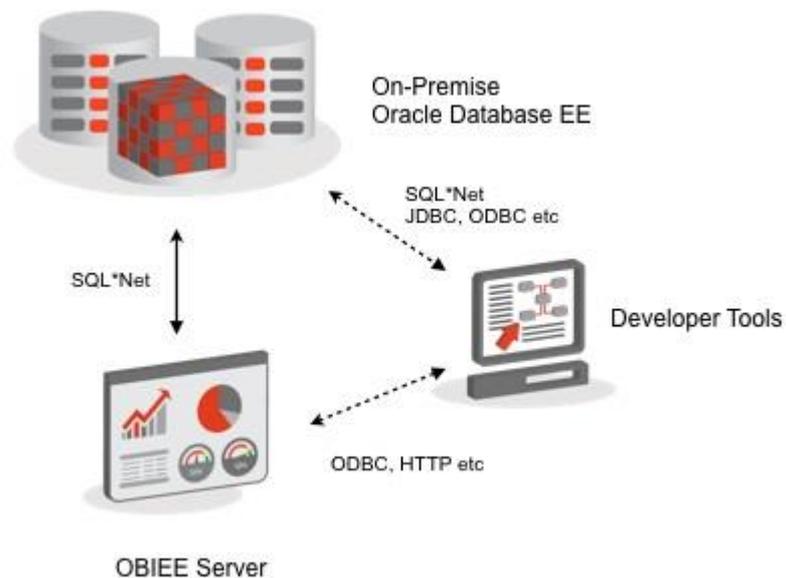


Figura 23: Oracle BI Server

Fuente: <http://www.rittmanmead.com/2013/09/using-non-oracle-and-cloud-hosted-databases-with-obiee-11g/>

Los clientes del servidor de Inteligencia de negocios trabajan con un conjunto de datos lógicos independientes de las fuentes de datos, y los entregan a un servidor BI. El servidor convierte los datos a una combinación de SQL física y los envía a las bases de datos finales. Las funciones de administración de servidores incluyen servicios como vigilancia, estadísticas de tala, cancelación, reunión y gestión de consultas, entre otros. La administración del servidor se gestiona internamente con el uso de protocolos estándar como LDAP (Lightweight Directory Access Protocol) protocolo compacto de acceso a directorios.

La compilación de la consulta tiene cinco fases:

- a) Análisis
- b) Generación petición lógica
- c) Navegación
- d) Volver a escribir
- e) La generación de código

2.12.5. Oracle Business Intelligence Server Administrator.

El servidor de administración contiene los componentes administrativos que permiten la administración de un solo o de múltiples nodos del dominio de BI, es decir administración distribuida. En concreto, el Servidor de Administración contiene la siguiente:

JMX MBeans - componentes Java que proporcionan acceso programático para la gestión de un dominio de BI.

Fusion Middleware de control - Una interfaz de usuario de administración que se utiliza para gestionar el dominio de BI.

WebLogic Server Administration Console - Una interfaz de usuario administrativo que proporciona gestión avanzada para Weblogic, componentes JEE y seguridad.

2.12.6. Oracle Business Intelligence Answers.

Autoservicio ad-hoc que permite a los usuarios finales crear fácilmente diagramas, tablas dinámicas, informes y cuadros de mando, y navegar con capacidades de drill up/down.

2.12.7. Oracle Business Intelligence Interactive Dashboards

Cuadros de mando interactivos para entornos de análisis. Los cuadros de mando son herramientas de gestión del rendimiento que se presentan ante los usuarios como una visualización de los indicadores empresariales. Permiten monitorizar, controlar y gestionar los procesos de una organización a través de excelentes gráficos tipo semáforo que utilizan códigos de colores que establecen alertas con las que disponer de una visión completa del rendimiento de la compañía.

Los cuadros de mando de gestión, con sus capacidades de alerta, proporcionan una gran ventaja y muchos beneficios. Los indicadores de cumplimiento, evaluación, eficiencia y eficacia contenidos en ellos ofrecen una visión completa de la organización y su rendimiento, permitiendo comprobar, por ejemplo, si la actividad diaria está alineada con la estrategia corporativa o interpretar lo que está ocurriendo y saber si debemos tomar medidas de mejora.

2.12.8. Oracle Business Intelligence Publisher.

Oracle Business Intelligence Publisher proporciona una única plataforma basada en la web para la creación, administración y entrega de informes interactivos y todo tipo de documentos con formato.

Los usuarios finales pueden diseñar fácilmente informes directamente en un navegador web o el uso de herramientas de escritorio familiares, lo que reduce drásticamente el tiempo y los costos necesarios para desarrollar y mantener informes.

Construido sobre estándares abiertos, el personal de tecnología de la información y los desarrolladores pueden crear modelos de datos compartibles contra prácticamente cualquier fuente de datos y el uso de las interfaces de programación de aplicaciones de BI Publisher para crear aplicaciones personalizadas que aprovechan las fuentes de datos existentes y la infraestructura extremadamente eficiente y altamente escalable, BI Publisher puede generar decenas de miles de documentos por hora, con un impacto mínimo en los sistemas transaccionales (Oracle, 2013).

2.12.9. Oracle Database Standard Edition One.

Oracle Database 11g Standard Edition One es una solución de gestión de datos con todas las funciones, ideal para las necesidades de las pequeñas y medianas empresas. Es fácil de instalar en Windows, Linux, y los sistemas operativos Unix, fácil de manejar con las capacidades de gestión automatizada y fácil de escalar a medida que su negocio y necesidades de tecnología de la información crecerán.

Oracle Database 11g Standard Edition One gestiona todos los tipos de datos y admite a todas las aplicaciones de negocio para aprovechar el rendimiento, la fiabilidad, la seguridad y escalabilidad para la que Oracle es conocido, a un bajo costo. Oracle Database 11g Standard Edition One ofrece compatibilidad ascendente completa con otra Oracle Database Edition, protegiendo su inversión como sus requisitos crecen al sentar las bases de TI para entregar con éxito más información con mayor calidad de servicio, reducir el riesgo de cambio dentro de ella, y hacer un uso más eficiente de sus presupuestos de TI (Oracle, 2009).

2.12.10. Oracle Warehouse Builder.

Herramienta para el diseño de base de datos multidimensional y para realizar procesos de extracción, transformación y carga (ETL) que ayuda a gestionar el ciclo de vida de los datos y metadatos.

CAPÍTULO III

METODOLOGÍA

3. PRINCIPALES METODOLOGÍAS ACTUALES

Desde los inicios de la Inteligencia de negocios hasta el momento, las grandes empresas desarrolladoras de software han creado sus propias metodologías para la implementación de soluciones en este campo, sin embargo cada una de estas empresas se han basado en las propuesta de Bill Inmon y Ralph Kimball con enfoques top-down y bottom-up respectivamente. Además de estas podemos mencionar la metodología Hefesto, Rapid Warehouse Builder, entre otras.

El desarrollo de un DW debe tener en cuenta las necesidades de los usuarios del negocio en lo que respecta a los análisis a realizar, que serán la base de los informes requeridos. Caso contrario no podremos hacerlo por cuanto solo será una bodega de datos de la que no se podrá extraer información que se pueda distribuir y utilizar en beneficio de la empresa. La utilidad de un DW radica en la selección de los procesos de negocio se con el fin de modelar cada uno de ellos, estableciendo su granularidad. Un aspecto muy importante que hay que tomar en cuenta son las relaciones que existen entre los datos que serán objeto de análisis.

El enfoque top-down, de arriba hacia abajo, inicia con una visión general de las necesidades del negocio para comenzar con el diseño del almacén de datos (DW) a nivel de toda la empresa, a partir de aquí se puede seleccionar áreas o secciones para elaborar los data marts (DM). De esta manera se minimiza los problemas de integración, pero su costo se eleva considerablemente.

El enfoque bottom-up, de abajo hacia arriba, empieza por una sección hasta llegar a un alto grado de depuración, luego se continua con las otras secciones y se enlazan estas hasta llegar a la construcción final del data warehouse. A diferencia del enfoque anterior los costos no son elevados porque se está aplicando a una sola sección de la

organización y se pueden ver los resultados en corto tiempo. Este enfoque parte de los requisitos de negocio los que serán validados tempranamente mientras que el enfoque top-down se lo hará al final cuando ya se termina la implementación.

3.1. Metodología propuesta por Bill Inmon

Bill Inmon es conocido mundialmente como el padre del Data Warehouse, es el responsable del desarrollo de una metodología que la presentó en el año 1992 en su libro *Building the Data warehouse*. Su diseño tiene un enfoque descendente (top-down), en él propone emprender el proyecto desde lo general a lo particular, construir lo que se conoce como la Fábrica Corporativa de Información.

La metodología de Inmon sigue etapas opuestas al ciclo de vida clásico de desarrollo de sistemas de apoyo a las decisiones llamados SDLC, conocido también como desarrollo en cascada, en el que se empieza por establecer primero los requerimientos para luego entrar en la etapa de diseño y desarrollo continuando con las pruebas y la implementación final del sistema.

Esta propuesta implica un gran riesgo para las organizaciones ya que tendrán que realizar fuertes inversiones para emprender el diseño y desarrollo de todo el Data warehouse empresarial, el mismo que tardará algún tiempo antes de que se pueda tener los primeros Data mart y comiencen a aportar sus primeros beneficios. De acuerdo a lo que afirma.

Los CLDS comienzan con datos. Una vez que los datos están en la mano, estos son integrados y luego se examina para ver el sesgo en los datos, si los hay. Programas se escriben en contra de los datos. El resultado de los programas se analizan, y finalmente se entienden los requisitos del sistema. Una vez que se entienden los requisitos, se realizaron ajustes en el diseño del sistema, y el ciclo comienza de nuevo por un conjunto diferente de datos. Debido a la constante reposición del ciclo de vida de desarrollo de diferentes tipos de datos, el enfoque

de desarrollo CLDS se suele llamar una metodología de desarrollo en espiral.
(Inmon, 2005, p.14)

Por esta razón siempre se relaciona a este autor con los almacenes de datos empresariales (data warehouse), en donde los mercados de datos (data marts) son partes dependientes del almacén que se diseñan una vez terminado el diseño de este. Lo dicho en esta sección se encuentra representado en la figura 20.

La intención de este enfoque es evitar la aparición peligros potenciales para el proyecto si llegaran a presentarse condiciones imprevistas, debido a que se conoce de antemano la estructura de todo el proyecto con sus puntos críticos perfectamente identificados (Rodríguez, 2010).

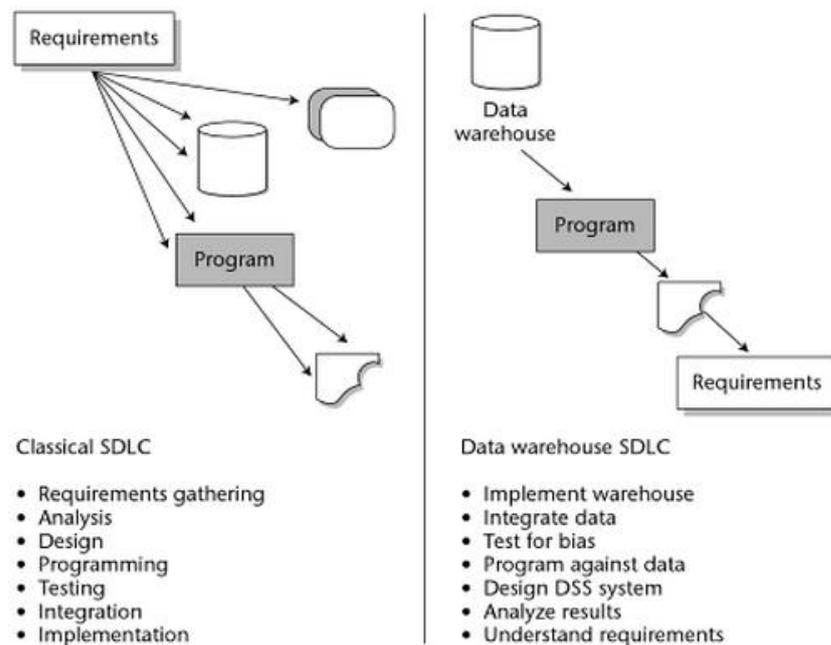


Figura 24: Building the Data warehouse

Fuente: Building the Data warehouse, Inmon, 2005

Inmon utiliza el modelo relacional de bases de datos para la implementación del data warehouse porque para él la base de un data mart es una base de datos con una ligera normalización ya que sostiene que las necesidades de información de una empresa no se pueden conocer en su totalidad por la dinámica del entorno en el que se desenvuelve.

Al utilizar el modelo entidad – relación con un enfoque global el tiempo empleado en el desarrollo se alarga, lo que viene a constituir una gran desventaja de la propuesta de Inmon para el diseño e implementación del data warehouse empresarial (Rodríguez, 2010).

3.2. Rapid Warehousing Methodology

Entre las metodologías que se emplean actualmente está la técnica “Rapid Warehousing Methodology (RWM)” propuesta por el Instituto SAS (EEUU), en el año 1998. Dicha metodología es iterativa y está basada en el desarrollo incremental del proyecto de DWH dividido en cinco fases:

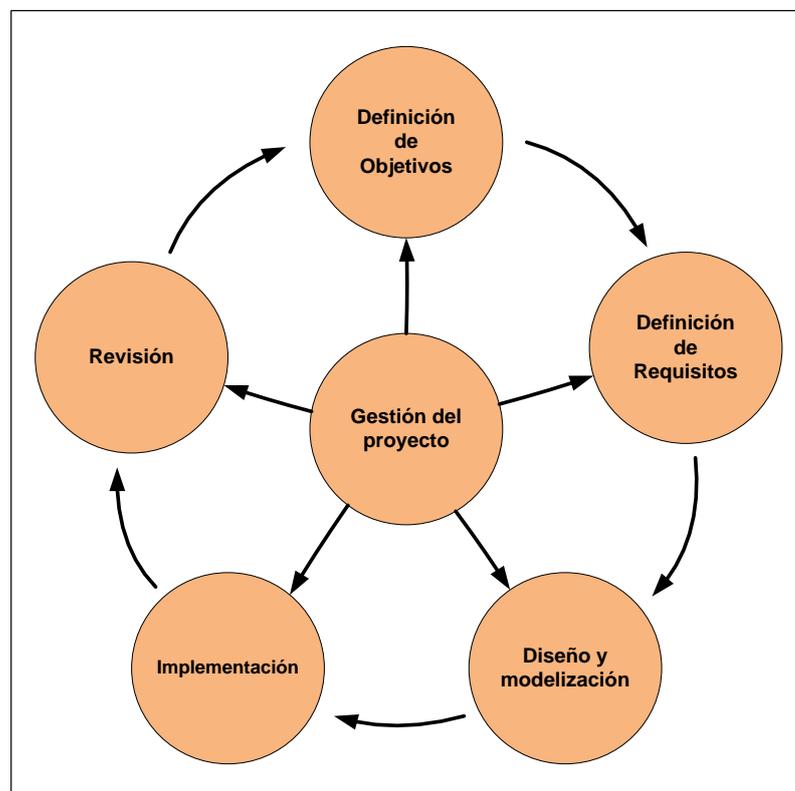


Figura 25: Metodología Rapid Warehouse Builder

Fuente: Elaboración propia

3.2.1. Definición de los objetivos

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 realizará como suministrador de información de negocio estratégica para la empresa. Se definirán así mismo, los parámetros que permitan evaluar el éxito del proyecto.

3.2.2. Definición de los requerimientos de información

Durante esta fase se mantendrán sucesivas entrevistas entre los representantes del departamento usuario final y los representantes del departamento de informática. Se realizará el estudio de los sistemas de información existentes, que ayudarán a comprender las carencias actuales y futuras que deben ser resueltas en el diseño del DWH.

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 DWH.

3.2.3. Diseño y modelización

Los requerimientos de información identificados durante la anterior fase proporcionarán las bases para realizar el diseño y la modelización del DWH. 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 DWH. Este modelo estará formado por entidades y relaciones que permitirán resolver las necesidades de negocio de la organización.

3.2.4. Implementación

La implantación de un DWH lleva implícitos los siguientes pasos:

- ✓ Extracción de los datos del sistema operacional y transformación de los mismos.
- ✓ Carga de los datos validados en el DWH. Esta carga deberá ser planificada con una periodicidad que se adaptará a las necesidades de refresco detectadas durante la fase de diseño del nuevo sistema.
- ✓ Explotación del data warehouse mediante diversas técnicas dependiendo del tipo de aplicación que se dé a los datos. Entre las técnicas más habituales podemos encontrar las siguientes:
 - Query & Reporting
 - On-line analytical processing (OLAP)
 - Executive Information System (EIS) ó Información de gestión
 - Decision Support Systems (DSS)
 - Visualización de la información

3.2.5. Revisión

La construcción del DWH 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 DWH 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.

3.2.6. Gestión del Proyecto

La gestión del proyecto debe encargarse de la coordinación y ejecución de las distintas fases que conforman la construcción e implantación de un DWH. Este proceso se tiene que apoyar en una metodología específica para este tipo de trabajos, si bien es más importante que la elección de la mejor de las metodologías, el realizar un control para asegurar el seguimiento de la misma.

En las fases que se establezcan es fundamental incluir una fase de formación en la herramienta utilizada, para un máximo aprovechamiento de la aplicación. Seguir los pasos de la metodología y comenzar el DWH por un área específica de la empresa permitirá obtener resultados tangibles en un corto espacio de tiempo.

3.3. Metodología propuesta por Ralph Kimball

La metodología propuesta por Ralph Kimball se compone de 6 fases bien definidas y que se representan en la figura 26. Las Fases son:

- ✓ Planificación
- ✓ Definición de requerimientos
- ✓ Línea tecnológica
- ✓ Línea de datos
- ✓ Línea de aplicación de BI
- ✓ Despliegue.

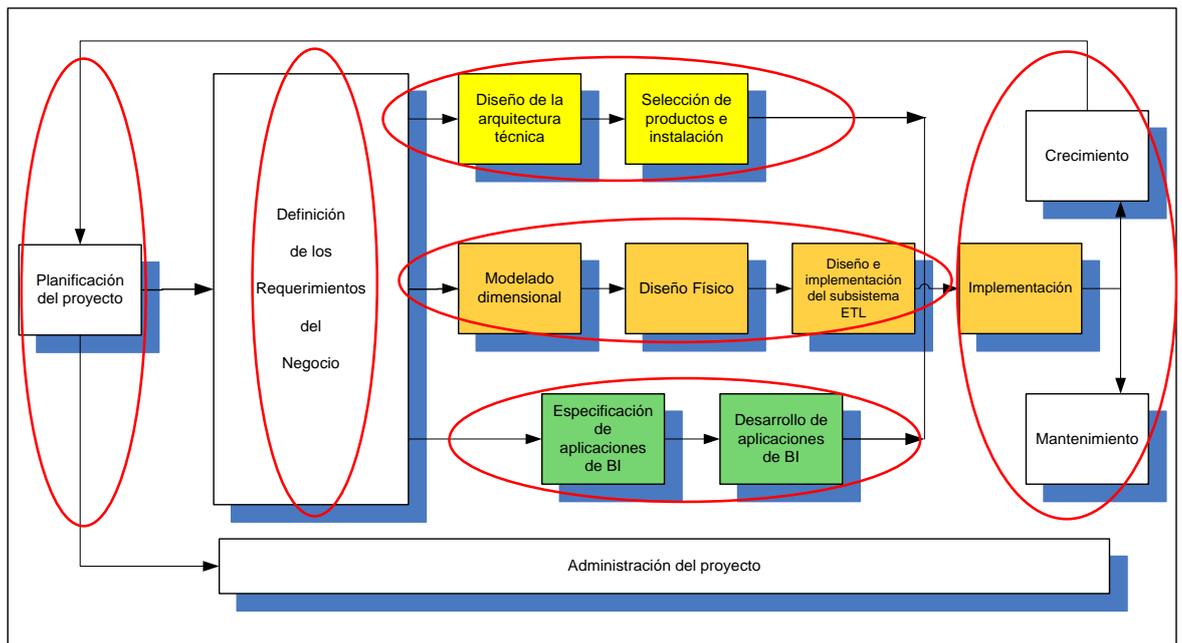


Figura 26: Metodología de Ralph Kimball

Fuente: Elaboración propia

3.3.1. Planificación del Proyecto

Durante la etapa de planificación definimos el alcance del proyecto y entendemos los requerimientos del negocio, identificamos las tareas a realizar durante el desarrollo del proyecto con su organización respectiva, debemos también decidir en que se utilizarán los recursos y durante cuánto tiempo, es decir, que carga de trabajo asignaremos a las personas involucradas en el proyecto.

3.3.2. Definición de los Requerimientos

De la misma forma que cualquier solución de software en una aplicación de inteligencia de negocios juegan un papel muy importante. Según Kimball, los requerimientos del negocio constituyen parte medular del data warehouse. El alcance de un proyecto de data warehouse es determinado por los requerimientos del negocio, a quien va dirigido, que personas pueden tener acceso y los lugares desde los que accederán, que datos serán trasladados de la fuente al destino.

Al respecto Kimball sostiene que el medio para poder determinar correctamente los requerimientos del negocio se centran en la aplicación de entrevistas debidamente planificadas de acuerdo al papel que desempeña cada uno de los entrevistados en la empresa.

3.3.3. Línea Tecnológica

Esta fase se subdivide en dos partes diseño de la arquitectura técnica y la selección de productos e instalación, como se muestra en la figura 27.

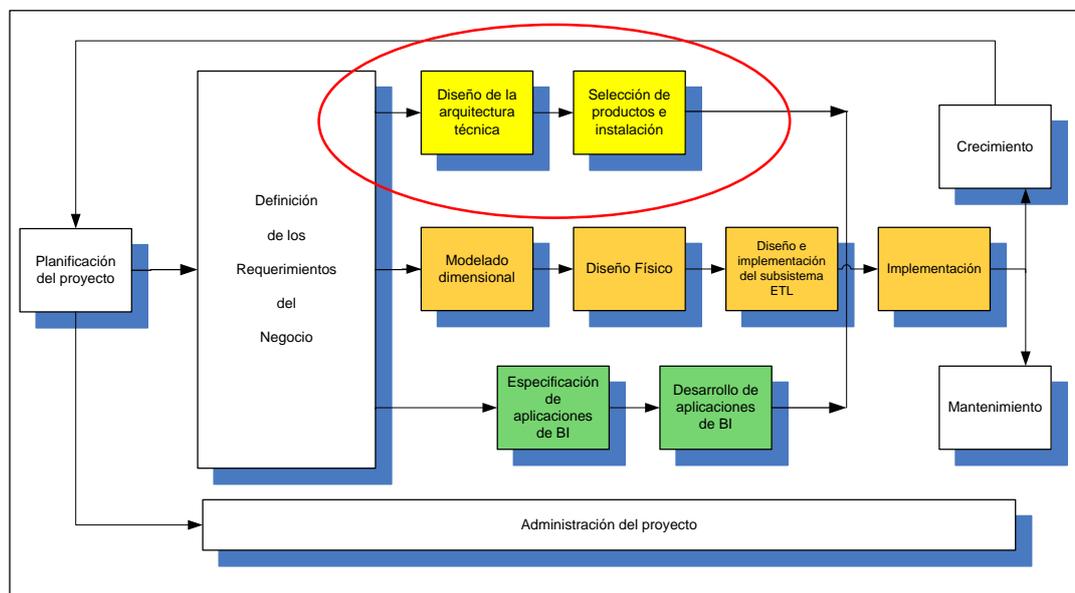


Figura 27: Línea Tecnológica

Fuente: Elaboración propia

3.3.3.1. Diseño de la Arquitectura Técnica

Las tecnologías que intervienen en proyectos de Warehousing son variadas, razón por la que se debe tener en cuenta en primer lugar, y como ya se mencionó anteriormente los requerimientos del negocio, en segundo lugar la plataforma tecnológica actual con la que cuenta la empresa, tanto en hardware como en software y en tercer lugar cual es la orientación de la empresa para los próximos años, cuáles son sus planes de desarrollo tecnológico para el futuro.

3.3.3.2. Selección de Productos e Instalación

Basándose en arquitectura técnica diseñada en la fase anterior se propone alternativas de los elementos de la arquitectura como pueden ser la plataforma de hardware, el motor de base de datos, la herramienta de ETL, sistema operativo, herramientas de reporteria.

3.3.4. Línea de datos

Esta fase la podemos subdividir en tres: modelado dimensional, diseño físico y diseño de procesos ETL como se puede observar en la figura 28.

3.3.4.1. Modelado Dimensional

La definición de los requerimientos del negocio determina los datos necesarios para cumplir los requerimientos de análisis de los usuarios. Diseñar los modelos de datos para soportar estos análisis requiere un enfoque diferente al usado en los sistemas operacionales. En un modelo dimensional se trabaja con tablas de hechos y dimensiones que se encuentran des normalizadas; en un modelo relacional tradicional, en cambio, se usa un conjunto de tablas que se encuentran relacionadas y normalizadas.

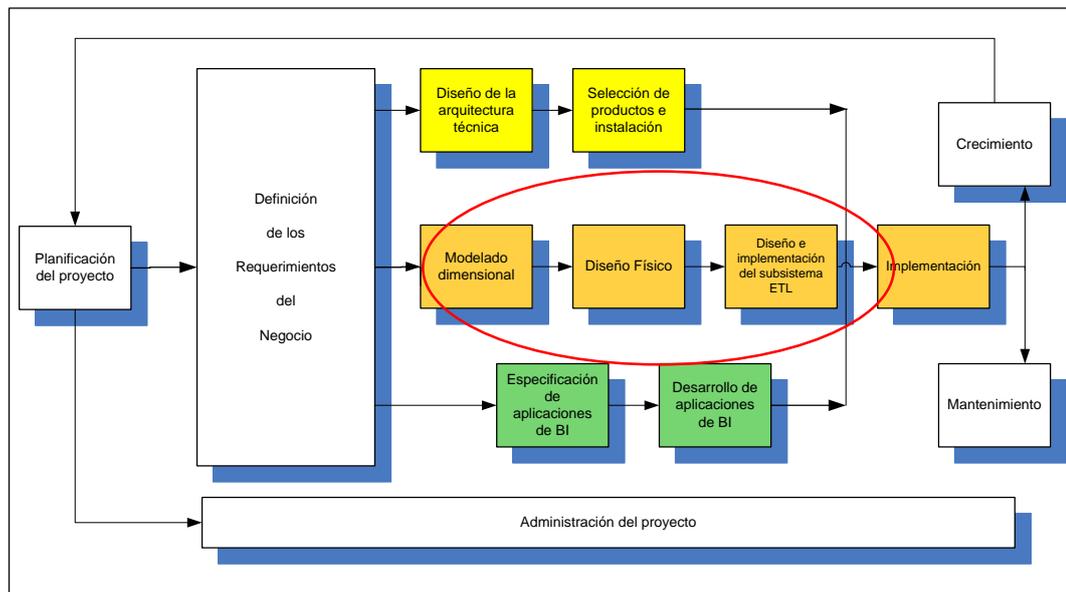


Figura 28: Línea de datos

Fuente: Elaboración propia

3.3.4.2. Diseño Físico

El diseño físico de las base de datos se centra en la selección de las estructuras necesarias para soportar el diseño lógico. Algunos de los elementos principales de este proceso son la definición de convenciones estándares de nombres y ajustes específicos del entorno de la base de datos.

3.3.4.3. Diseño del subsistema de procesos ETL

Se definen como procesos de transformación para convertir o recodificar los datos fuente a fin poder efectuar la carga efectiva del modelo físico. Por otra parte, los procesos de carga de datos son los procesos requeridos para poblar el data warehouse.

3.3.5. Línea de aplicaciones del BI

Esta se compone de dos etapas diseño de aplicaciones BI e implementación de las aplicaciones de BI. En la figura 29 se muestra estos componentes.

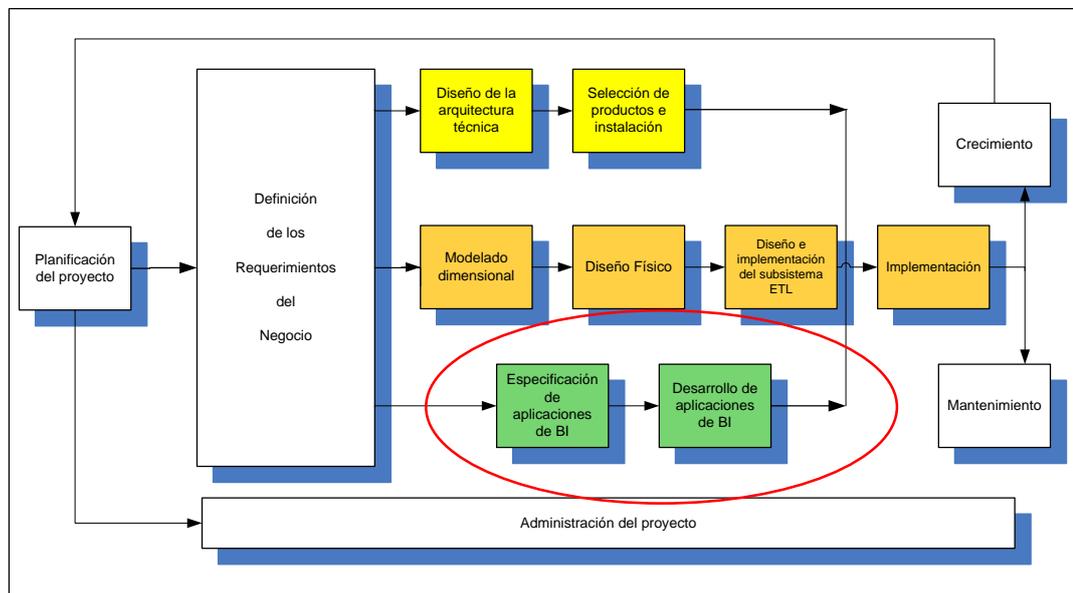


Figura 29: Línea de aplicación del BI

Fuente: Elaboración propia

3.3.5.1. Diseño de aplicaciones del BI

Utilizando el diseño de arquitectura técnica como marco, es necesario evaluar y seleccionar componentes específicos de la arquitectura como la plataforma de hardware, el motor de base de datos, la herramienta de ETL o el desarrollo pertinente, herramientas de acceso, entre otros.

3.3.5.2. Desarrollo de aplicaciones del BI

No todos los usuarios del warehouse no necesitan el mismo nivel de análisis. Es por ello que en esta etapa se identifican los diferentes roles o perfiles de usuarios para determinar los diferentes tipos de aplicaciones necesarias en base al alcance de los diferentes perfiles gerencial, analista del negocio, vendedor, etc.

3.3.6. Despliegue

Esta fase está subdividida en tres Implementación, mantenimiento y crecimiento, a las dos últimas se las tratará en un solo ítem. En la figura 30 se muestra esta sección.

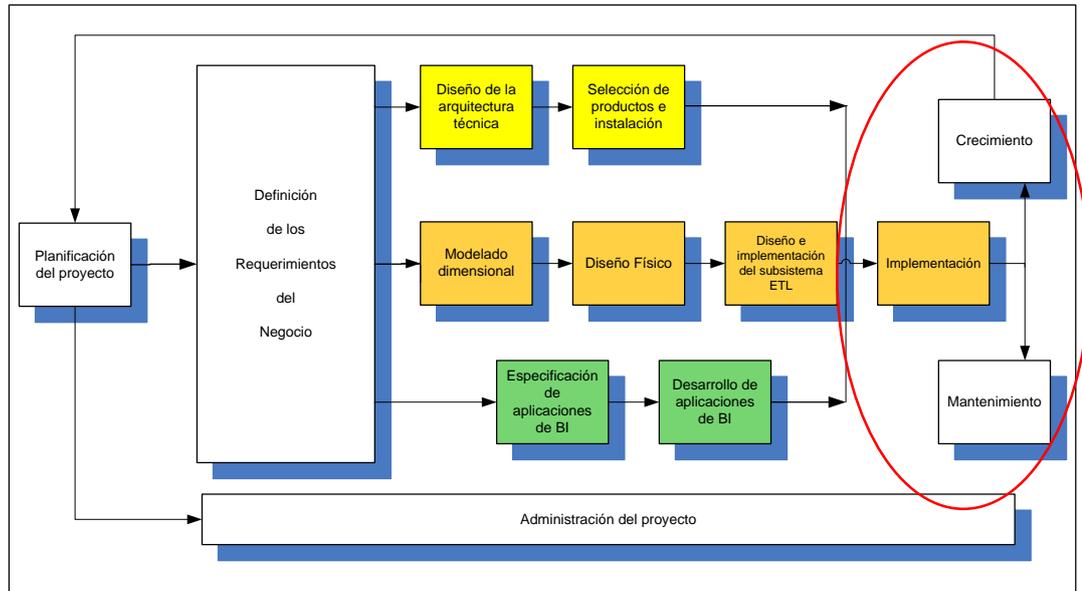


Figura 30: Despliegue

Fuente: Elaboración propia

3.3.6.1. Implementación

La implementación representa la convergencia de la tecnología, los datos y las aplicaciones de usuarios finales accesible desde el escritorio del usuario del negocio. Hay varios factores extras que aseguran el correcto funcionamiento de todas estas piezas, entre ellos se encuentran la capacitación, el soporte técnico, la comunicación, las estrategias de retroalimentación.

3.3.6.2. Mantenimiento y Crecimiento

La construcción de un Data warehousing es un proceso de etapas bien definidas, con comienzo y fin, pero de naturaleza espiral pues acompaña a la evolución de la organización durante toda su historia. Se necesita continuar con los relevamientos de forma constante para poder seguir la evolución de las metas por conseguir.

Según afirma Kimball, “si se ha utilizado el ciclo de vida de desarrollo dimensional en el data warehouse, está preparado para evolucionar y crecer”. Al contrario de los sistemas tradicionales, los cambios en el desarrollo deben ser vistos como signos de éxito y no de falla.

3.3.7. Gerenciamiento del Proyecto

El gerenciamiento del proyecto asegura que las actividades del ciclo de vida de desarrollo del data warehouse se realicen a tiempo en forma sincronizada, que los recursos tecnológicos sean los apropiados, que no se descuide ningún detalle y que se acompañe el procesos desde el inicio hasta su finalización, tal y como lo indica figura 31 del modelo propuesto por el autor, Ralph Kimball quien sostiene que las actividades principales se encuentra el monitoreo del estado del proyecto y la comunicación entre los requerimientos del negocio y las restricciones de información para poder manejar correctamente las expectativas en ambos sentidos.

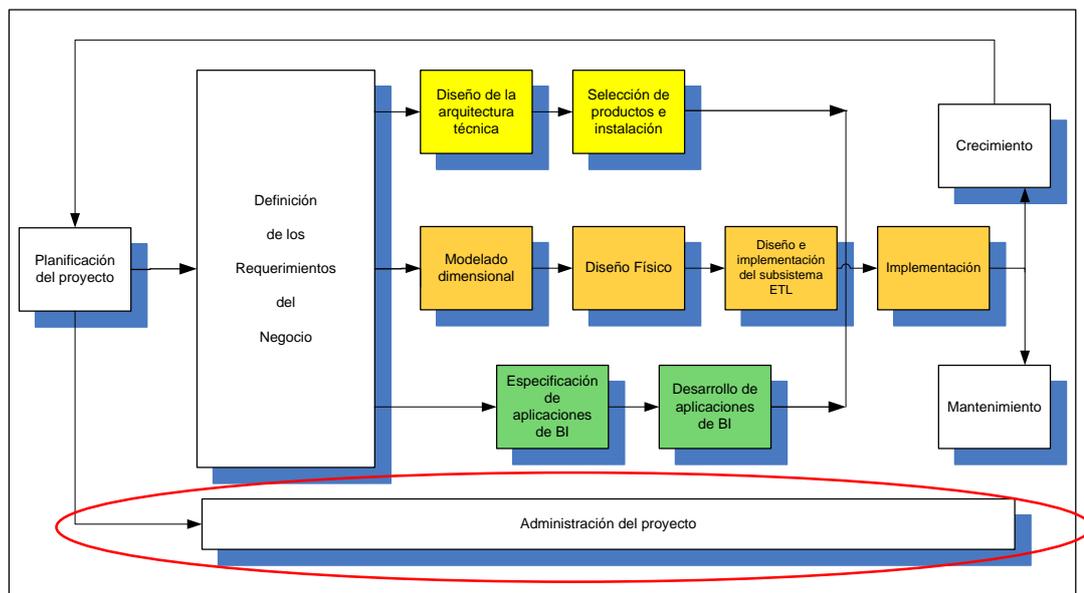


Figura 31: Administración del proyecto

Fuente: Elaboración propia

3.4. Selección de la metodología

La metodología seleccionada en nuestro caso es la de Kimball, ya que su enfoque de abajo hacia arriba se ajusta perfectamente al tema del proyecto, construir una data mart aplicado a la toma de decisiones académicas de la Universidad Técnica del Norte, ya que no se intenta construir el Data Warehouse de toda la Institución de esta manera podemos obtener los beneficios en un lapso de tiempo relativamente corto, además se podrá desarrollar de a poco los data marts restantes para completar con el Data warehouse de toda la Universidad. Otro factor que se tomó en cuenta en la selección es la cantidad de documentación, casos de uso existente sobre esta metodología.

CAPÍTULO IV

IMPLEMENTACIÓN Y DESARROLLO

4. Desarrollo de la metodología seleccionada

En los apartados siguientes se irán describiendo cada una de las etapas por la que tendrá que pasar la construcción del Data mart Académico propuesto, siguiendo las etapas que se proponen en la metodología de Ralph Kimball, que es la elegida por las razones expuestas en la sección correspondiente.

4.1. Planeación del proyecto

Es necesario recordar que durante esta etapa se define el objetivo principal del proyecto y cuál su alcance que ya ha sido mencionado en el capítulo I. El presente proyecto tiene como finalidad la construcción de un data mart aplicado a la gestión académica de la Universidad Técnica del Norte, los usuarios beneficiarios directos son: Rectorado, Vicerrectorado Académico, Vicerrectorado Administrativo, Decanatos, Sub Decanatos y Coordinadores de Carrera.

El objetivo principal del proyecto es: Implementar una Solución de Inteligencia de Negocios mediante el diseño de una aplicación en Oracle Business Intelligence para contribuir en la mejora de la toma de decisiones de la Gestión Académica de la Universidad Técnica del Norte.

4.2. Definición de requerimientos

Los requerimientos de usuario se han resumido previamente en la Tabla N°1 que se expone en el capítulo I, en esta tabla se pueden observar que las necesidades se enfocan en los procesos que involucran a estudiantes y docentes específicamente de los que se

desprenden los principales indicadores requeridos por el Vicerrectorado Académico, obtenidos en los diálogos mantenidos con su titular.

Para iniciar con definición de los requerimientos analizamos los procesos que se dan en torno a los estudiantes y docentes, seleccionamos aquellos que nos permitirán responder a los requerimientos que se detallan en la tabla N°1, estos se describen a continuación.

4.2.1. Matrícula

Este proceso comienza con la inscripción de los aspirantes y una vez concluida esta fase tienen que someterse a un examen de ingreso, luego se publican los aprobados quienes continúan el proceso para obtener su matrícula. El análisis que se quiere realizar consiste en determinar cuántos estudiantes se inscriben en un determinado período académico, de estos cuántos se matricularon, agrupados por género, etnia, facultad, escuela y carrera.

4.2.2. Promoción.

El proceso de promoción tiene que ver con los resultados de la actividad académica del estudiante y está bajo la responsabilidad de los docentes, quienes registran las notas conforme se desarrolla el proceso de enseñanza aprendizaje cada cierto período de tiempo perfectamente establecido en el calendario académico de la universidad, una vez alcanzada la fecha máxima para ingresar notas y finalizado el ciclo académico el sistema calcula la nota final para cada estudiante y en cada materia que se encuentra matriculado, cabe señalar que ciertos estudiantes deciden retirarse o anular su matrícula, siempre tomando en cuenta las leyes y reglamentos vigentes.

El análisis a realizar en este punto consiste en determinar el número de estudiantes que aprobaron, perdieron, se retiraron, anularon agrupados por género, etnia, facultad, escuela, carrera y materia.

4.2.3. Asignación de carga horaria a docentes.

Basándose en el plan de estudio de la carrera, aprobado por la autoridad educativa nacional, este proceso se inicia con la creación de los cursos a partir del primer nivel y los subsiguientes, la asignación de las materias, horas prácticas, horas teóricas, esto por cada curso considerado dentro del plan curricular institucional, a continuación la asignación de docentes para los cursos creados, luego está la elaboración de horarios.

El análisis que se desea realizar en esta parte consiste en determinar el número de docentes que firmaron contratos, con su carga horaria, su dedicación, tiempo que dedican a la formación, horas de vinculación, número de horas de investigación, número de horas en cada facultad, número de docentes a tiempo completo, medio y cuarto de tiempo. Además de los docentes a contrato.

4.2.4. Graduación.

Inicia el momento en el que el estudiante ha aprobado todos los niveles, cumple con los demás requisitos y realiza los trámites necesarios para poder graduarse y posterior titulación. Se inicia con la presentación de un anteproyecto para empezar con su trabajo de grado que una vez aprobado pone en marcha el proyecto que culmina con la defensa privada y la pública, luego de cumplidos los requisitos se le confiere el título profesional por el que ha optado.

En reuniones periódicas con el jefe de proyectos de la Dirección de Desarrollo Tecnológico e Informático, se ha podido establecer que la institución cuenta con un potente sistema transaccional integrado, en el que la universidad se apoya para realizar sus operaciones a diario, las que se enmarcan en plan estratégico de la Universidad así como en su estatuto, reglamento interno y las leyes y reglamentos expedidos por la autoridad educativa nacional.

4.2.5. Análisis de la fuente de datos

En los proyectos de inteligencia de negocios la materia prima se constituye la información que se puede producir con los datos que se encuentran almacenados en las bases de datos de sus sistemas transaccionales, a lo que se le llama fuente de datos. Es necesario analizar estas bases de datos para poder determinar la calidad de los mismos y si es necesario realizar un proceso de limpieza para su posterior extracción y carga hacia el data mart para lo que se hace imprescindible diseñar los procesos ETL correspondientes.

La base de datos de la Universidad Técnica del Norte se compone de 1359 tablas que almacenan los datos de todas operaciones que se realizan a diario en la universidad. De la totalidad de estas tablas serán tomadas en cuenta 26 de ellas que corresponden al área académica, de recursos humanos, bienestar universitario e institucional como se detalla en la tabla 2.

N°	Nombre de la tabla
1	ACA_TAB_ESTUDIANTES
2	ACA_TAB_MATRICULAS
3	ACA_TAB_EJES_FORMACION
4	ACA_TAB_CICLOS_ACADEMICOS
5	ACA_TAB_NIVELES
6	ACA_TAB_ESPECIALIDADES
7	ACA_TAB_ESTUDIANTES
8	ACA_TAB_MODALIDADES_ESTUDIOS
9	ACA_TAB_DEPENDENCIAS
10	ACA_TAB_MATERAS
11	ACA_TAB_ACTIVIDADES
12	ACA_TAB_SISTEMA_ESTUDIOS
13	ACA_TAB_AREAS_ACADEMICAS
14	ACA_TAB_PERIODEOS_ACADEMICOS

Nº	Nombre de la tabla
15	ACA_TAB_COLEGIOS
16	ACA_TAB_TIPOS_INSTITUCIONES
17	ACA_TAB_TIPOS_FINANCIAMIENTO
18	ACA_TAB_TIPOS_CICLOS_ACADEMICOS
19	ACA_TAB_SUBAREAS_UNESCO
20	DBU_TAB_ETNIA
21	DBU_TAB_SUBGRUPO_DISCAPACIDAD
22	RHU_TAB_PERSONAS
23	RHU_TAB_EMPLEADOS
24	RHU_TAB_ESCALAFONES
25	INS_TAB_LOCALIDADES
26	INS_TAB_TIPOS_LOCALIDADES

Tabla 2: fuente de datos para los data mart

Fuente: Elaboración propia

4.2.6. Calidad de datos

Al hacer un análisis de los datos almacenados en las tablas listadas anteriormente, se pudo encontrar campos con valores nulos y en otros casos datos faltantes, o que no han sido registrados.

4.3. Línea Tecnológica

En esta parte se define cuál es la Arquitectura tecnológica con la que se trabajará además se seleccionan las herramientas de software y se procede a su instalación.

4.3.1. Arquitectura tecnológica

En el departamento de desarrollo tecnológico e informático se cuenta con un servidor en donde está instalada la base de datos Oracle 11g sobre el sistema operativo Oracle

Linux 6.6, además se cuenta con un servidor de aplicaciones llamado aplicaciones2 que proporciona el acceso al sistema integrado de la Universidad.

Se requiere contar con un servidor dedicado específicamente para las aplicaciones de inteligencia de negocios, esto para asegurar el éxito del proyecto al no tener que interferir con la base de datos del sistema transaccional evitando innumerables fallas. Este servidor debe estar en concordancia con los requerimientos de la plataforma de software a instalar.

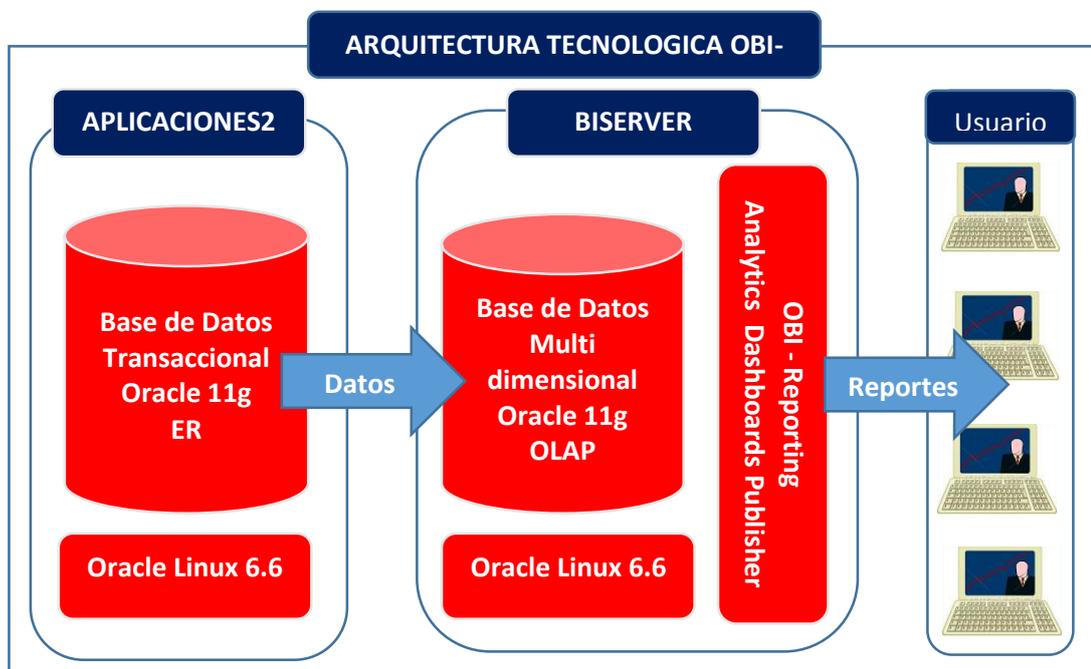


Figura 32: Arquitectura Tecnológica OBI - UTN

Fuente: Elaboración propia

4.3.2. Selección de productos e instalación

La herramienta de inteligencia de negocios seleccionada es Oracle Business Intelligence 11g Standart Edition One, los factores determinantes para esta decisión fueron: se cuenta actualmente con licencias para el uso de este software, además como se cuenta ya con el ERP desarrollado con productor Oracle la integración no se dificultará y se reducirán los costos tanto de adquisición, licenciamiento,

mantenimiento y su crecimiento. El sistema operativo del servidor será de igual forma Oracle Linux en su versión 6.6.

4.4. Línea de Datos (Modelamiento Dimensional)

En esta parte Kimball define ciertas fases para el modelamiento dimensional que listamos a continuación:

- a) Definición de la granularidad.
- b) Elección de las dimensiones.
- c) Identificación de los hechos.
- d) Detalle de las tablas de dimensión.

4.4.1. Elección de la granularidad

La granularidad, como ya se ha expuesto en apartados anteriores, consiste en definir la profundidad del grado de detalle con el que se hará el análisis de la información, en otras palabras, se define la estructura de las tablas de hechos. Como parte de la metodología de Kimball es necesario recordar que él recomienda que debe ser lo más baja posible, de manera tal que ya no se pueda desglosar más en nivel de detalle. Esta forma de definir la granularidad nos brinda la flexibilidad para que los usuarios puedan ellos mismos realizar consultas directas.

Bajo esta perspectiva y atendiendo a las necesidades o requerimientos de la gestión académica en la institución expuestos en la tabla N°1, el análisis que se espera realizar abarca a la historia general del estudiante que se matricula y su trayectoria general en la universidad hasta su titulación, para lograrlo es necesario obtener una fila por cada estudiante en la tabla de hechos logrando una estructura en la que se puede juntar la siguiente información.

ANÁLISIS DE ESTUDIANTES

- Fecha de inscripción.
- Fecha de matrícula.
- Estado de su matrícula.
- Es hombre o mujer.
- Etnia a la que pertenece.
- Facultad en la que estudia.
- Escuela a la que pertenece su carrera.
- Carrera que estudia.
- Materias que estudia.
- Materias aprobadas.
- Materias reprobadas.
- Número de estudiantes retirados.
- Que matrículas anuló.
- Tiene 1ra, 2da o 3ra matrícula.

Tabla 3: Análisis de Estudiantes

Fuente: Elaboración propia

Además se requiere un análisis de los docentes en lo que respecta a su condición dentro de la institución de acuerdo a su contrato, carga horaria, vinculación e investigación, la estructura de la información en relación al análisis que se pretende realizar se puede describir de la siguiente manera:

ANÁLISIS DE DOCENTES

- Fecha de firma del contrato.
 - Tipo de contrato.
 - Docente hombre o mujer.
 - Etnia a la que pertenece.
 - Facultad en la que imparte.
 - Escuelas en las que es docente.
 - Carreras en las que enseña.
-

ANALISIS DE DOCENTES
➤ Materias que dicta.
➤ Carga horaria.
➤ Horas de investigación.
➤ Horas de vinculación.

Tabla 4: análisis de docentes

Fuente: Elaboración propia

4.4.2. Identificación de hechos

Una vez definidos los análisis a realizar se debe determinar los hechos del negocio que serán medidos los mismos que se describen en la tabla 5.

HECHOS
Número_estudiantes.
Nota_Grado.
Título_Grado.
Número_docentes.
Horas_clase.
Horas_gestión.
Horas_Preparación.
Horas_vinculación.
Horas_tutoría.
Horas_capoacitación.
Antigüedad.
Número_libros
Número_artículos
Número_estudiantes

Tabla 5: Hechos o medidas a considerar en el proyecto

Fuente: Elaboración propia

4.4.3. Elección y detalle de las dimensiones.

Las dimensiones son tablas que describen las medidas de la tabla de hechos, estas tablas se caracterizan por tener varias columnas de atributos pudiendo llegar fácilmente a 100 columnas, como contraste tienen un número reducido de filas.

Según Ralph Kimball, la calidad de un almacén de datos se mide por la calidad de las dimensiones, calidad que está determinada por una correcta elección de sus atributos. Los atributos de las dimensiones deben tener nombres entendibles y no siglas o contracciones de palabras.

Una vez que se ha escogido el grano se analizó que dimensiones tiene asociadas para este proyecto, y se pudo determinar que las dimensiones más apropiadas son: tiempo, estudiantes, graduados, aprobados, docentes, matrículas, dependencias, nómina y localidades. A continuación se describen las tablas de dimensiones y las de hechos que se crearán.

4.4.3.1. Dimensión Estudiante

La dimensión estudiante almacenará los datos de todos los estudiantes, con algunos de sus datos personales, así como también de sus estudios secundarios. El atributo ID es la clave primaria de tipo numérico.

Dimensión ESTUDIANTE
ID (PK)
CEDULA
NOMBRE
GENERO
DIRECCION
TIPO_IDENTIFICACION
EMAIL
EMAIL_INSTITUCIONAL
ETNIA
FECHA_NACIMIENTO

Dimensión ESTUDIANTE
ESTADO_CIVIL
TELEFONO
CELULAR
DISCAPACIDAD
COLEGIO
FECHA_DE_GRADO_COLEGIO
ESPECIALIDAD_COLEGIO
TIPO_DE_COLEGIO
PORCENTAJE_DISCAPACIDAD
CARNET_CONADIS
NOTA_BACHILLER

Figura 33: Dimensión Estudiante

Fuente: Elaboración propia

4.4.3.2. Dimensión Tiempo

Esta dimensión se considera de gran importancia, ya que por medio de esta se podrán presentar los datos que corresponden a la historia del estudiante durante su paso por la institución. Durante el proceso de enseñanza aprendizaje el estudiante puede verse inmerso en una serie de sucesos como matrícula, aprobación o no de materias, anulación de matrícula, entre otros.

Dimensión TIEMPO
ID (PK)
AÑO
MES
DÍA
SEMESTRE
TRIMESTRE
CICLO ACADÉMICO

Figura 34: Dimensión Período

Fuente: Elaboración propia

4.4.3.3. Dimensión Matricula

Esta tabla almacenará la información de la matrícula de los todos los estudiantes con la que se podrá obtener el número de estudiantes con primera, segunda o tercera matrícula.

Dimensión MATRÍCULA
ID (PK)
CODIGO_MATRICULA
CEDULA
MODALIDAD
TIPO_DE_CICLO
TIPO_DE_FINANCIAMIENTO
SISTEMA_ESTUDIOS
NIVEL
NUMERO_MATRICULA
ESTADO
TIPO_MATRICULA
TIPO_BECA
USUARIO

Figura 35: Dimensión Matrícula

Fuente: Elaboración propia

4.4.3.4. Dimensión Nómina

Contiene la información correspondiente a los datos laborales de los docentes de la Universidad..

Dimensión NÓMINA
ID (PK)
CEDULA
CICLO_ACAD_CODIGO
DESIGNACION
DEDICACION
RELACION_LABORAL

Figura 36: Dimensión Período

Fuente: Elaboración propia

4.4.3.5. Dimensión Materias

En esta tabla se guardan todas las materias que se imparten en todas las carreras de todas las escuelas y por supuesto de todas las carreras.

Dimensión MATERIAS
ID (PK)
CODIGO
NOMBRE
NIVEL
AREA_ACADEMICA
EJE_FORMACION
CODIFICACION

Figura 37: Dimensión Asignatura

Fuente: Elaboración propia

4.4.3.6. Dimensión Dependencia

Guarda la información sobre todas las carreras que se ofertan en toda la institución.

Dimensión DEPENDENCIA
ID (PK)
CODIGO
NOMBRE
SIGLA
SUBAREA
ESTADO

Figura 38: Dimensión Dependencia

Fuente: Elaboración propia

4.4.3.7. Dimensión Docentes

En esta dimensión se guarda la información personal de todos los docentes de la Institución.

Dimensión DOCENTES
ID (PK)
CEDULA
NOMBRE
TIPO_IDENTIFICACION
DIRECCION
FECHA_NACIMIENTO
GENERO
ESTADO_CIVIL
EMAIL
EMAIL_INSTITUCIONAL
ETNIA
TELEFONO
CELULAR
DISCAPACIDAD
PEROCENTAJE_DISCAPACIDAD
CARNET_CONADIS
CODIGO_UNICO

Figura 39: Dimensión Dependencia

Fuente: Elaboración propia

4.4.3.8. Dimensión Ciclos Académicos

En la Universidad las labores se caracterizan por que se desarrollan cada cierto intervalo de tiempo, el más largo es de años y después vienen los semestres, bimestres, meses y días. Estos lapsos de tiempo tienen una fecha de inicio y una de finalización, el estado del ciclo activo mientras la fecha está en el intervalo de tiempo una vez que se alcanza la fecha de finalización pasa al estado Inactivo.

Dimensión CICLO_ACADEMICO
ID (PK)
CODIGO
NOMBRE
SIGLA
SUBAREA
ESTADO

Figura 40: Dimensión Período

Fuente: Elaboración propia

4.5. Modelo físico

Usando la herramienta Oracle Warehouse Builder que viene integrada en el servidor de bases de datos Oracle 11g se inicia el proceso creación de las conexiones a la base de datos. En primer lugar creamos ubicaciones como se muestra en la figura 41.

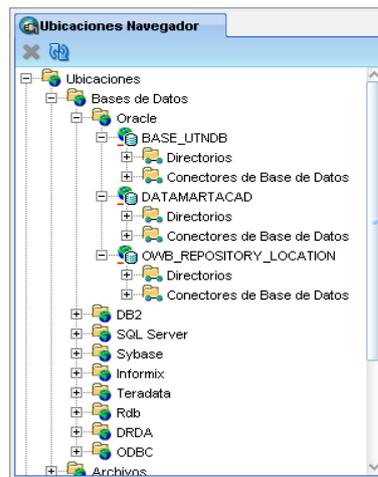


Figura 41: Creación de Ubicaciones en OWB 11g

Fuente: elaboración propia

Como se puede observar se crearon dos ubicaciones una para el origen de datos BASE_UTNDB y otra para el destino DATAMARTACAD, la primera esta ubicación servirá para recuperar los metadatos de la fuente y la segunda para almacenar los datos seleccionados en la base de datos multidimensional. A continuación crearemos dos módulos de igual forma un módulo para el origen ORIGEN_DM_UTN y otro para el destino DESTINO_DM_UTN. A continuación creamos los módulos uno para el origen de datos y otro para el destino, estos nos ayudarán a crear los procesos de extracción, transformación y carga de datos desde la base de datos de la UTN hasta el modelo dimensional. Lo descrito se muestra en la figura 42.

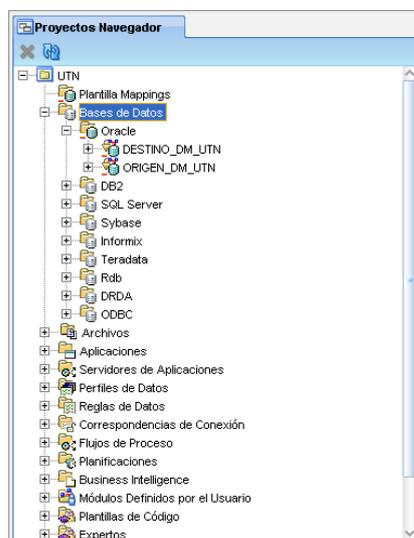


Figura 42: Creación de módulos en OWB 11g

Fuente: Elaboración propia

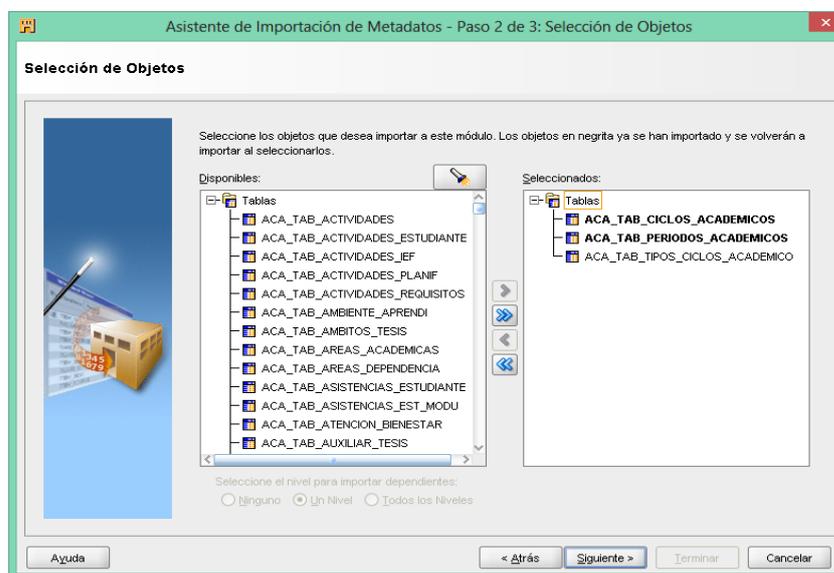


Figura 43: Importación de metadatos en OWB 11g

Fuente: Elaboración propia

En la figura 43 se muestra la selección de las tablas de la base de datos de la Universidad que servirán para alimentar las dimensiones y tablas de hechos diseñadas en pasos anteriores y que se encuentran listadas en la tabla 2. Esto lo realizamos desde el módulo origen en donde creamos una correspondencia.

Ahora ya podemos iniciar con la creación de las dimensiones, para ello primero creamos una correspondencia en el navegador de proyectos y haciendo clic derecho sobre el módulo de destino y luego hacemos clic derecho sobre dimensiones y seleccionamos nueva dimensión.

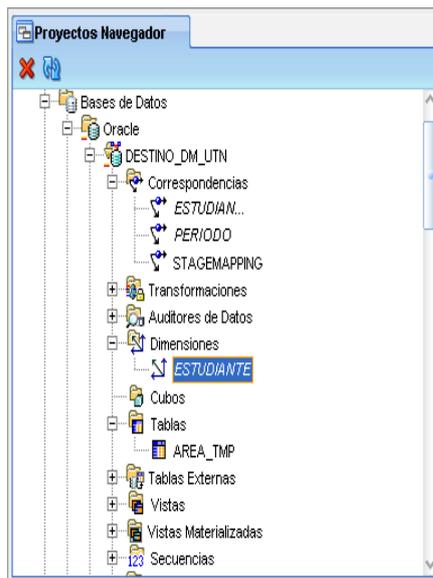


Figura 44: Creación de dimensiones

Fuente: Elaboración propia

De la misma forma se procede para la creación de las dimensiones faltantes Y finalmente crear los cubos que se requieren. Hay que señalar que todavía no hay nada físico, solo están creados los elementos a nivel lógico, para que puedan ser creados físicamente se debe desplegar cada elemento creado en el centro de control de warehouse builder. Este proceso se describe en detalle en el anexo C. Al terminar el trabajo con esta herramienta ETL el producto final se lo puede ver tal como se muestra en la figura 45.

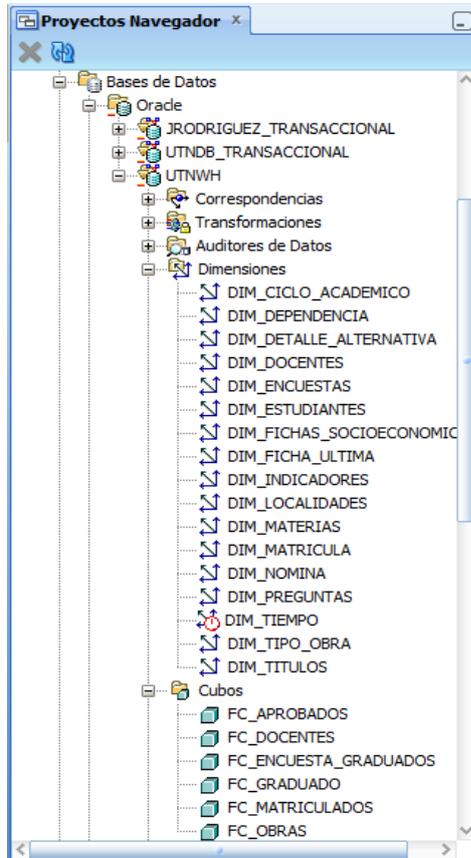


Figura 45: Árbol del proyecto

Fuente: Elaboración propia

4.5.1. Cubos Estudiantes y Docentes

De acuerdo a la elección de la granularidad realizada en secciones anteriores se diseñó el modelo dimensional en estrella que se presenta a continuación para cumplir con los requerimientos.

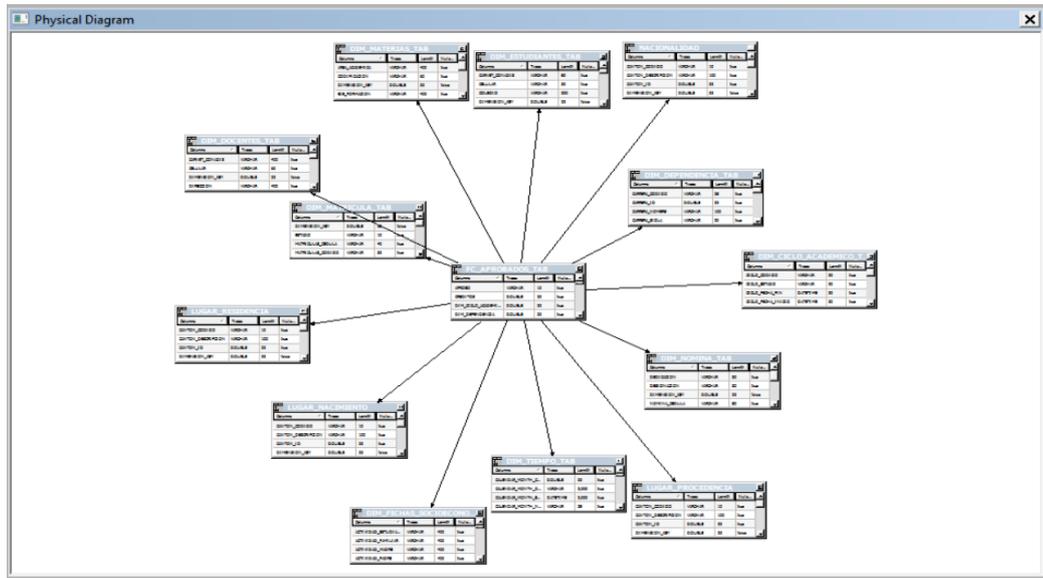


Figura 46: Esquema en estrella Aprobados

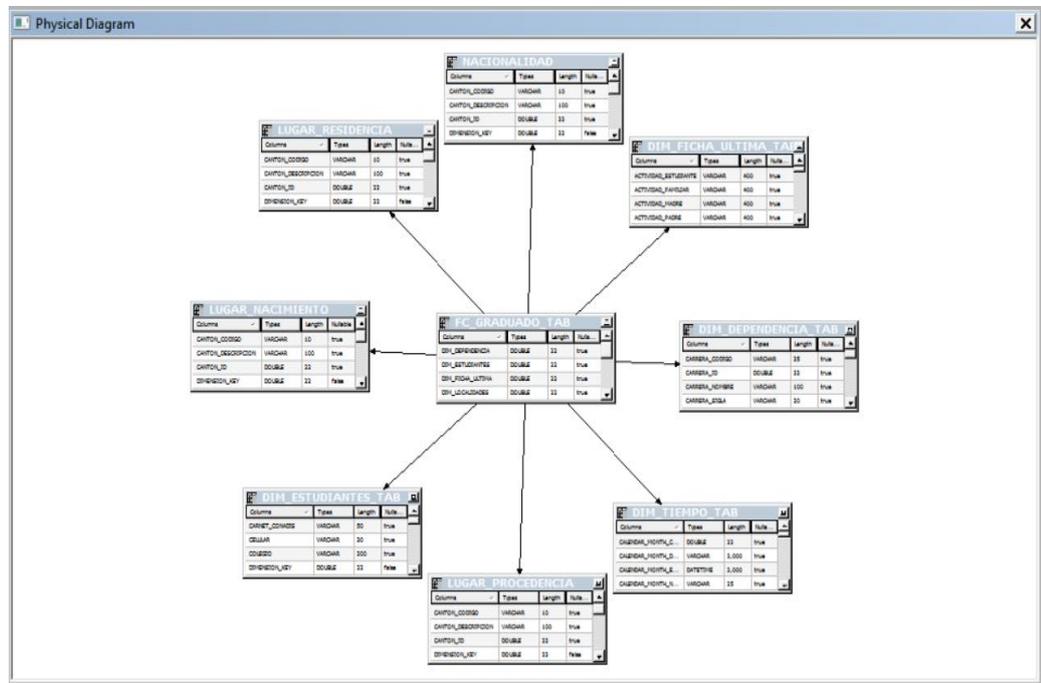


Figura 47: Esquema en estrella Graduados

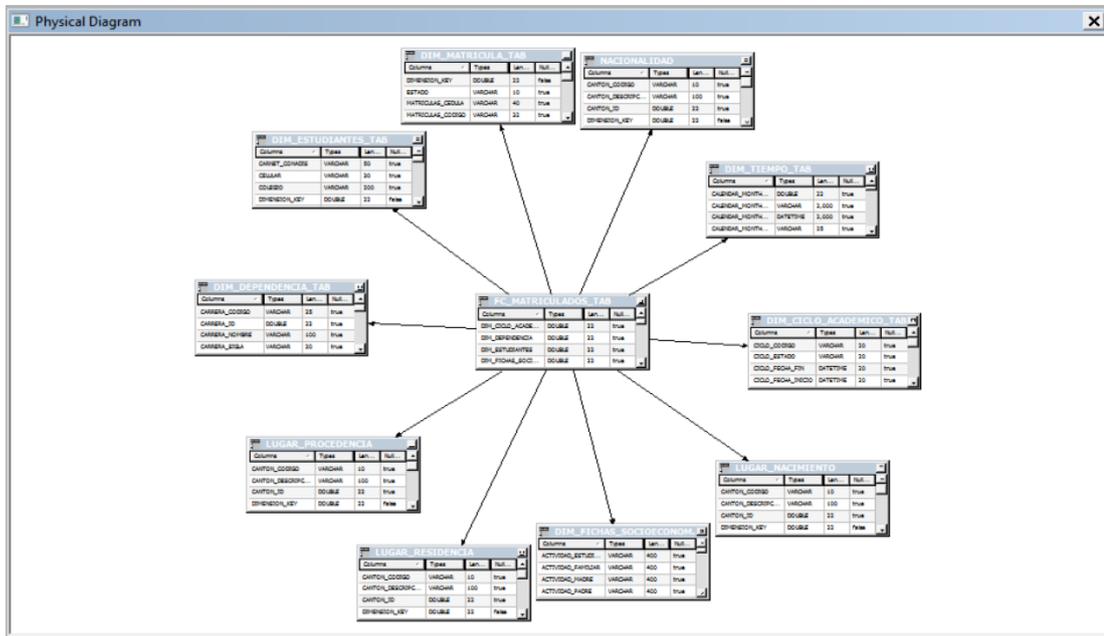


Figura 48: Esquema en estrella Matriculados

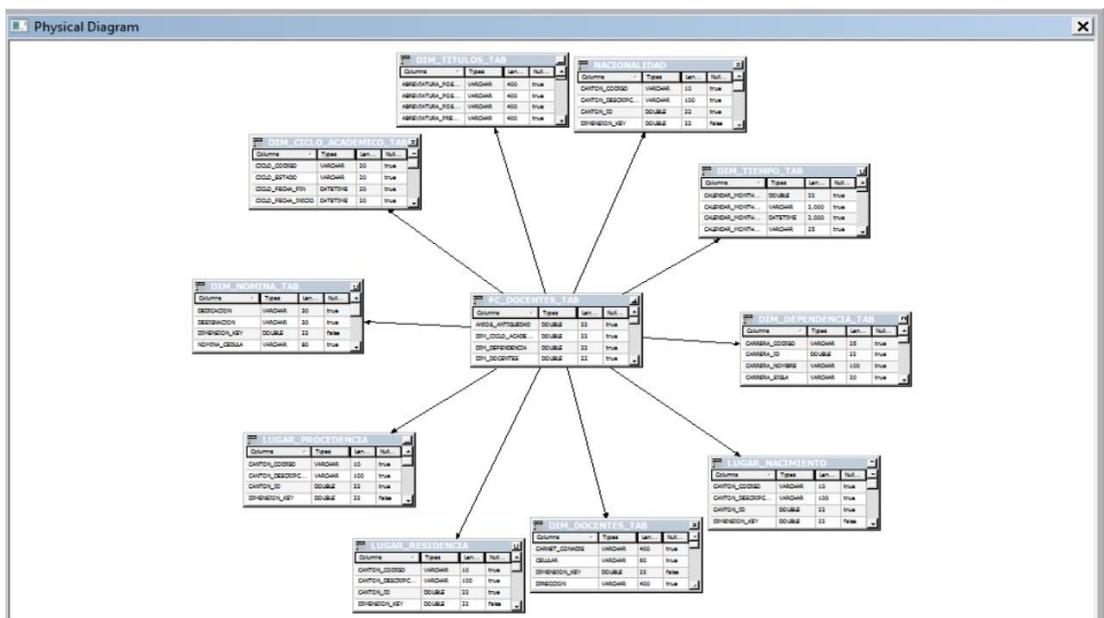


Figura 49: Esquema en estrella Docentes

4.6. Extracción, transformación y carga (ETL)

Para realizar los procesos ETL se han de crear correspondencias, que nos son otra cosa que objetos de datos enlazados en un cierto orden, esto determina el flujo que han de tomar los datos desde la fuente, datos transaccionales, hacia el destino o data marts y data warehouse.

En estas correspondencias se insertan objetos de datos fuente y destino además de objetos para realizar transformaciones necesarias antes de llenar el almacén de datos, estos disponen de una amplia gama de funciones de transformación para garantizar de esta forma la calidad de la información. Una vez concluido el diseño de los ETL's el árbol del proyecto se puede ver en la figura 50

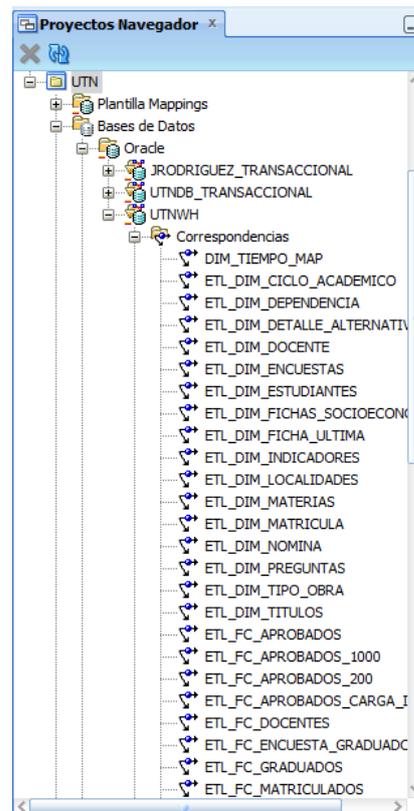


Figura 50: Procesos ETL

Fuente: Elaboración propia

Terminado el diseño de los procesos de extracción, transformación y carga es necesario automatizarlos, y para esto es preciso crear flujos de proceso que realizarán este trabajo de forma automática. Haciendo clic derecho en el nodo flujos de proceso iniciamos la creación al seleccionar luego Nuevo flujo de proceso, el producto terminado se muestra en la figura 51.

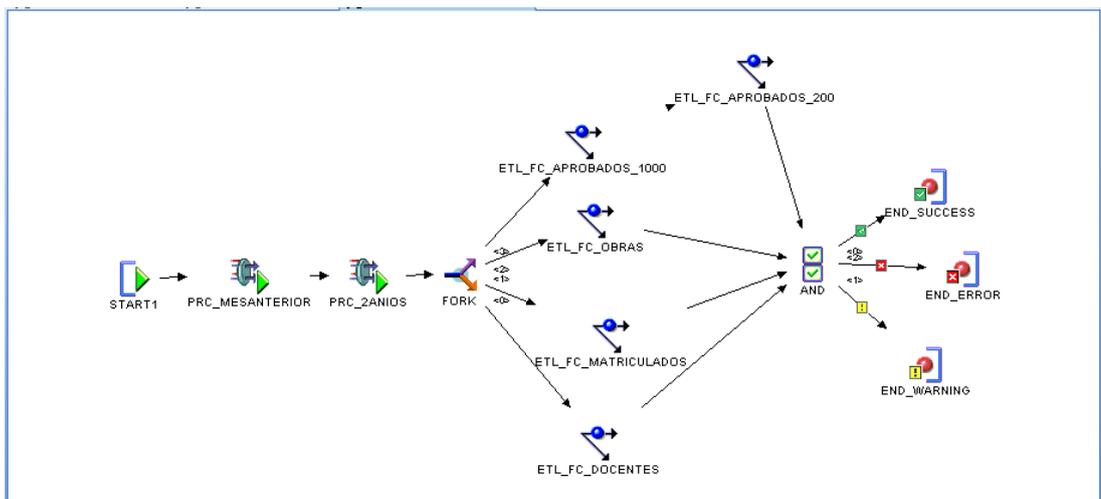


Figura 51: Flujo de Procesos
Fuente: Elaboración propia

4.7. Línea de aplicación de Inteligencia de Negocios

En esta parte estamos a punto de terminar la implementación del almacén de datos propuesto en secciones anteriores, para el efecto se usará la herramienta de administración de inteligencia de negocios con la que trasladaremos todo el diseño realizado hacia la plataforma de presentación de Oracle Business Intelligence.

4.7.1. Diseño del Business Intelligence (BI)

Una vez abierta la herramienta de administración se despliega en pantalla una ventana con tres secciones: Modelo físico, Modelo de Negocio y Modelo de presentación, en primer lugar se crearan las conexiones necesarias para obtener acceso al modelo creado

en warehouse builder, esto se la hará en la sección de Modelo físico y posteriormente se traslada este hacia la sección de Modelo del negocio en donde se seleccionan los datos que se han de presentar y como se lo hará, el producto final se muestra en la figura 52.

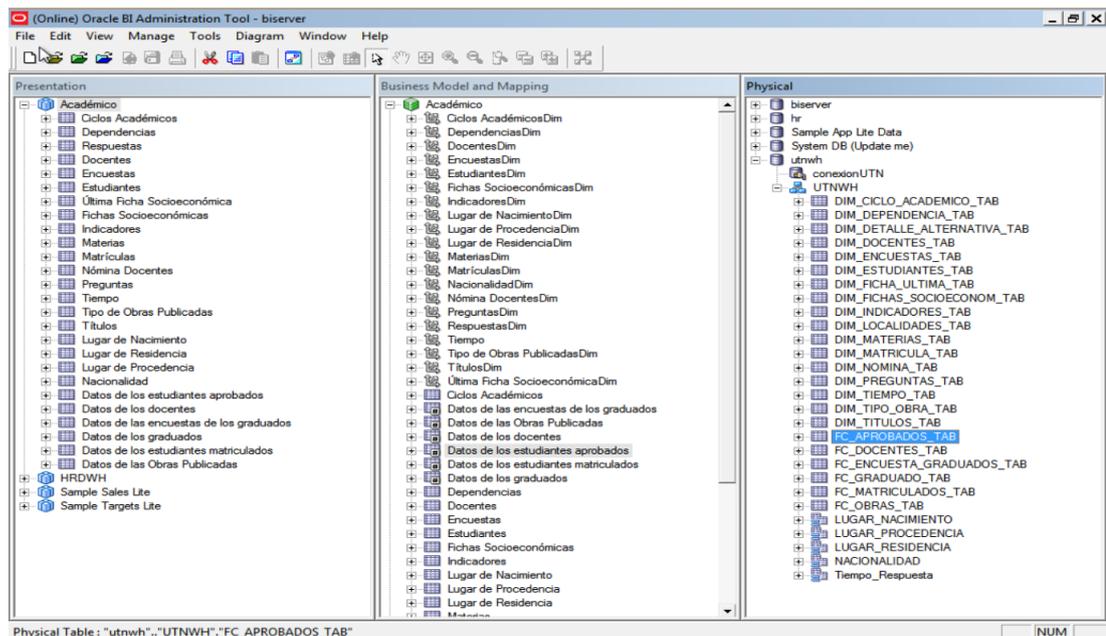


Figura 52: Herramienta de administración de Oracle Business Intelligence
Fuente: elaboración propia

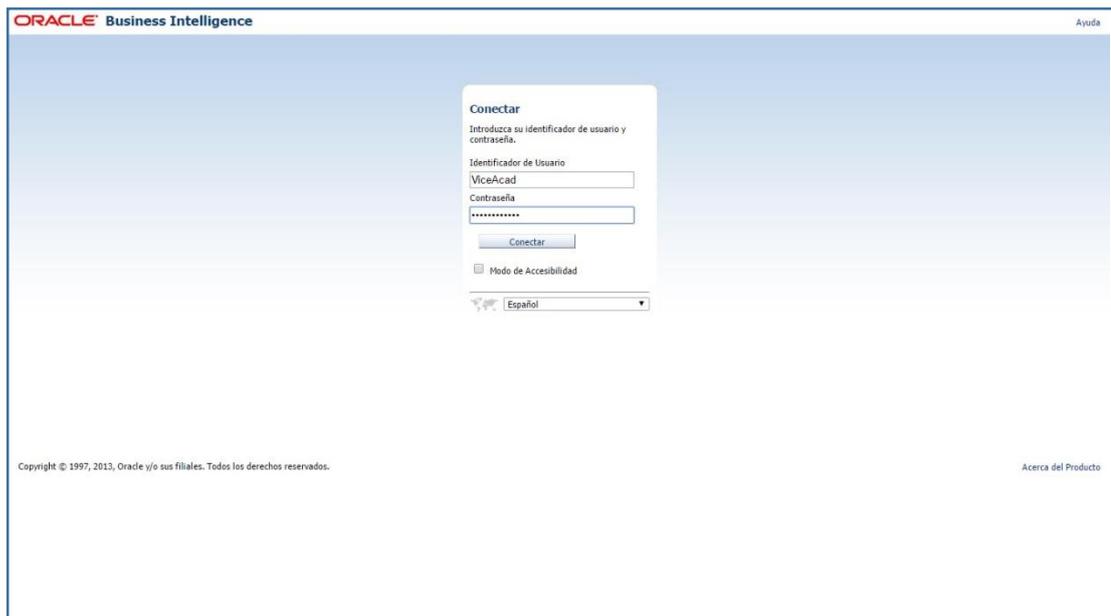
4.7.2. Desarrollo del BI

Para esta última parte se inicia la consola del weblogic y se recupera el archivo rpd generado previamente con la herramienta de administración del BI, se guarda los cambios. Ahora estamos listos para comenzar a desarrollar nuestras aplicaciones y reportes, para ello utilizar la herramienta analytics que se carga desde cualquier browser con la dirección <http://localhost:7001/analytics>.

Aquí podemos utilizar análisis, paneles de control, gráficos, etc. Para construir nuestras aplicaciones de Inteligencia de negocios. Este proceso inicia con la creación de objetos llamados análisis, que son diseñados de acuerdo a cada uno de los indicadores a satisfacer y que previamente se establecieron en la segunda fase de la

metodología. En estos objetos se puede especificar los criterios para seleccionar los datos a mostrar mediante sentencias SQL o filtros además de las opciones de formato. La información se puede organizar en tablas dinámicas acompañadas de gráficos muy variados, estos análisis pasarán a formar parte de páginas y estas a su vez formarán parte de los paneles de control, estructurando de esta manera los dashboard de nuestra solución.

Para ingresar al sistema se debe abrir un explorador de internet y escribir a continuación <http://localhost:7001/analytics> inmediatamente se muestra la ventana de inicio en el que se debe ingresar el nombre de usuario ViceAcad y su contraseña tal como se muestra en la figura 53.



*Figura 53: Ingreso al sistema
Fuente: elaboración propia*

Una vez ingresado al sistema se puede navegar por el panel de control para seleccionar el reporte que se desea mirar, seleccionando en la parte izquierda los parámetros que filtrará la información según los requerimientos del usuario en ese instante. A continuación se muestran algunos de esos reportes se muestran en las siguientes figuras.

Reporte que muestra el total de estudiantes matriculados por facultad, escuela y carrera, por género estado civil y etnia, filtrado desde el panel izquierdo de la ventana.

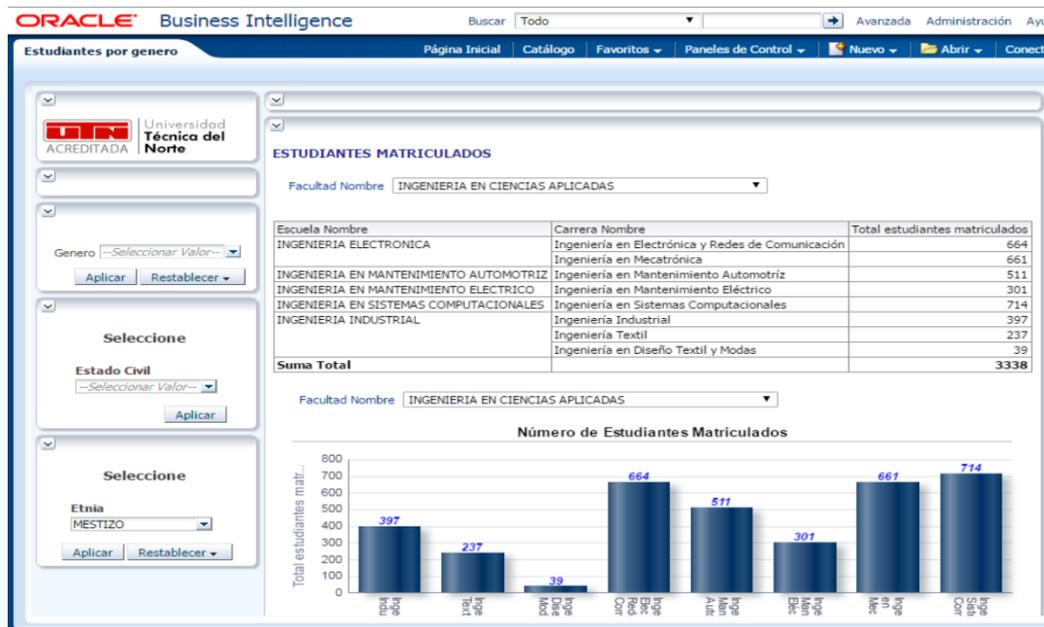


Figura 54: Estudiantes matriculados

Fuente: elaboración propia

Reporte para mostrar el total de docentes por facultad, escuela y carrera, se puede filtrar por ciclo académico, período, tipo de ciclo.

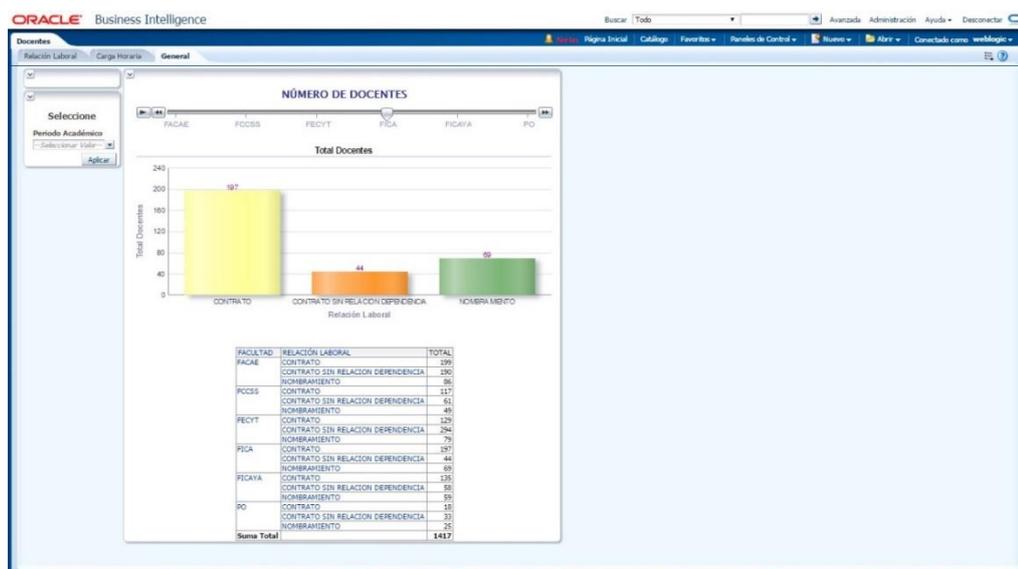


Figura 55: Reporte de estudiantes matriculados

Fuente: elaboración propia

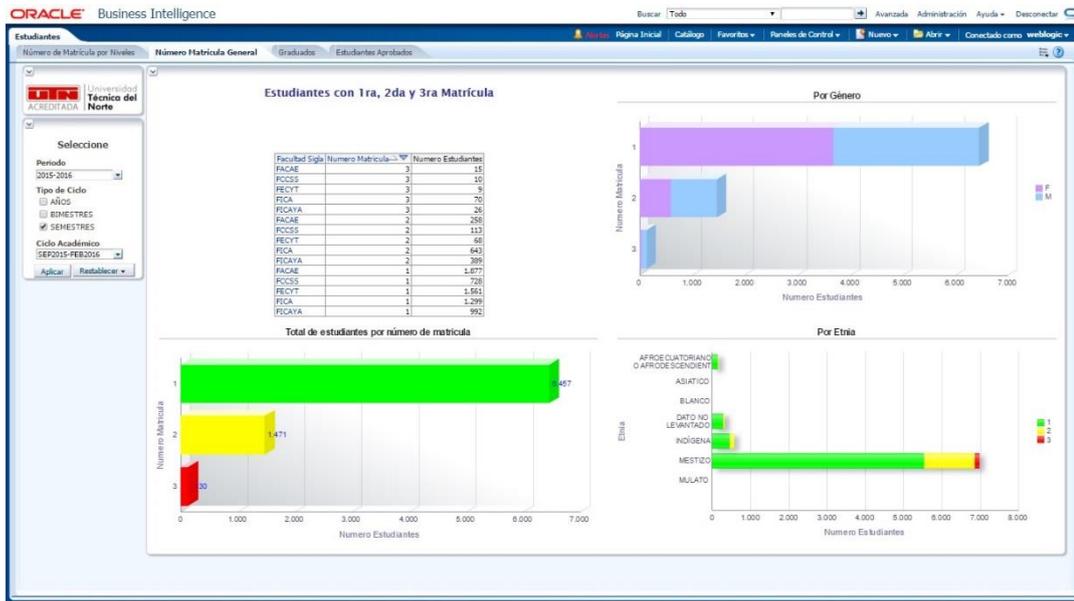


Figura 56: Estudiantes con 1ra, 2da y 3ra matrícula por género y etnia
Fuente: elaboración propia

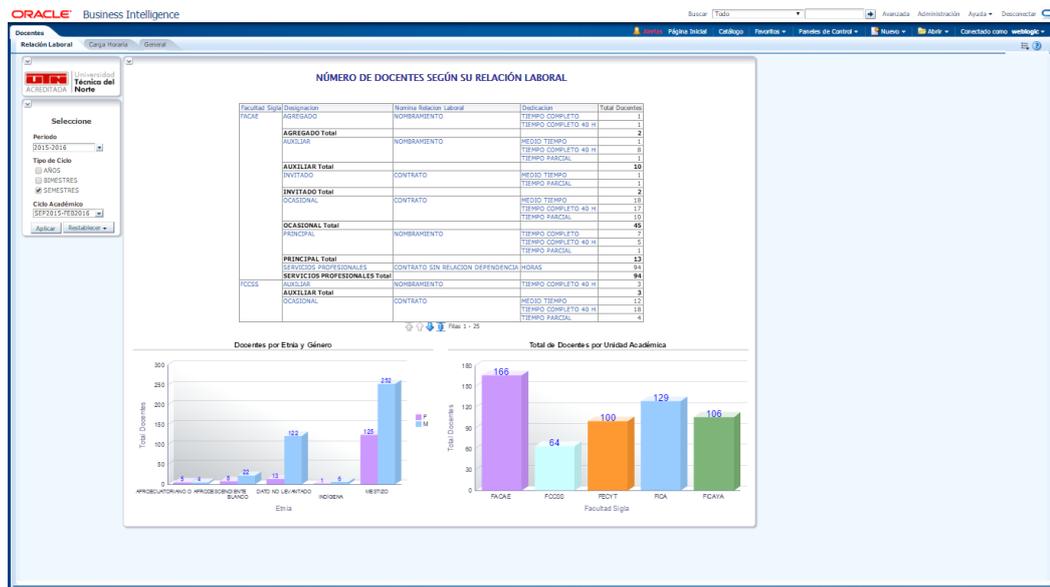


Figura 57: Docentes según su relación laboral
Fuente: elaboración propia

CAPÍTULO V

CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

- a) La toma de decisiones organizacionales tienen un soporte muy fuerte en las soluciones de inteligencia de negocios, éstas, ya no serán producto de la intuición y buen juicio de los ejecutivos.
- b) Las organizaciones, incluidas las educativas, pueden quedarse relegadas y perder su espacio, si no emprenden proyectos de inteligencia de negocios con el fin de liderar el competitivo ambiente de la Educación.
- c) Para la construcción de los almacenes de datos se utilizó el modelo dimensional en estrella, debido a su simplicidad y tiempo relativamente corto para su implementación y sobre todo porque se puede obtener resultados casi inmediatos.
- d) La metodología de construcción de Data warehouse y Data marts que más se acopla a entornos educativos es la de Ralph Kimball, porque permite empezar por un área específica de la organización y crecer paulatinamente hasta cubrir todas las áreas de la institución.
- e) Emprender proyectos de inteligencia de negocios con las herramientas de la Corporación Oracle tiene algunas ventajas, debido a que el sistema integrado de la UTN está desarrollado sobre esta plataforma y se dispone del licenciamiento.
- f) La presentación de la información a los usuarios mejora de manera sustancial, es más completa, entendible y oportuna, proporcionando una visión del estado de la actividad académica en la institución.

- g) No se pudo satisfacer completamente con el requerimiento acerca de los graduados y sus temas de tesis, debido a que esta información no está en la base de datos a pesar de que el sistema cuenta con un módulo para el registro de esta información, tan solo se cuenta con una referencia en el módulo de biblioteca.
- h) Los grandes volúmenes de datos almacenados que tiene la UTN, contienen información no visible a simple vista, sin embargo se pueden obtener con técnicas de minería de datos.

5.2. Recomendaciones

- a) Aprovechar el producto de este proyecto de investigación en futuras decisiones a tomarse en el área académica, apoyando con información para que las mismas sean las más acertadas y oportunas.
- b) Continuar con el diseño e implementación de los data marts para las restantes áreas de la universidad, lo que permitirá la construcción del data warehouse de toda la institución, mediante proyectos conjuntos entre el departamento de desarrollo tecnológico e informático y la carrera de ingeniería en sistemas computacionales.
- c) Incrementar en la malla de la carrera de ingeniería en sistemas computacionales la asignatura de inteligencia de negocios de manera oficial, no como está actualmente de forma optativa, que propicie ambientes de investigación en este campo.
- d) Incursionar en el área de minería de datos con la finalidad de encontrar patrones, tendencias y comportamientos del proceso de enseñanza aprendizaje, impulsando proyectos en este campo desde el departamento de desarrollo tecnológico e informático y la carrera de ingeniería en sistemas computacionales.

- e) Para futuros desarrollos de inteligencia de negocios seguir usando las herramientas de Oracle, y de esta manera evitar problemas de integración entre el software de diferentes marcas y versiones.
- f) Continuar aprovechando y explotando las ventajas de la herramienta de inteligencia de negocios de Oracle, y promover cursos de capacitación sobre el uso y aplicación de estas herramientas de software.
- g) Establecer los mecanismos necesarios para poder contar con la información necesaria completa acerca de los graduados y sus temas de tesis, y sea registrada en la base de datos de la universidad.

BIBLIOGRAFÍA

Inmon, W. H. (1992). Building the Data warehouse, Technical Publishing Group.

Kimball, R. (1996). The Data warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing, and Deploying Data warehouses New York, USA:Ed. John Wiley.

Inmon, W. H. (2005). Building the data warehouse. New York, USA:John wiley

Institute SAS, Inc. (2001). Rapid Warehouse Methology, White Paper.

Kimball, R. (1998). The Data warehouse Lifecycle Tool Kit, 1998, New York, USA:Ed. John Wiley,

Araníbar, J. C., (2003). Inteligencia de Negocios. *Revista Ciencia y Cultura* , n. 12
95-101.

Peña, A., (2006). Inteligencia de Negocios: Una Propuesta para su Desarrollo en las organizaciones. Instituto Politécnico Nacional. México D. F.

Calzada, L. y Abreu, J. L., (2009). Daena: International Journal of Good Conscience.
4(2), 16-52.

Soto, B. M., & Barrios, F. N. (2006). Gestión del conocimiento. Parte I: revisión crítica del estado del arte. Cuba: Scielo Cuba. <http://www.ebrary.com/>

Pastor, J. A., (2002). Usos de Los Sistemas de Información en la Organización.
Editor Editorial UOC, ISBN 8484296784, 9788484296782

Pueyo, D.C., Ortiz, M. y Amo, M.O., (2010). Sistemas de información en la organización: casos prácticos UOC, Universitat Oberta de Catalunya. Recuperado de <https://books.google.com.ec/books?id=aZxHMwEACAAJ>

Heredero, C., López J. J., Romo, S. M. y Medina, S. (2004). Informática y comunicaciones en la empresa. Madrid, España: Esic Editorial.

Fernández, C., (2009). DataPrix: Las Suites de Oracle Business Intelligence <http://www.dataprix.com/las-suites-business-intelligence-oracle>

Oracle, (2013). Oracle Business Intelligence Publisher 11g. Oracle Data Sheet. Recuperado de <http://www.oracle.com/technetwork/middleware/bi-publisher/overview/bipublisherdatasheet-129370.pdf>

Oracle, (2009). Descripción del producto Oracle. Oracle Data Sheet. Recuperado de <http://www.oracle.com/es/solutions/midsize/090114-pcv-dbse1-product-459590-es.pdf>

Oracle, (2009). Oracle Database 11g Standard Edition One. Oracle Data Sheet. Recuperado de <http://www.oracle.com/technetwork/database/database-11g-standard-edition-one-d-132019.pdf>

Oracle, (2013). Oracle Database Data Warehousing Guide, 11g Release 2. Help Center Recuperado de https://docs.oracle.com/cd/E11882_01/server.112/e25554/

Sieber, S., Valor, J., y Porta, V. (2006). Los sistemas de información en la empresa actual: aspectos estratégicos y alternativas tácticas. España: McGraw-Hill. Recuperado de <http://www.ebrary.com/>

Curto, D. J., y Curto, D. J. (2010). Introducción al Business Intelligence. España: Editorial UOC. Recuperado de <http://www.ebrary.com/>

Donohue M., (2010). Oracle Business Intelligence Standard Edition One 11g

Rodríguez, S., (2010), *Análisis Y Diseño De Un Data mart Para El Seguimiento Académico De Alumnos En Un Entorno Universitario*. (Tesis de Posgrado). Universidad Carlos III de Madrid Escuela Politécnica Superior Ingeniería en Informática, España, Madrid.

Reinosa, E. J., Maldonado, C. A., y Muñoz, R. (2012). Bases de datos. México: Alfaomega Grupo Editor. Recuperado de <http://www.ebrary.com/>

Trujillo, J. C. (2013). Diseño y explotación de almacenes de datos: conceptos básicos de modelado multidimensional. España: ECU. Recuperado de <http://www.ebrary.com/>

Nader, J. (2003). *Sistema de Apoyo Gerencial Universitario*. (Tesis de Maestría). Instituto Tecnológico de Buenos Aires, Argentina, Buenos Aires.

