UNIVERSIDAD AMAZÓNICA DE PANDO

DIRECCIÓN DE POSTGRADO E INVESTIGACIÓN



TESIS PARA OPTAR EL GRADO ACADÉMICO DE MAGISTER EN INGENIERÍA INFORMÁTICA

"Implementación de un data warehouse para el padrón electoral departamental de Potosí"

POSTULANTE: LIC. JUAN JOSÉ CERVANTES AVILES

TUTOR: LIC. RAMIRO VILLA

Potosí – Bolivia

2012

DEDICADO:

A mi hija Anel, la cual alegra los días de mi vida la misma que con su cariño me da la fuerza necesaria para estar de pie y con la cabeza en alto para enfrenta cualquier situación por difícil que parezca.

AGRADECIMIENTOS:

Agradezco primero a Dios por iluminarme en cada momento en cada triunfo y caída.

Mil palabras no bastarían para agradecerles su apoyo, su comprensión y sus consejos en los momentos difíciles a:

Sariel, espero no defraudarla y contar siempre con su valioso apoyo, sincero e incondicional.

Mis padres y hermanos gracias por haber fomentado en mí el deseo de superación y el deseo de triunfar en la vida, sigo alcanzando metas, ya que siempre estuvieron impulsándome en los momentos más difíciles de mi vida, y porque el orgullo que sienten por mí, me inspiro a llegar mucho más lejos. Va por ustedes, por lo que valen, porque admiro su fortaleza y por lo que han hecho de mí.

ÍNDICE DE CONTENIDO

Pág.

INTRODUCCIÓN.....1

CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA, OBJETIVOS Y ESTRATEGIA METODOLÓGICA. 3

1.1. ANTECEDENTES.....3

1.2. PROBLEMA DE INVESTIGACIÓN. 3

1.3. OBJETO DE ESTUDIO. 4

1.4. CAMPO DE ACCIÓN. 4

1.5. OBJETIVOS.....4

1.5.1. Objetivo General. 4

1.5.2. Objetivos Específicos.....4

1.6. HIPÓTESIS.....4

1.6.1. Operacionalización de Variables. 5

1.7. JUSTIFICACIÓN.....5

1.8. APORTE TEÓRICO. 5

1.9. APORTE SOCIAL. 5

1.10. APORTE METODOLÓGICO. 6

1.11. MÉTODOS DE INVESTIGACIÓN. 6

1.11.1. Métodos a Nivel Empírico 6

1.11.1.1. Entrevista a Vocales y Directores.....6

1.11.1.2. Criterio o Consulta a Personal Operativo. 6

1.11.2. Métodos a Nivel Teórico 6

1.11.2.1. Estudio Documental. 6

1.11.2.2. Análisis y Síntesis 7

1.11.2.3. Inducción y Deducción. 7

1.11.2.4. Histórico – Lógico. 7

1.11.2.5. Enfoque Sistémico. 7

CAPÍTULO II: MARCO TEÓRICO.....8

2.1. BUSINESS INTELLIGENCE......8

2.1.1. Introducción. 8

2.1.2. Definición. 8

2.1.3. Importancia y Beneficios de Business Intelligence en la Organización. 8

2.1.4. Arquitectura de Business Intelligence 10

2.1.5. Etapas de Business Intelligence.....11

2.1.5.1. Dirigir y Planear.....11

2.1.5.2. Identificación de Información 13

2.1.5.3. Procesamiento de Datos 14

2.1.5.4. Análisis y Producción.....14

2.2. DATA WAREHOUSE. 15

2.2.1. Introducción......15

2.2.2. Características 15

2.2.3. Arquitectura De Un Data Warehouse. 16

2.2.3.1. Base de Datos Operacional y Base de Datos Externa 17

2.2.3.2. Nivel de Acceso a la Información 17

2.2.3.3. Nivel de Acceso a los Datos 17

2.2.3.4. Nivel de Directorio de Datos (Metadatos).....18

2.2.3.5. Nivel de Gestión de Procesos 18

2.2.3.6. Nivel de Mensaje de la Aplicación.....18

2.2.3.7. Nivel Data Warehouse (Físico) 18

2.2.3.8. Nivel de Organización de Datos 19

2.2.4. Estructura de un Data Warehouse. 19

2.2.5. Proceso de Extracción, Transformación y Carga.....20

2.2.5.1. Extracción. 22

2.2.5.2. Transformación. 23

2.2.5.3. Carga. 24

2.2.6. Modelo Dimensional de un Data Warehouse 24

2.2.6.1. Esquema Estrella (Star Schema)......25

2.2.6.2. Esquema Copo de Nieve (Snowflake Schema) 26
2.2.6.3. Tablas 26
2.2.6.4. Nivel. 28
2.2.6.5. Jerarquía. 29
2.2.6.6. Dimensiones Lentamente Cambiantes (SCD). 30
2.2.6.7. Granularidad. 31
2.2.6.8. Relaciones. 31
2.2.6.9. Llaves Subrogadas. 31
2.2.7. Explotación del Data Warehouse. 31
2.2.7.1. Reportes y Consultas. 32
2.2.7.2. Minería de Datos (Data Mining). 32
2.2.7.3. Análisis Olap. 33
2.3. METODOLOGÍA RALPH KIMBALL. 36
2.3.1. Planificación del proyecto. 37
2.3.2. Definición de los Requerimientos del Negocio. 40
2.3.3. Diseño Técnico de la Arquitectura. 42
2.3.3.1. Back Room. 42
2.3.3.2. Front Room. 42
2.3.4. Selección e Instalación del Producto. 43
2.3.5. Modelamiento Dimensional. 43
2.3.5.1. Cuatro Pasos Para el Proceso de Diseño Dimensional. 44
2.3.5.2. Data Warehouse Bus Matrix 45
2.3.6. Diseño Físico. 45
2.3.7. Diseño y Desarrollo del Proceso ETL. 46
2.3.7.1. Paso 1 – Trazar un Plan de Alto Nivel. 46
2.3.7.2. Paso 2 – Seleccionar la Herramienta ETL. 46
2.3.7.3. Paso 3 – Desarrollar Estrategias por Defecto. 46
2.3.7.4. Paso 4 – Profundizar la Tabla Objetivo. 47
2.3.7.5. Paso 5 – Poblar las Tablas de Dimensiones con los Datos Históricos. 47
2.3.7.6. Paso 6 – Realizar la Carga Histórica de la Tabla de Hechos. 47

2.3.7.7. Paso 7 – Procesamiento Incremental de la Dimensiones. 47
2.3.7.8. Paso 8 – Procesamiento Incremental de los Hechos. 47
2.3.7.9. Paso 9 – Operación y Automatización del Sistema ETL. 48
2.3.8. Diseño y Desarrollo de Aplicación de BI. 48
2.3.8.1. Tipo de Aplicación de BI. 48
2.3.8.2. Preparar el Desarrollo de la Aplicación. 49
2.3.8.3. Crear las Aplicaciones 49
2.3.9. Despliegue 49
2.3.9.1. Despliegue ETL 49
2.3.9.2. Despliegue de BI 49
2.3.10. Mantenimiento y Crecimiento 50
2.4. HERRAMIENTAS PARA EL DESARROLLO DE DATA WAREHOUSE. 50
2.4.1. SQL Server 2008. 50
2.4.2. SQL Server Integration Services (SSIS). 51
2.4.3. SQL Server Analysis Services (SSAS) 51
2.4.4. SQL Server Reporting Service (SSRS) 53
2.4.5. Microsoft Office Excel 2010 53
CAPITULO III: DESARROLLO E IMPLEMENTACIÓN DEL DATA WAREHOUSE PARA EL PADRÓN ELECTORAL DEPARTAMENTAL DE POTOSÍ 55
3.1. PLANIFICACIÓN DEL PROYECTO. 55
3.1.1. Definición del Proyecto. 55
3.1.2. Alcance y Justificación del Proyecto. 55
3.2. DEFINICIÓN DE LOS REQUERIMIENTOS DEL NEGOCIO. 56
3.2.1. Estadísticas. 56
3.3. DISEÑO TÉCNICO DE LA ARQUITECTURA. 57
3.3.1. Back Room. 57
3.3.2. Front Room. 59

3.4. SELECCIÓN DEL PRODUCTO. 59
3.4.1. Plataforma DBMS. 59
3.4.2. Herramienta ETL. 60
3.4.3. Herramienta de BI. 60
3.5. MODELAMIENTO DIMENSIONAL. 60
3.5.1. Paso1. Selección del Proceso de Negocio. 60
3.5.2. Paso2. Definición de la Granularidad. 60
3.5.3. Paso3. Identificar las Dimensiones y Mapear los Datos Origen. 60
3.5.3.1. Dimensión Tiempo. 61
3.5.3.2. Dimensión Geografía. 61
3.5.3.3. Dimensión Grado de Instrucción. 62
3.5.3.4. Dimensión Estado Civil. 63
3.5.3.5. Dimensión Sexo. 63
3.5.3.6. Dimensión Tipo Cedula. 64
3.5.4. Paso4. Identificar los Hechos y Mapear los Datos Origen. 64
3.5.4.1. Ciudadano. 65
3.5.5. Data Warehouse Bus Matrix 65
3.6. DISEÑO FÍSICO. 66
3.7. DISEÑO Y DESARROLLO DEL PROCESO ETL. 66
3.7.1. Paso1. Trazar un Plan de Alto Nivel. 66
3.7.1.1. Geografía. 66
3.7.1.2. Estado Civil. 67
3.7.1.3. Grado Instrucción. 67
3.7.1.4. Sexo. 68
3.7.1.5. Tipo Cedula. 68
3.7.1.6. Ciudadano. 69
3.7.2. Paso2. Selección de la Herramienta ETL. 69
3.7.3. Paso3. Desarrollar Estrategias por Defecto. 69
3.7.4. Paso4. Profundizar la Tabla Objetivo. 70
3.7.5. Paso5. Poblar las Tablas de Dimensiones Con los Datos Históricos. 71

3.7.6. Paso6. Realizar la Carga Histórica de la Tabla de Hechos. ... 71

3.7.7. Paso7. Procesamiento Incremental de las Dimensiones. ... 72

3.7.8. Paso8. Procesamiento Incremental de los Hechos. ... 72

3.7.9. Paso9. Operación y Automatización del Sistema ETL. ... 72

3.8. DISEÑO Y DESARROLLO DE LA APLICACIÓN DE BI. ... 73

3.8.1. Tipo de Aplicación de BI. ... 73

3.8.2. Preparar el Desarrollo de la Aplicación. ... 73

3.8.3. Crear la Aplicación. ... 73

3.8.3.1. Desarrollo de la Aplicación BI con SSAS. ... 73

3.9. DESPLIEGUE. ... 77

3.9.1. Despliegue ETL. ... 78

3.9.2. Despliegue BI. ... 78

3.10. MANTENIMIENTO Y CRECIMIENTO. ... 80

CAPITULO IV: RESULTADOS ... 81

4.1. TIEMPO DE ELABORACIÓN DE INFORMES. ... 81

4.2. SATISFACCIÓN DEL USUARIO ESTRATÉGICO. ... 83

4.2.1. En una Escala de 1 a 5 cuál es el nivel de satisfacción con los reportes y vistas OLAP. ... 83

4.2.2. ¿Cómo considera el tiempo de respuesta en la obtención de las consultas de los informes (Bajo, Medio y Alto). ... 84

4.2.3. ¿La información que obtiene al consultar los informes son los correctos? ... 85

4.2.4. ¿Considera fácil el uso del Data Warehouse? ... 86

4.2.5. ¿Para obtener informes usted prefiere realizar consultas o utilizar el Data Warehouse? ... 87

CONCLUSIONES Y RECOMENDACIONES

BIBLIOGRAFÍA

REFERENCIAS BIBLIOGRÁFICAS

ANEXOS

ÍNDICE DE FIGURAS

FIGURA 2.1: OBJETIVOS BI 9

FIGURA 2.2: ARQUITECTURA BI 10

FIGURA 2.3: PROCESO BI [2] 11

FIGURA 2.4: ARQUITECTURA DE UN DATA WAREHOUSE 17

FIGURA 2.5: ESTRUCTURA DE UN DATA WAREHOUSE 19

FIGURA 2.6: PROCESO ETL 21

FIGURA 2.7: SUBPROCESO ETL – EXTRACCIÓN 22

FIGURA 2.8: SUBPROCESO ETL – TRANSFORMACIÓN 23

FIGURA 2.9: SUBPROCESO ETL – CARGA 24

FIGURA 2.10: ESQUEMA ESTRELLA (STAR SCHEMA) 25

FIGURA 2.11: ESQUEMA COPO DE NIEVE (SNOWFLAKE SCHEMA) 26

FIGURA 2.12: TABLA DE HECHO 27

FIGURA 2.13: TABLAS DE DIMENSIONES 28

FIGURA 2.14: NIVELES EN LA DIMENSIÓN. 29

FIGURA 2.15: JERARQUÍA EN LA DIMENSIÓN. 30

FIGURA 2.16: OLAP (ON-LINE ANALYTICAL PROCESSING) 34

FIGURA 2.17: CUBO OLAP 35

FIGURA 2.18: THE KIMBALL LYFECYCLE DIAGRAM 37

FIGURA 2.19: DATA WAREHOUSE BUS MATRIX 45

FIGURA 3.1: BACK ROOM – PROYECTO 58

FIGURA 3.2: FRONT ROOM – PROYECTO 59

FIGURA 3.3: TABLA DE DIMENSIÓN TIEMPO 61

FIGURA 3.4: TABLA DE DIMENSIÓN GEOGRAFÍA 62

FIGURA 3.5: TABLA DE DIMENSIÓN GRADO DE INSTRUCCIÓN 63

FIGURA 3.6: TABLA DE DIMENSIÓN ESTADO CIVIL 63

FIGURA 3.7: TABLA DE DIMENSIÓN SEXO 64

FIGURA 3.8: TABLA DE DIMENSIÓN TIPO CEDULA 64
FIGURA 3.9: TABLA DE HECHOS CIUDADANO 65
FIGURA 3.10: DATA WAREHOUSE BUS MATRIX DETALLADO 65
FIGURA 3.11: PLAN ETL GEOGRAFÍA 66
FIGURA 3.12: PLAN ETL ESTADO CIVIL 67
FIGURA 3.13: PLAN ETL GRADO DE INSTRUCCIÓN 67
FIGURA 3.14: PLAN ETL SEXO 68
FIGURA 3.15: PLAN ETL TIPO DE CEDULA 68
FIGURA 3.16: PLAN ETL CIUDADANO 69
FIGURA 3.17: RELACIÓN ENTRE LA TABLA DE DIMENSIONES Y DE HECHOS 70
FIGURA 3.18: JERARQUÍA TABLA DE DIMENSIÓN DE GEOGRAFÍA 71
FIGURA 3.19: JERARQUÍA TABLA DE DIMENSIÓN DE TIEMPO 71
FIGURA 3.20: CONFIGURACIÓN DE ORIGEN DE DATOS 74
FIGURA 3.21: VISTA DE ORÍGENES DE DATOS CREADOS 75
FIGURA 3.22: DIMENSIONES AGREGADAS 75
FIGURA 3.23: DIMENSIÓN DE GEOGRAFÍA CON SU JERARQUÍA ESPECÍFICA 76
FIGURA 3.24: DIMENSIÓN DE TIEMPO CON SU JERARQUÍA ESPECÍFICA 76
FIGURA 3.25: CUBO DEL NEGOCIO GENERADO 77
FIGURA 3.26: REPORTE GENERADO CON REPORTING SERVICE 79
FIGURA 3.27: REPORTE GENERADO CON EXCEL 2010 80
FIGURA 4.1: COMPARACIÓN DEL TIEMPO DE ELABORACIÓN DE INFORMES SISTEMA TRANSACCIONAL VS DATA WAREHOUSE 82
FIGURA 4.2: TIEMPO DE ELABORACIÓN DE INFORMES SISTEMA TRANSACCIONAL VS DATA WAREHOUSE 82
FIGURA 4.3: SATISFACCIÓN CON LOS REPORTES Y VISTAS OLAP 83
FIGURA 4.4: PORCENTAJE DE SATISFACCIÓN CON LOS REPORTES Y VISTAS OLAP 84
FIGURA 4.5: TIEMPO DE RESPUESTA EN LA OBTENCIÓN DE LAS CONSULTAS 84
FIGURA 4.6: PORCENTAJE TIEMPO DE RESPUESTA EN LA OBTENCIÓN DE LAS CONSULTAS 85
FIGURA 4.7: LOS INFORMES SON LOS CORRECTOS 85

FIGURA 4.8: QUE PORCENTAJE DE LOS INFORMES SON LOS CORRECTOS 86

FIGURA 4.9: CONSIDERA FÁCIL EL USO DEL DATA WAREHOUSE 86

FIGURA 4.10: EL PORCENTAJE DE LA INFORMACIÓN QUE CONSIDERA FÁCIL EL USO DEL DATA WAREHOUSE 87

FIGURA 4.11: PARA OBTENER INFORMES USTED PREFIERE REALIZAR CONSULTAS O UTILIZAR EL DATA WAREHOUSE 87

FIGURA 4.12: PORCENTAJE QUE PREFIERE UTILIZAR EL DATA WAREHOUSE 88

RESUMEN

El presente proyecto está orientado al soporte a la toma de decisiones mediante la generación de información analítica la cual será presentada de manera oportuna en el Tribunal Electoral Departamental de Potosí.

Para la metodología de desarrollo se estableció usar la metodología de Ralph Kimball debido a que la misma permite el desarrollo durante todo el ciclo de vida empezando desde la planificación del proyecto hasta la finalización de la misma.

La propuesta fue revisada mediante una encuesta que se realizó a directores y vocales del Tribunal Electoral Departamental de Potosí, los cuales dieron pleno consentimiento en el desarrollo de un Data Warehouse para el Padrón Electoral Departamental de Potosí.
La propuesta considero las siguientes actividades:

- Implementar un Data Warehouse para generar información analítica y oportuna.
- Generación de reportes en base a encuestas realizadas a directores y vocales del Tribunal Electoral Departamental de Potosí.

Palabras clave: información oportuna, Data Warehouse, Generación de Información.

INTRODUCCIÓN

En la actualidad el Tribunal Electoral Departamental(TED) de Potosí es una institución encargada de llevar a cabo los procesos electorales nacionales o departamentales del departamento, para ello cuenta con una Base de Datos denominada Padrón Electoral el cual almacena la información de las personas que se encuentran habilitadas o no para emitir el voto.

El Tribunal Electoral Departamental de Potosí cuenta con varios departamentos los cuales contribuyen a la institución, en ellas se maneja información que ayuda a la toma de decisiones por parte de los Vocales y Directores de cada área.

Para el apoyo de toma de decisiones los vocales y directores, requieren Información de sus distintas áreas y éstas a su vez del Departamento de Informática de la Institución, lo cual es un proceso complejo para el sencillo efecto de contar con información precisa y resumida; este proceso produce pérdida de tiempo, información no oportuna, y la falta de disponibilidad del Departamento de Informática para apoyar a los Vocales y Directores que requieren la información para soportar los procesos de la institución.

Durante muchos años el Tribunal Electoral Departamental de Potosí ha acumulado la información que las áreas operacionales generan al resolver las transacciones que hacen posible su gestión cotidiana. Esta información es sumamente valiosa a la hora de tomar decisiones, que necesitan conocimientos sobre datos históricos y tendencias.
El caso típico de una persona buscando información estratégica implica que luego de obtener cierta información, usará el resultado obtenido para dirigir su próxima consulta, por lo que se debe contar con una herramienta que permita esa interacción.

El uso de herramientas tradicionales para obtener información estratégica para la toma decisiones presenta inconvenientes de diferente índole, como: tiempos de respuesta extremadamente lentos, dificultad o imposibilidad de expresar una consulta compleja, e incapacidad de obtención de resultados que recopilen información de fuentes diversas.

Las planillas electrónicas si bien son muy amigables para el usuario final, presentan problemas de eficiencia y no son confiables para grandes volúmenes de información.

Una solución encontrada a estos inconvenientes en los últimos años ha sido la construcción de Sistemas de Data Warehouse que almacenan la información en estructuras especiales, apropiadas para favorecer la eficiencia en las operaciones de consulta, y que con herramientas adecuadas de extracción de datos, ofrecen un efectivo apoyo en la toma de decisiones.

La presente tesis está constituida por 4 Capítulos; el Capítulo I está constituido por el planteamiento del problema, los objetivos y las metodologías.

En el Capítulo II está dedicado al Marco Teórico, en el cual se estudia Business Intelligence, Data Warehouse, el desarrollo del Data Warehouse basado en la Metodología de Ralph Kimball, así mismo un estudio completo de la herramienta de desarrollo de Data Warehouse de Microsoft Sql Server 2008.

En el Capítulo III se realiza el desarrollo e implementación del Data Warehouse para el Padrón Electoral Departamental de Potosí en base a la metodología de Ralph Kimball.

En el Capítulo IV se desarrolló los resultados donde se muestra la conformidad de la utilización del Data Warehouse

El final está dedicado a las conclusiones, recomendaciones y anexos.

CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA, OBJETIVOS Y ESTRATEGIA METODOLÓGICA.

1.1. Antecedentes.

Desde que empezó a introducir los ordenadores para automatizar la gestión de las organizaciones en la década de los sesenta, la evolución de los sistemas de información ha tenido una considerable repercusión en la gestión de datos, las tecnologías de información y comunicación constituyen la infraestructura fundamental de grandes organizaciones, permitiendo el registro de múltiples datos.

Las bases de datos posibilitan almacenar cada operación así como otros muchos elementos que refleja la organización. Es necesario convertir grandes volúmenes de datos en conocimiento para que sea útil a la toma de decisiones.

El desarrollo de nuevas técnicas para encontrar la información requerida a partir de grandes cantidades de datos es uno de los principales desafíos de esta época. La idea que emerge de todo esto es que el resultado final sea que esta tecnología, llegue a formar parte de las aplicaciones que ayuden a la toma de decisiones.

La implementación de una solución de Data Warehouse con la información analítica y estratégica que proporcionan todas las unidades del Tribunal Electoral Departamental de Potosí a directores y vocales, cuya información se encuentra almacenada en bases de datos, garantiza aspectos de accesibilidad, integración de tecnología y toma de decisiones.

1.2. Problema de Investigación.

Se plantea el siguiente ***problema de investigación****:* ¿Cómo implementar un Data Warehouse para generar información a tiempo en diferentes niveles Operativos, Tácticos, Gerenciales además de apoyar a la toma de decisiones en el Tribunal Electoral Departamental del Departamento de Potosí.?

1.3. Objeto de Estudio.

Como ***Objeto de estudio*** se tiene: Business Intelligence

1.4. Campo de Acción.

Como ***Campo de acción*** el estudio de Data Warehouse aplicado al Padrón Electoral Biométrico del Tribunal Electoral Departamental de Potosí.

1.5. Objetivos.

1.5.1. Objetivo General.

Para dar respuesta al problema formulado se plantea el siguiente ***Objetivo general:*** "Implementar un Data Warehouse para apoyar a la toma de decisiones en el Tribunal Electoral Departamental de Potosí obteniendo información a tiempo para los procesos electorales. Y de esta forma coadyuvar con la obtención rápida y oportuna de la información el mismo que ayudara a una toma de decisiones mucho más rápida a nivel gerencial."

1.5.2. Objetivos Específicos.

Como ***Objetivos Específicos*** se tienen:

- Implementar el Data Warehouse para generar información analítica de los ciudadanos habilitados, depurados y nuevos inscritos en el Padrón Electoral.
- Generar reportes de habilitados, depurados y nuevos inscritos por Geografía Electoral (provincia, municipio, localidad, recinto electoral y mesas)
- Realizar reportes de habilitados, depurados y nuevos inscritos por tiempo(año, cuatrimestre, mes, semana, día)
- Efectuar reportes de habilitados, depurados y nuevos inscritos por grado de instrucción.
- Generar reportes de inteligencia de negocios el cual permita la generación de reportes a nivel gerencial.

1.6. Hipótesis.

En base a los lineamientos anteriores se plantea la siguiente ***hipótesis***: La implementación de un Data Warehouse como parte de una solución de Business Intelligence en el Tribunal

Electoral Departamental de Potosí, brindará información útil, disponible, analítica y oportuna, a nivel Táctico, Estratégico – Gerencial, lo que permitirá el apoyo a la toma de decisiones, en beneficio del Tribunal Electoral Departamental de Potosí.

1.6.1. Operacionalización de Variables.

Variable Dependiente 1: Información útil, disponible, analítica y oportuna.
Variable Dependiente 2: Apoyo a la toma de decisiones.
Variable Independiente 1: Data Warehouse
Variable Independiente 2: Business Intelligence.
Variable de Relación 1: Brindara.
Variable de Relación 2: Permitirá.

1.7. Justificación.

Se plantea una solución basada en tecnologías de información mediante la implementación de Data Warehouse, para obtener un producto que permita Información tratada y útil, la cual apoyará el análisis de los procesos electorales que se requieran, lo que permitirá dar soporte en la toma de decisiones a vocales y directores. Dentro la Institución se aumentará la capacidad de generar y recoger datos, los cuáles serán identificados, categorizados y resumidos para aportar con Información analítica, la misma que podrá ser utilizada eficiente y oportunamente; esto apoyará la toma de decisiones permitiendo incrementar la credibilidad institucional.

1.8. Aporte Teórico.

El ***Aporte Teórico*** de esta investigación se centra en la investigación y escritura de esta tesis. Si bien el concepto de Data Warehouse se originó un poco antes de a mediados de los años 80, en esencia, salvo desconocimiento de mi parte la presente investigación es la primera obra que trata sobre Data Warehouse en nuestro medio.

1.9. Aporte Social.

En cuanto al ***Aporte Social***, si bien el Tribunal Electoral Departamental de Potosí es el encargado de llevar a cabo la inscripción y actualización de datos de ciudadanos al Padrón

Electoral Biométrico, con la implementación del Data Warehouse para el Padrón Electoral Biométrico de Potosí, permitirá dar información estratégica para la toma de decisiones a nivel gerencial, en consecuencia la implementación del Data Warehouse será una contribución a la sociedad el cual con base a toma de decisiones contribuirá a la sociedad con la información que proporcione el Data Warehouse.

1.10. Aporte Metodológico.

El ***Aporte Metodológico*** se centra en la metodología de Ralph Kimball, el mismo que es reconocido como uno de los padres del concepto de Data Warehouse, el cual permitirá futuros proyectos de investigación e implementación de Data Warehouse.

1.11. Métodos de Investigación.

Los ***Métodos de Investigación*** a ser empleados son:

1.11.1. Métodos a Nivel Empírico

1.11.1.1. Entrevista a Vocales y Directores.

Este método fue empleado con la finalidad de obtener opiniones y profundizar acerca de reportes que necesiten para llevar a cabo procesos electorales. El instrumento a utilizar fue la guía de entrevista. La población fue 7 personas entre 4 vocales y 3 directores.

1.11.1.2. Criterio o Consulta a Personal Operativo.

Con la finalidad de conocer la evaluación de la viabilidad del sistema propuesto así como las recomendaciones para su perfeccionamiento continúo. Para este cometido el instrumento que se utilizó fueron cuestionarios. La población fue de 6 personas, entre 4 técnicos con ítem y 2 técnicos eventuales los mismos que forman parte del departamento de informática.

1.11.2. Métodos a Nivel Teórico

1.11.2.1. Estudio Documental.

El método fue empleado con la finalidad de recopilar y registrar los principales usos y normas relacionados con el empleo de medios de desarrollo de Data Warehouse y el desarrollo de inteligencia de negocios, así como sistemas de extracción transformación y carga.

1.11.2.2. Análisis y Síntesis

Este método fue utilizado con la finalidad de estudiar de forma independiente los fundamentos teóricos del desarrollo de Data Warehouse e inteligencia de negocios, rescatando los conceptos importantes para asumir posiciones. Diagnosticando la situación actual para determinar las características de la propuesta que permitan a su vez, ver la propuesta, a partir de sus relaciones como un medio que favorezca el proceso de empadronamiento en el TED. Probando su viabilidad a partir del cual es posible determinar ideas conclusivas y recomendaciones para posteriores investigaciones.

1.11.2.3. Inducción y Deducción.

La inducción existió con la finalidad de realizar generalizaciones y diagnosticar a partir de cada uno de los métodos aplicados en la recopilación de información, las opiniones de vocales, directores y personal operativo, la acción y uso actual de los procesos de empadronamiento. La deducción se refleja en la utilización de estos resultados, es decir dichos rasgos generales determinados teóricamente serán propiedades generales del sistema propuesto.

1.11.2.4. Histórico – Lógico.

Este método fue utilizado en el desarrollo del objeto de investigación, mediante el análisis concreto y funcionamiento del objeto.

1.11.2.5. Enfoque Sistémico.

Este método fue utilizado con la finalidad de identificar los componentes del Data Warehouse, cuyas relaciones existentes entre dichos componentes permitieron diseñar el Data Warehouse.

CAPÍTULO II: MARCO TEÓRICO

2.1. Business Intelligence.

2.1.1. Introducción.

Business Intelligence es un término ideado por Gartner Group en los años 80 para escribir la capacidad de una organización para acceder y explotar la información residente en una base de datos, de manera que los usuarios puedan analizar esa información y desarrollar y adquirir, con base en ella, teorías y conocimientos que apoyen la toma de decisiones del negocio.

El cambiante entorno económico y la problemática de sistemas impulsan al surgimiento del llamado Business Intelligence, el cual es un concepto que trata de englobar todos los Sistemas de Información de una organización para obtener de ellos no solo información o conocimiento, sino una verdadera inteligencia que le confiera a la organización una ventaja competitiva por sobre sus competidores.

2.1.2. Definición.

Business Intelligence (BI) o también conocido mediante su traducción al español como Inteligencia de Negocio, pretende dar una visión de información y conocimiento en las Organizaciones. Para definir Business Intelligence se partirá de la definición de Salvador Ramos:

"Es el conjunto de estrategias y tecnologías que nos van ayudar a convertir los datos en información de calidad, y dicha información en conocimiento que nos permita una toma de decisiones más acertadas y nos ayude así a mejorar nuestra competitividad"[1].

2.1.3. Importancia y Beneficios de Business Intelligence en la Organización.

Business Intelligence es la habilidad para transformar los datos en información, y la información en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones en los negocios, como se muestra en la Figura Nro. 2.1.



Figura 2.1: Objetivos BI

Fuente: FABIAN, A. Implementación De Un Datawarehouse Para El Instituto Geográfico Militar. Pag.7

Dentro de los beneficios que conlleva la implementación de una solución de este tipo, el principal beneficio es que ayuda a tomar decisiones en función de la información que produce una aplicación de BI.

A continuación se darán unos ejemplos de beneficios que suelen ser producto de la implementación de Business Intelligence en las organizaciones, entre las cuales la mejora de competitividad es punto a resaltar.

- Beneficios Tangibles
    - Aumentar Ingresos de Ventas.
    - Aumentar compras de Artículos de mayor uso o necesidad.
    - Eliminar sobreproducción de bienes.
    - Aumentar producción de bienes con más Ventas.
    - Aumento en la información de calidad con tiempos de respuesta cortos.
    - Disminuir el gasto innecesario de productos.
- Beneficios Intangibles
    - Menor dependencia a los sistemas que se desarrolle.
    - Mayor conocimiento de los empleados sobre la Organización.
    - Mejora de la atención al cliente.
    - Aumento en el control de costos.

- Aprender de los errores que se han cometido en el pasado.

- Beneficios Estratégicos
  - Aumento por parte de los Directivos en la visión del(los) negocio(s)
  - Aumento en la gestión de los Directivos.
  - Incremento de habilidades de los Directivos en la toma de decisiones.

2.1.4. Arquitectura de Business Intelligence

Las aplicaciones analíticas incorporan las mejores prácticas tecnológicas, orientadas a entregar una solución completa y eficiente. A continuación se muestra en la Figura 2.2, la Arquitectura de una solución de Business Intelligence.

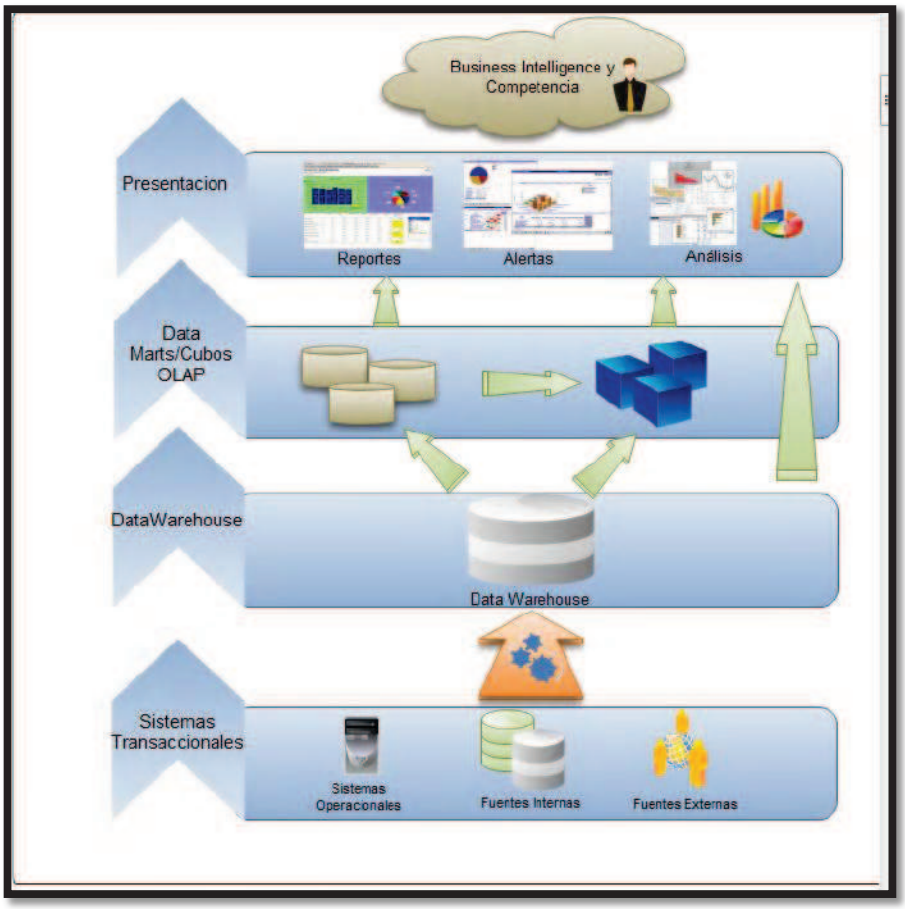


Figura 2.2: Arquitectura BI

Fuente: FABIAN, A. Implementación De Un Datawarehouse Para El Instituto Geográfico Militar. Pag.9

Como se observó en la figura anterior una solución de Business Intelligence inicia desde los sistemas de origen de una organización (bases de datos, ERPs, archivos de texto...), sobre los que suele ser necesario aplicar una transformación estructural para optimizar su proceso analítico (datos de calidad), posteriormente estos datos son almacenados, después pueden ser estructurados específicamente y finalmente ser presentados de diferente manera.

2.1.5. Etapas de Business Intelligence

Según CANALES INOCENCIO, B. en su publicación de Inteligencia de Negocio el proceso de Business Intelligence garantiza un ciclo, que permite generar información analítica constante, es decir obtener la información oportuna que el usuario busca.

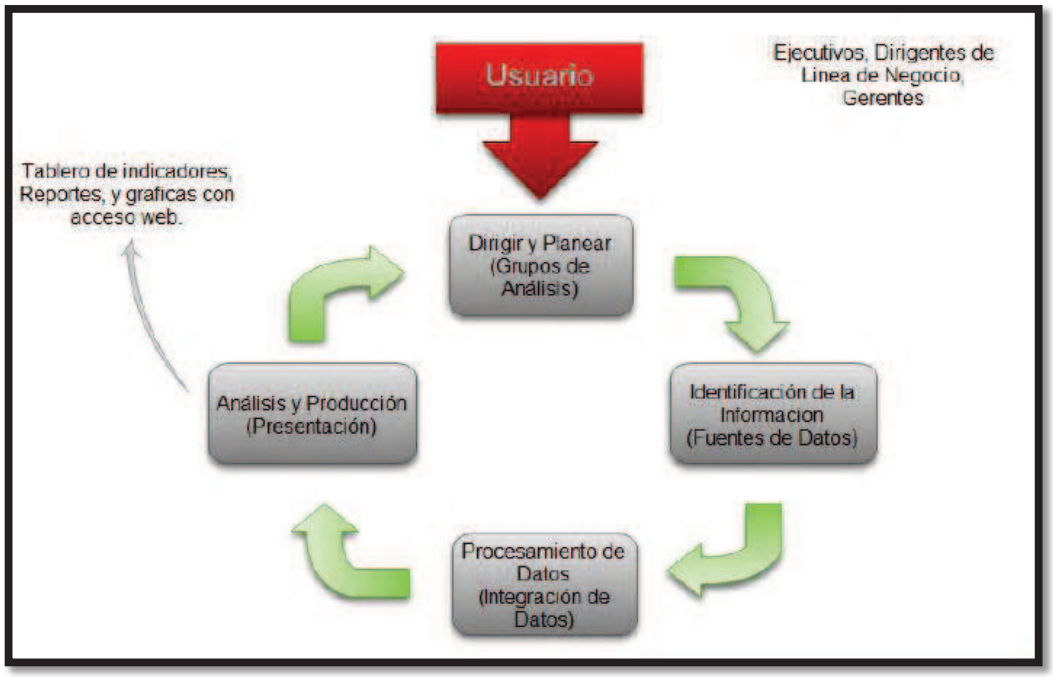


Figura 2.3: Proceso BI [2]

Fuente: http://www.gestiopolis.com/canales2/gerencia/1/busint.htm

2.1.5.1. Dirigir y Planear

Como se observó en la Figura 2.3, la primera etapa es en la cual se inicia y se finaliza el proceso de Business Intelligence mediante la continuidad de la misma y según las necesidades Organizacionales, la cual se inicia desde la redacción de las necesidades de los usuarios y/o

Organización. Estas necesidades son planteadas o generadas en el análisis a preguntas como por ejemplo: ¿Qué producto se vendió más en el primer semestre?, ¿Cuál es el margen de ganancia entre un mes y el otro?

El análisis de las preguntas, produce la redacción de necesidades que se plasmarán en una solución de BI.

Así mismo al finalizar un ciclo del proceso de BI, se obtiene la respuesta a preguntas que van a ser sujetas a análisis por parte de los usuarios, y las cuales servirán de soporte a la toma de decisiones.

2.1.5.1.1. Toma de Decisiones.

Definición.

Es una actividad intelectual realizada por todas las personas para seleccionar una acción u objeto entre varias opciones o alternativas, en beneficio de resolver diferentes situaciones que se presenten. Las decisiones pueden ser:

Estructuradas.

También son conocidas como programadas, ya que son decisiones repetitivas, lo que se convierte en rutinas del día a día.

Por lo general los Gerentes que realizan con frecuencia este tipo de decisiones no necesitan diseñar algún tipo de solución que les ayuden, por el hecho que son decisiones de manera rutinaria y simplemente se rigen a acciones establecidas en una Organización por ejemplo reordenar inventario mínimo.

No Estructuradas.

También conocidas como no programadas, debido a que las decisiones tomadas se presentan de una manera no rutinaria y no se puede predecir ni predefinir las necesidades de información. Tienen que ver con problemas complejos que involucran toda la corporación. Ejemplo: Aumento de producción de un artículo en función de ser el de mayor venta a clientes.

En esta primera etapa se procede también a realizar el planeamiento del proyecto de BI y tomar puntos en consideración como por ejemplo:

- Si se cuenta o no con el patrocinador del negocio.
- Se cuenta con el apoyo de las personas interesadas a las cuales va a llegar una solución de BI.
- Se cuenta con los recursos necesarios para una solución de BI.
- Considerar el tiempo necesario de acuerdo al negocio, en el que se pretende proponer la solución.

2.1.5.2. Identificación de Información

Para empezar con la segunda etapa se debe tomar en cuenta lo planeado en la etapa anterior. Posterior al planeamiento se inicia con lo que son los requerimientos del negocio, como:

Requisitos Funcionales:

- ¿Qué tipo de información necesitan las personas inmersas al negocio?
- ¿En base a que reportes las personas realizan los análisis del negocio?
- ¿Qué tipo de reportes son los que desean?

Requisitos de Datos:

- ¿Qué datos del negocio necesitan?
- ¿Qué datos son considerados críticos para el negocio?
- ¿Hasta qué nivel general y de detalle es requerido los datos?

Requisitos Históricos:

- ¿Desde qué año necesitan ser almacenados los datos?

Finalizado la toma de requerimientos, se da énfasis a la recolección de los datos y a la identificación de la misma en las fuentes de los diferentes sistemas dentro de la organización, y las que van a ser analizas, aquí es importante tomar en cuenta la etapa anterior desde el punto de vista de cuales fueron la necesidades del negocio planteadas, y podar dar inicio a la identificación de los datos.

Para poder identificar los datos es recomendable tomar en cuenta puntos como:

- Saber dónde reside los datos que van a ser analizados, en qué Sistema, en que base de datos o archivos.
- Se cuenta con manuales o diccionarios de datos de los negocios a los que se va a analizar.
- Saber qué persona se encuentra encargada sobre el negocio, es decir quien tiene autoridad sobre esas fuentes de datos.
- Saber que tan limpio se encuentran los datos en las fuentes de origen.
- Saber identificar como se evidencia las reglas de negocio en los sistemas y base de datos.

2.1.5.3. Procesamiento de Datos

En esta etapa se da lugar al proceso de integración de datos, aquí se analiza, diseña y determina en profundidad los datos relevantes, las transformaciones a realizar, y las estructuras de destino como es un Data Warehouse (Almacén de datos).

La integración de los datos se lo realiza mediante un proceso principal que es llamado ETL (Extracción, Transformación y Carga), el cual es considerado como base para la calidad de información con que se va a contar al final de una implementación de Business Intelligence, ya que se extraerán los datos debidamente identificados y los necesarios de acuerdo al negocio, posteriormente se realizará la transformación de datos erróneos, inconsistentes alojados en diferentes sistemas transaccionales, para que finalmente estos datos sean cargados en un Data Warehouse.

2.1.5.4. Análisis y Producción

En esta etapa se da lugar al análisis de la información, que va a ser generada a partir de los datos que se encuentran almacenados de una manera estructurada, de acuerdo al negocio especificado en la etapa de planeamiento. Aquí se podrá visualizar el resultado de Business Intelligence mediante la presentación de la información del negocio, como es por ejemplo mediante la aplicación de reportes, tableros o cuadros de mando, así como también de análisis de la información como son los cubos OLAP, o minería de datos (Data Mining) en busca de tendencia o patrones del negocio.

2.2.Data Warehouse.

2.2.1. Introducción.

Teniendo una mejor perspectiva acerca de Business Intelligence, y mediante la apreciación de la Arquitectura de BI, ya se puede tener una idea del papel que juega un almacén de datos (Data Warehouse) en una solución de este tipo. Un Data Warehouse es una base de datos accesible por los usuarios el cual tiene un registro de datos históricos y actuales acerca de todas las entidades Importantes que se encuentran en la empresa y de acuerdo a negocios específicos.

El Data Warehouse organiza y aloja los datos necesarios, para ser utilizados en el procesamiento analítico dentro de una perspectiva de tiempo. Según "Ralph Kimball y Bill Inmon" un Data Warehouse es:

"Una copia de la transacción de datos específicamente estructurado para consulta y análisis."

"Orientado al tema, integrado, de tiempo variante, de colección de datos no volátil en apoyo a la gestión del proceso de toma de decisiones." [3]

El desarrollo de un Data Warehouse facilita información útil como por ejemplo las ventas obtenidas mensualmente, las cuales pueden ser interpretadas mediante herramientas gráficas, o la cantidad de Existencia de un artículo.

2.2.2. Características

"Un Data Warehouse es mucho más que lo que hemos comentado hasta el momento. Según Bill Inmon se caracteriza por ser" [4]:

- **Orientado a temas**.

  Los datos están organizados por temas para facilitar el entendimiento por parte de los usuarios, de forma que todos los datos relativos a un mismo elemento de la vida real queden unidos entre sí. Por ejemplo, todos los datos de un cliente pueden estar consolidados en una misma tabla, todos los datos de los productos en otra, y así sucesivamente.

- **Integrado**.

Los datos se deben integrar en una estructura consistente, debiendo eliminarse las inconsistencias existentes entre los diversos sistemas operacionales. La información se estructura en diversos niveles de detalle para adecuarse a las necesidades de consulta de los usuarios. Algunas de las inconsistencias más comunes que nos solemos encontrar son: en nomenclatura, en unidades de medida, en formatos de fechas, múltiples tablas con información similar (por ejemplo, hay varias aplicaciones con tablas de clientes).

- **Histórico (variante en el tiempo).**

  Los datos, que pueden ir variando a lo largo del tiempo, deben quedar reflejados de forma que al ser consultados reflejen estos cambios y no se altere la realidad que había en el momento en que se almacenaron, evitando así la problemática que ocurre en los sistemas operacionales, que reflejan solamente el estado de la actividad de negocio presente. Un Data Warehouse debe almacenar los diferentes valores que toma una variable a lo largo del tiempo. Por ejemplo si un cliente ha vivido en tres ciudades diferentes, debe almacenar el periodo que vivió en cada una de ellas y asociar los hechos (ventas, devoluciones, incidencias, etc.) que se produjeron en cada momento a la ciudad en la que vivía cuando se produjeron, y no asociar todos los hechos históricos a la ciudad en la que vive actualmente.

- **No volátil**.

  La información de un Data Warehouse, una vez introducida, debe ser de sólo lectura, nunca se modifica ni se elimina, y ha de ser permanente y mantenerse para futuras consultas.

2.2.3. Arquitectura De Un Data Warehouse.

Según BUIGUES, A. en su publicación sobre arquitectura de un Data Warehouse dice una arquitectura de Data Warehouse es una forma de representar la estructura global de los datos, la comunicación, los procesos y la presentación al usuario final. La arquitectura está constituida por las siguientes partes interconectadas [5]:

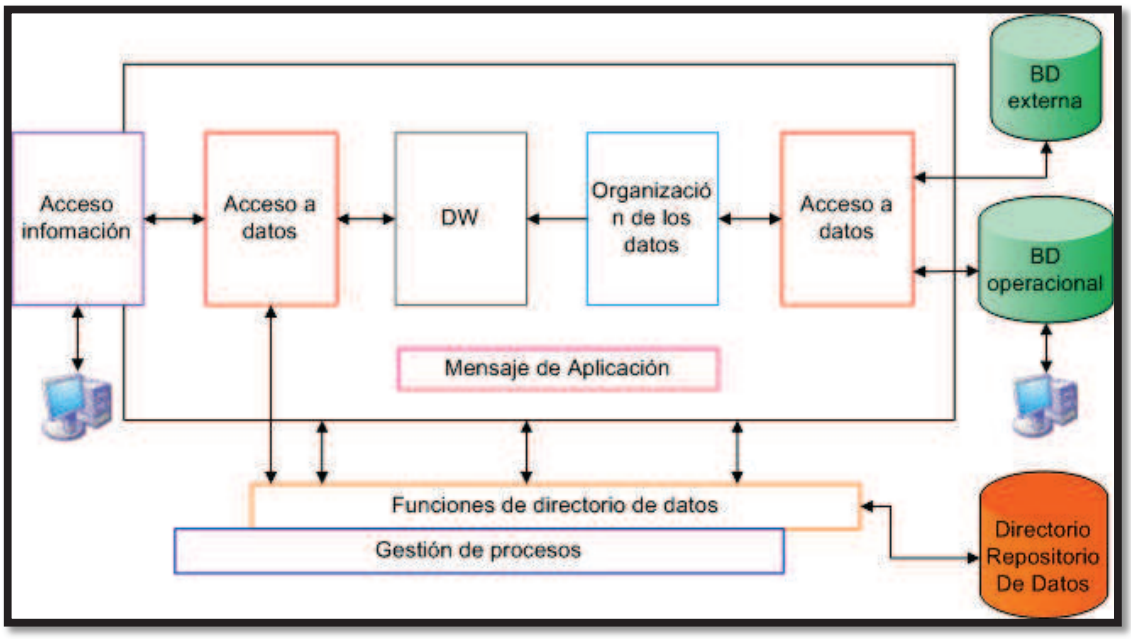


Figura 2.4: Arquitectura de un Data Warehouse

Fuente: http://anabuigues.com/2010/03/05/arquitectura-de-un-data-warehouse/; Mejorado

2.2.3.1. Base de Datos Operacional y Base de Datos Externa

Las organizaciones adquieren datos de bases de datos externas a la propia organización, que incluyen datos demográficos, económicos, datos sobre la competencia, etc.

Mediante el proceso de Data Warehousing se extrae la información que está en la bases de datos operacionales y se mezcla con otras fuentes de datos. Enriquecemos la información.

2.2.3.2. Nivel de Acceso a la Información

Es la capa con la que trata el usuario final. La información almacenada se convierte en información fácil y transparente para las herramientas que utilizan los usuarios. Se obtienen informes, gráficos, diagramas, etc.

2.2.3.3.Nivel de Acceso a los Datos

Comunica el nivel de acceso a la información con el nivel operacional, es el responsable de la interfaz entre las herramientas de acceso a la información y las bases de datos.

La clave de este nivel está en proveer al usuario de un acceso universal a los datos, es decir, que los usuarios sin tener en cuenta la ubicación de los datos o la herramienta de acceso a la información, deberían ser capaces de acceder a cualquier dato del Data Warehouse que les fuera necesario para realizar su trabajo.

2.2.3.4. Nivel de Directorio de Datos (Metadatos)

Para proveer de un acceso universal, es absolutamente necesario mantener alguna clase de directorio de datos o repositorio de información de metadato que ayude a mantener un control sobre los datos. El metadato aporta información sobre los datos de la organización, de dónde proviene, qué formato tenía, cuál era su significado y si se trata de un agregado, cómo se ha calculado éste.

Para mantener un almacén completamente funcional, es necesario disponer de una amplia variedad de metadatos, información sobre las vistas de datos para los usuarios finales y sobre las bases de datos operacionales.

2.2.3.5. Nivel de Gestión de Procesos

Este nivel tiene que ver con la planificación de las tareas que se deben realizar, no sólo para construir, sino también para mantener el Data Warehouse y la información del directorio de datos. Es o el controlador de alto nivel de los procesos que se han de llevar a cabo para que el Data Warehouse permanezca actualizado.

2.2.3.6. Nivel de Mensaje de la Aplicación

Este nivel es el encargado del transporte de la información a lo largo del entorno, se puede pensar en él como un middleware.

2.2.3.7. Nivel Data Warehouse (Físico)

Es el núcleo del sistema, el repositorio central de información donde los datos actuales usados principalmente con fines informacionales residen. En el Data Warehouse físico se almacenan

copias de los datos operacionales y/o externos, en una estructura que optimiza su acceso para la consulta y que es muy flexible.

2.2.3.8.Nivel de Organización de Datos

Incluye todos los procesos necesarios para seleccionar, editar, resumir, combinar y cargar en el Data Warehouse y en la capa de acceso a la información los datos operacionales y/o externos.

2.2.4. Estructura de un Data Warehouse.

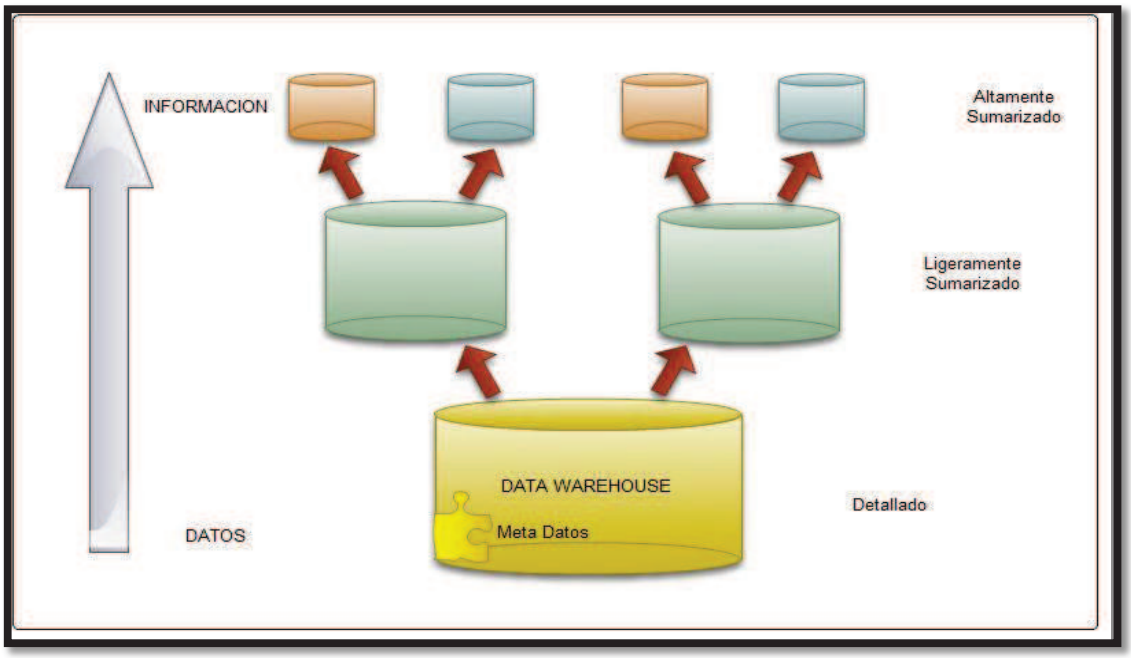


Figura 2.5: Estructura de un Data Warehouse

Fuente: FABIAN, A. Implementación De Un Datawarehouse Para El Instituto Geográfico Militar. Pag.19

La estructura de un Data Warehouse, es básicamente referente a la lógica de almacenamiento de los datos, los cuales se caracterizan por tener diferentes niveles de esquematización y de detalle, cómo se muestra en la Figura 1.5, en la que se puede identificar los niveles de datos, además de los metadatos (metadata) del repositorio. Tomando en cuenta estos aspectos, se estructura al Data Warehouse de la siguiente manera:

- **Datos**

Los datos son claves en un almacén, es lo que genera la información del negocio, estos datos por lo general se encontrarán previamente tratados e integrados.

- **Metadatos**

Son datos estructurados y codificados que describen características de instancias, conteniendo información para ayudar a identificar, descubrir, valorar y administrar las instancias descritas. Estos metadatos por ejemplo contienen:

  - Algoritmos utilizados en las transformaciones de datos.
  - Mapeo de los datos del proceso de integración de datos.
  - Funciones, procedimientos almacenados, vistas q sirven al desarrollo del Data Warehouse.
  - Especificaciones y datos propios del repositorio (datos de sistema).

Con lo descrito anteriormente, se puede dar una idea general en la que los metadatos permiten mantener información de la procedencia de los datos, la periodicidad de refresco, su fiabilidad, forma de cálculo, etc., relativa a los datos de nuestro almacén. Así mismo en cuanto a los datos, poco, ligera y altamente sumarizados, esto se puede identificar de mejor manera en ambientes multidimensionales como los cubos OLAP ya que mediante esta implementación, de explotación de un Data Warehouse se puede apreciar, agregados, navegación de niveles, etc.

2.2.5. Proceso de Extracción, Transformación y Carga

Existe un conjunto básico de procesos de suma importancia detrás de una arquitectura de Data Warehouse, que garantizan la calidad de datos que en ella se almacenarán.

Este proceso de Extracción, Transformación y carga, también conocido como ETL o proceso de integración de datos, cumple con la función principal de organizar e integrar el flujo de datos desde múltiples fuentes, hacia un destino que es el almacén de datos (Data Warehouse).

El proceso ETL brinda soporte a la Gestión de Datos que se va a realizar, obteniendo calidad de los mismos dentro de un almacén como se puede apreciar en la Figura Nro. 2.6.

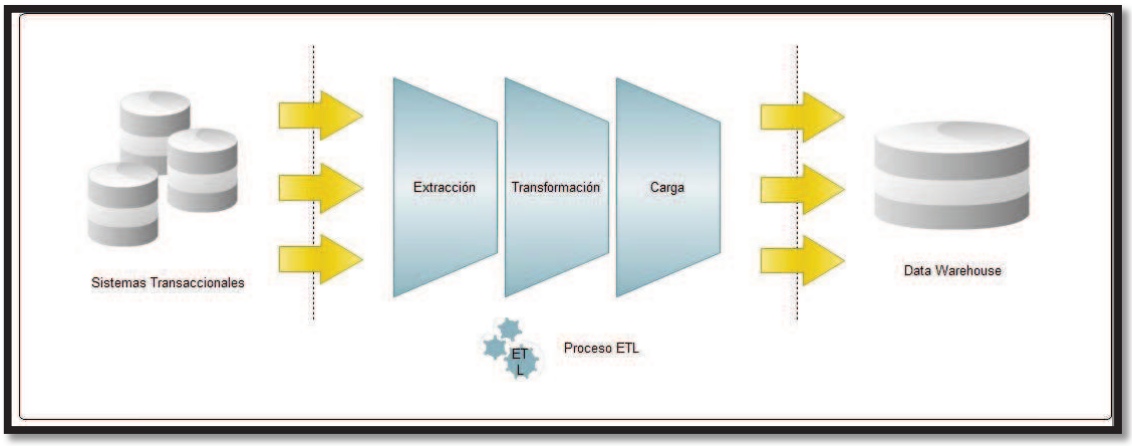


Figura 2.6: Proceso ETL

Fuente: FABIAN, A. Implementación De Un Datawarehouse Para El Instituto Geográfico Militar. Pag.21

Todo el proceso que se lleva a cabo, se debe especificar los tiempos en los cuales se deberá realizar el mismo, lo que garantiza que se mantenga al día los datos en el almacén. Aunque esto va a ser definido de acuerdo a las necesidades de la Organización, ya que por ejemplo se puede definir cargas diarias, así como semanales o mensuales. Este proceso general se encuentra subdividido en 3 subprocesos fundamentales como se detalla a continuación.

2.2.5.1. Extracción.

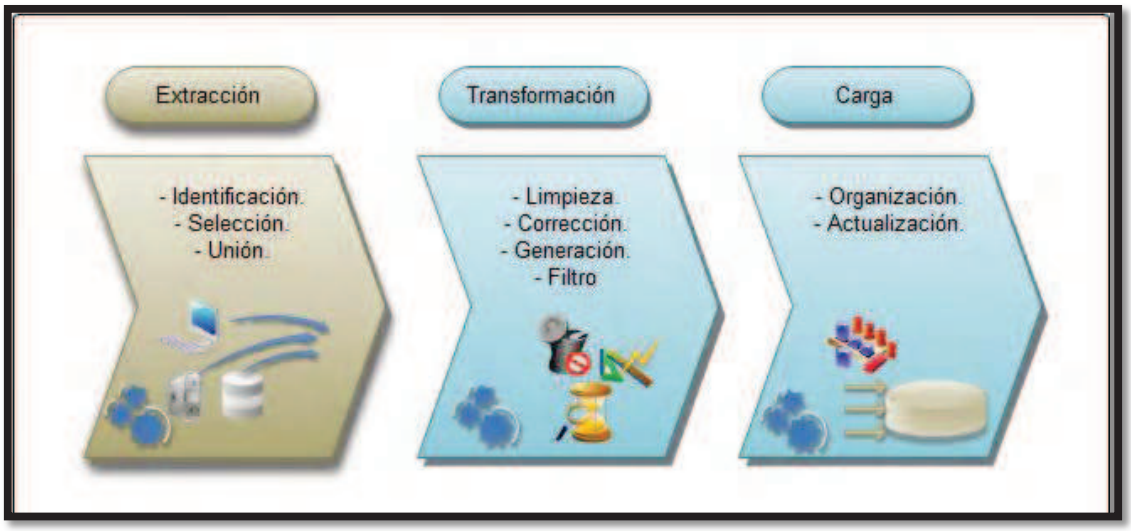


Figura 2.7: Subproceso ETL – Extracción

Fuente: FABIAN, A. Implementación De Un Datawarehouse Para El Instituto Geográfico Militar. Pag.22

La extracción se refiere a la adquisición de los datos, los cuales pueden ser recogidos de diferentes fuentes, como archivos planos, hojas de cálculo, ó bases de datos.

La extracción de los datos se almacenarán en una área temporalmente o Staging Area, vale recalcar que solo se extraerán datos necesarios, es decir de acuerdo a lo que se haya especificado en los requerimientos, ya que en el ambiente transaccional se encuentra gran cantidad innecesaria de datos, por lo que es indispensable la extracción de los mismos, y que serán útiles en el ambiente del Data Warehouse.

2.2.5.2. Transformación.

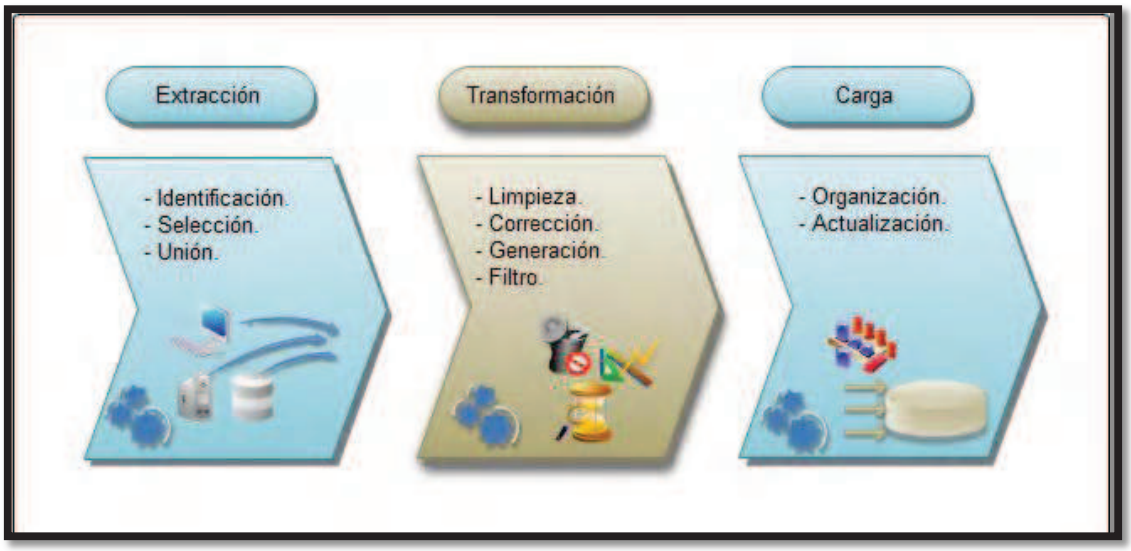


Figura 2.8: Subproceso ETL – Transformación

Fuente: FABIAN, A. Implementación De Un Datawarehouse Para El Instituto Geográfico Militar. Pag.22

Es el subproceso más laborioso con respecto a los otros dos, debido a que en esta etapa se realiza el refinamiento de los datos que han sido extraídos de las diferentes fuentes, por lo que aquí se especificará pasos de acuerdo a los datos que van a ser tratados dando valor para los usuarios. Este proceso incluye corrección de errores, decodificación, borrado de campos que no son de interés, generación de claves, agregación de información, etcétera, lo que es más conocido como limpieza de los datos fuentes.

2.2.5.3.Carga.

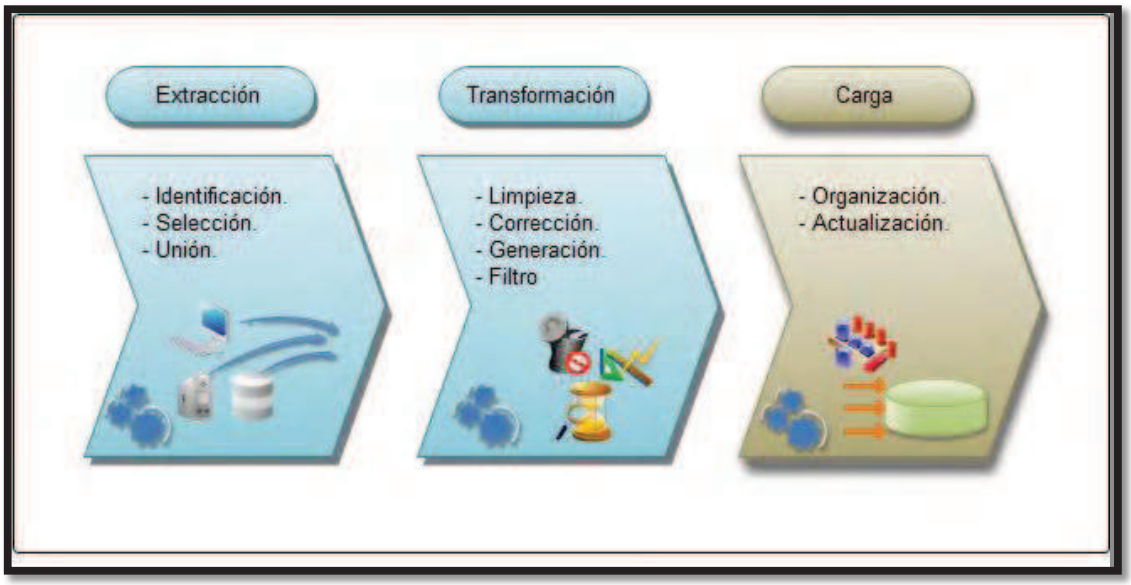


Figura 2.9: Subproceso ETL – Carga

Fuente: FABIAN, A. Implementación De Un Datawarehouse Para El Instituto Geográfico Militar. Pag.23

El último subproceso se caracteriza por realizar la carga hacia el Data Warehouse, los datos que previamente han sido extraídos y tratados en los dos subprocesos anteriores para contar con datos de calidad, ahora se procederá a realizar la carga de los mismos a un nuevo ambiente que es el de almacén de datos, para ello es importante implementar métodos y/o maneras de carga de datos con el fin de controlar por ejemplo datos actualizados ó históricos.

Bien, finalizado todo el proceso ETL, lo que se pretende es contar con datos relevantes para el negocio, los mismos que deben ser de valor sin ningún tipo de codificación, es decir datos transparentes y entendibles por los usuarios finales. Ya contado con la calidad de datos en el ambiente del almacén se termina un ciclo del proceso ETL.

2.2.6. Modelo Dimensional de un Data Warehouse

Un Data Warehouse adopta un Modelo Dimensional en su estructura de almacenamiento, caracterizado por ser un esquema en estrella o copo de nieve, lo que permite maximizar el rendimiento de las consultas. Según CARMEN GLORIA WOLFF el Modelo Dimensional es:

"Una técnica para modelar bases de datos simples y entendibles al usuario final. La idea fundamental es que el usuario visualice fácilmente la relación que existe entre los distintos componentes del negocio" (Carmen Gloria Wolff)

Un diseño dimensional en estrella o copo de nieve es muy diferente del diseño de un esquema de base de datos operacional, en este último, los datos están altamente normalizados para soportar frecuentes actualizaciones y para mantener la integridad referencial, en cambio en un diseño de Data Warehouse, los datos están desnormalizados o redundantes para proporcionar acceso inmediato a los datos sin tener que realizar una gran cantidad de relaciones.

2.2.6.1.Esquema Estrella (Star Schema).

Es la técnica de diseño más popular usada para un Data Warehouse. Es un paradigma en el cual un único objeto en el centro (conocido como tabla de hecho) se encuentra conectado radialmente con otros objetos circundantes llamados tabla de dimensiones formando una estrella.

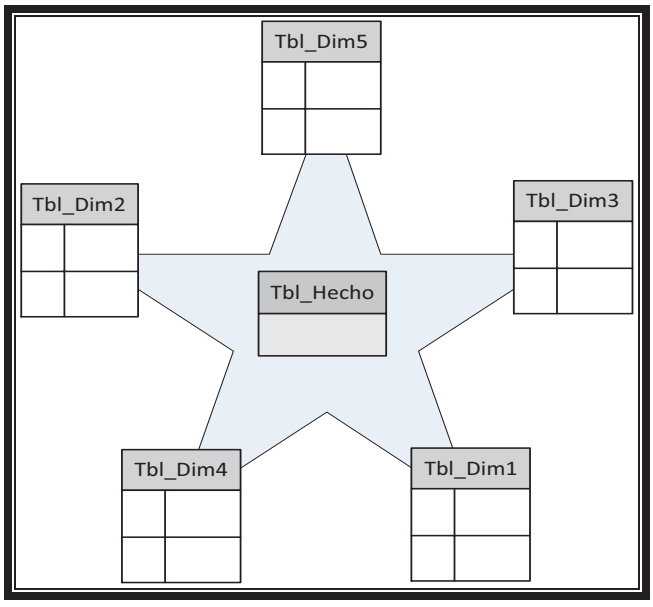


Figura 2.10: Esquema Estrella (Star Schema)

Fuente: Propia

Este tipo de esquema es la más utilizada en un ambiente de Data Warehouse debido a q que los datos se encuentran desnormalizados y por ende las consultas que se pueden realizar son menos complejas ya que no existe la necesidad de realizar muchas relaciones entre tablas.

2.2.6.2.Esquema Copo de Nieve (Snowflake Schema)

El esquema copo de nieve es una extensión del esquema estrella, donde cada punta se explota en más puntas y su denominación se debe a que el diagrama se asemeja a un copo de nieve.

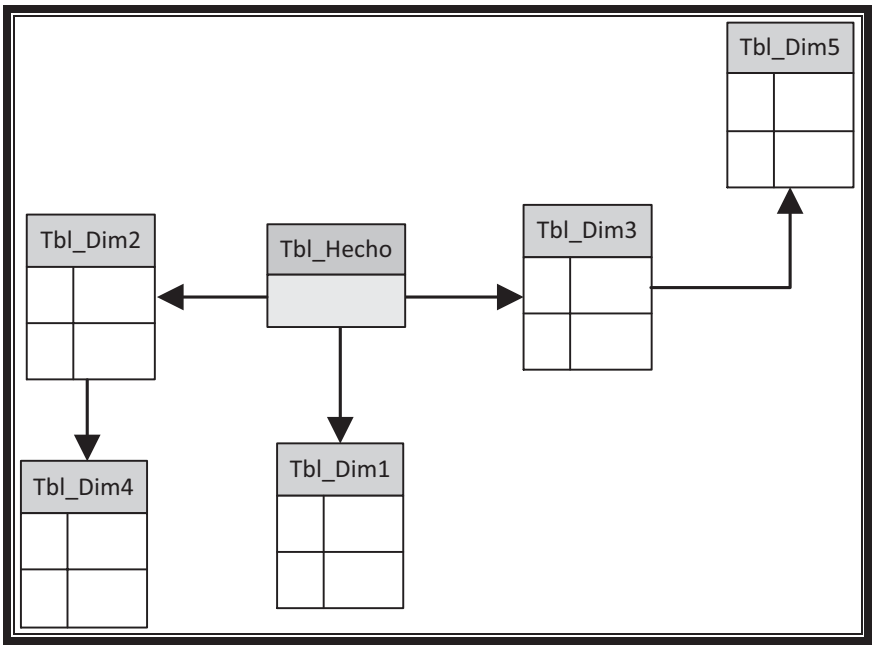


Figura 2.11: Esquema Copo De Nieve (Snowflake Schema)

Fuente: Propia

Como se puede observar en la Figura Nro. 1.11, en el esquema copo de nieve se normalizan dimensiones para eliminar redundancia, permitiendo que los datos de las dimensiones se agrupen en múltiples tablas en lugar de una tabla grande.

Tras la especificación del tipo de esquema que se implemente, un modelo de Data Warehouse está estructurado de la siguiente manera:

2.2.6.3.Tablas

2.2.6.3.1. Tabla de Hechos

Tabla de hechos (fact table), es la tabla central de un esquema dimensional (en estrella o en copo de nieve), que contiene los valores de las medidas de negocio a ser analizadas como se

puede observar en la Figura 2.12 en un modelo de datos tipo estrella. Este tipo de tabla representa el hecho o actividad del negocio, como por ejemplo, Ventas, Movimientos, Pedidos, etc. Por lo general estos datos son numéricos y pueden agruparse (agregación) en un valor total, las medidas pueden ser por ejemplo, cantidad vendida, costo, precio unitario, etc. Es decir son los indicadores que permitirán medir los hechos q se realizan en el negocio.

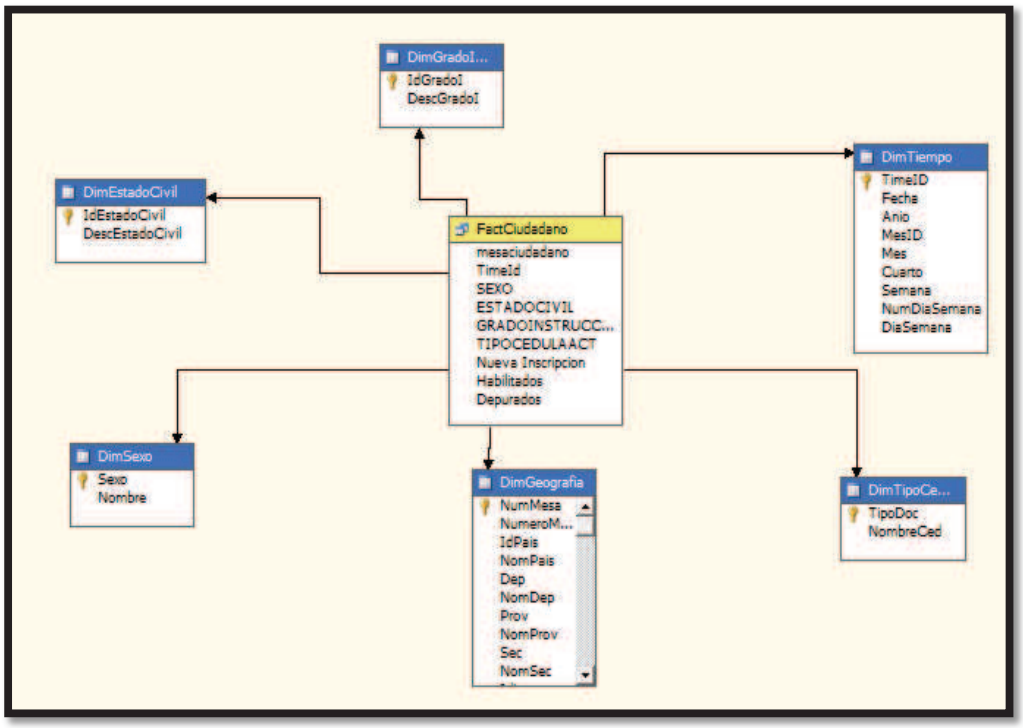


Figura 2.12: Tabla de Hecho

Fuente: Propia

Una tabla de hecho está dividido por 2 tipos de Atributos, estos son: las claves foráneas provenientes de las Dimensiones, y por los indicadores o medidas del hecho.

1.2.6.3.2. Tabla de Dimensiones

Son tablas que describen a la tabla de hecho, mediante atributos descriptores que poseen de acuerdo a un tema específico del negocio como por ejemplo: Clientes, Productos, Ubicación Geográfica, etc. Esto se puede apreciar de mejor manera a continuación

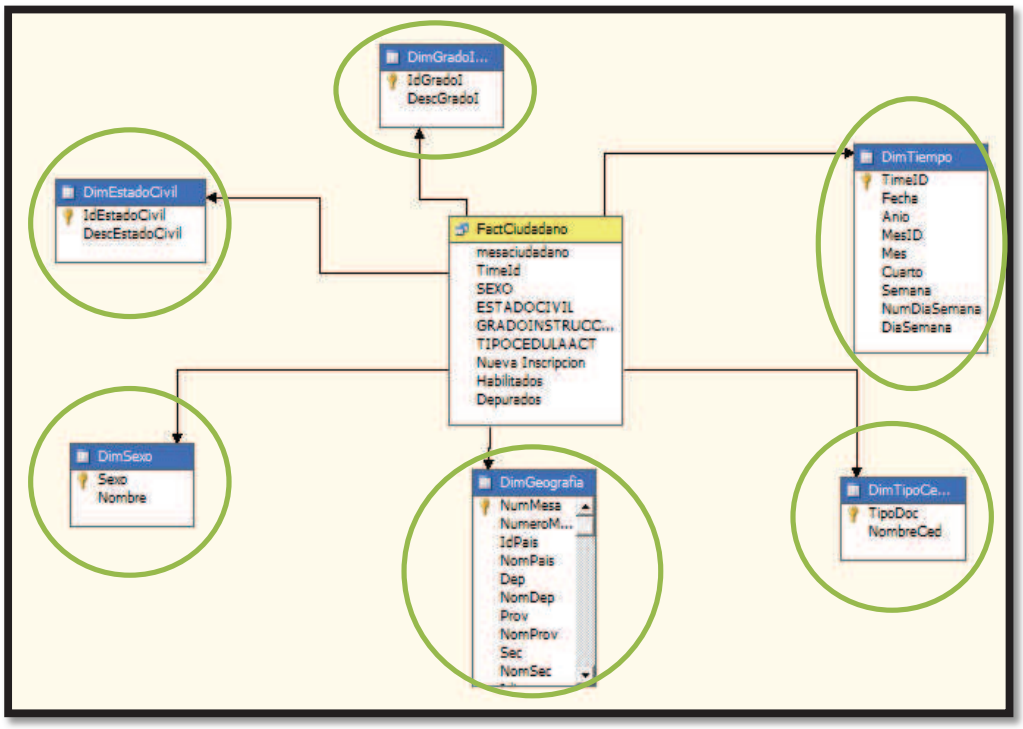


Figura 2.13: Tablas de Dimensiones

Fuente: Propia

Al igual que la tabla de hecho, en la dimensión se cuenta con 2 tipos de atributos estos son: la clave primaria, única para cada registro, y de los atributos descriptores de ellas. Los atributos por lo general son campos textuales o descriptores numéricos cortos.

2.2.6.4.Nivel.

En las dimensiones se pueden especificar características (Atributos) en particular de acuerdo a cada dimensión. Estos atributos necesitan ser ordenados o agrupados de acuerdo a un fin en específico como por ejemplo se puede observar en la Figura 2.14, en la cual dos dimensiones se encuentran agrupadas de cierta manera, a estas agrupaciones son llamadas Niveles.

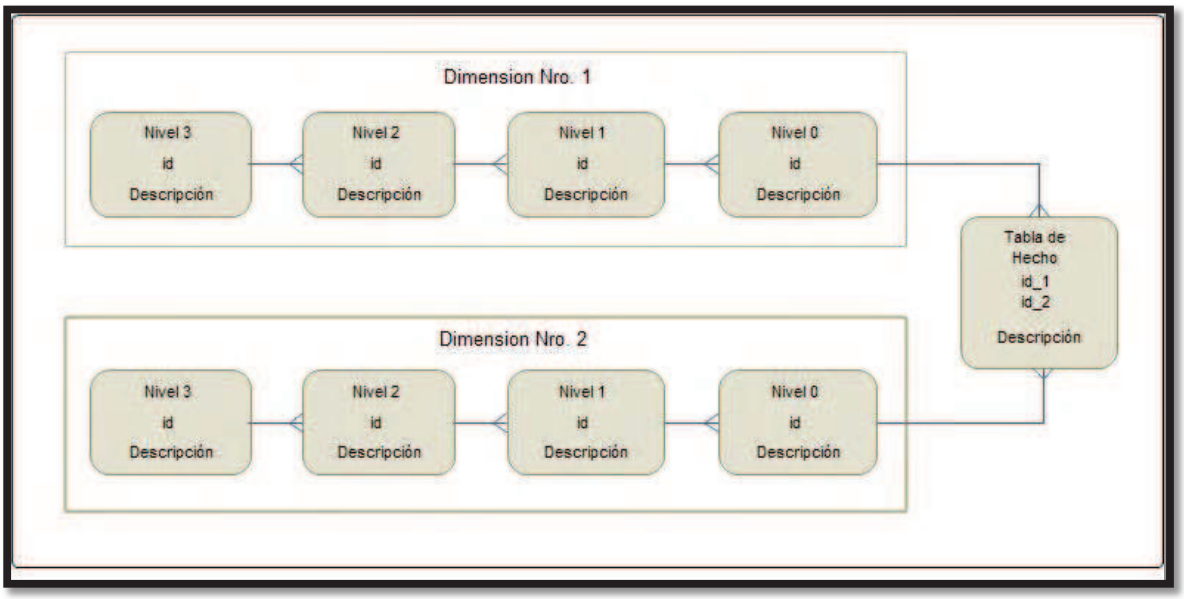


Figura 2.14: Niveles en la Dimensión.

Fuente: FABIAN, A. Implementación De Un Data Warehouse Para El Instituto Geográfico Militar. Pag.29

Para garantizar que existan registros únicos en cada Dimensión es necesario contar con una clave única para cada nivel de agregación, la misma que se recomienda sea secuencial.

2.2.6.5.Jerarquía.

Las jerarquías son estructuras lógicas usadas para categorizar los datos, de acuerdo a los Niveles definidos, la jerarquía también ayuda a definir la navegación por la información que va a tener la Dimensión en un orden (de los niveles) definido.

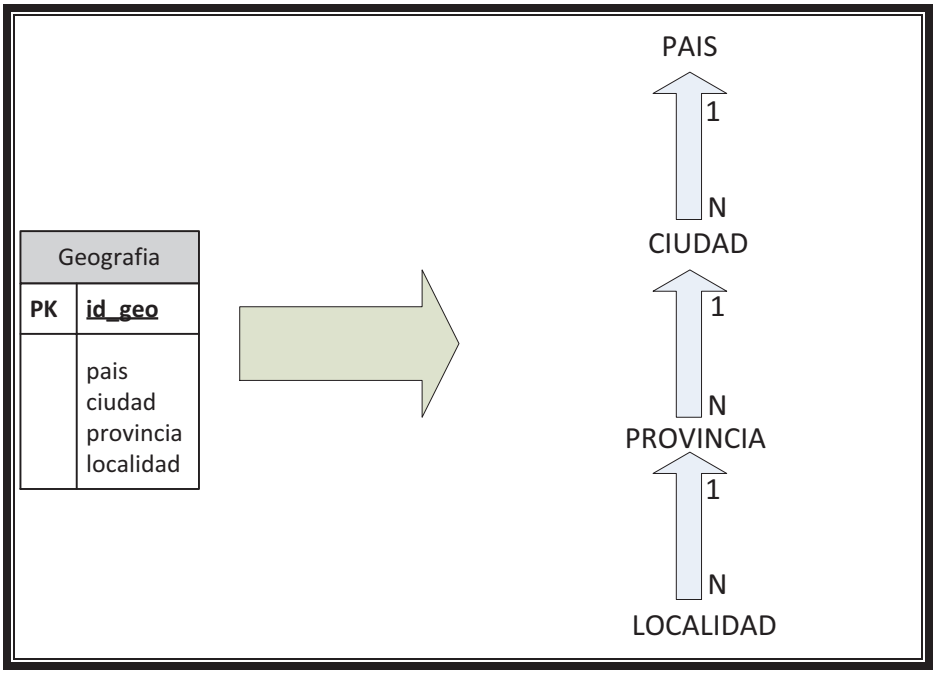


Figura 2.15: Jerarquía en la Dimensión.

Fuente: Propia

2.2.6.6. Dimensiones Lentamente Cambiantes (SCD).

Respecto este punto los registros de las tablas destino (dimensiones) se pueden clasificar de acuerdo al tipo de carga como:

- **Tipo 0 (SCD):** No se tiene en cuenta la gestión de los cambios históricos y no se realiza esfuerzo alguno. De manera que alguna información será sobrescrita, pero otra no. Sin planificación alguna.
- **Tipo 1 (SCD):** No se guardan históricos. La nueva información sobrescribe la antigua siempre. Principalmente la sobre escritura se realiza por errores de calidad de datos. Este tipo de dimensiones es fácil mantener y son usadas cuando la información histórica no es importante.
- **Tipo 2 (SCD):** Toda la información histórica se guarda en el Data Warehouse. Cuando hay un cambio se crea una nueva entrada con su fecha y clave subrogada apropiadas. A partir de ese momento será el valor usado para las futuras entradas. Las antiguas usaran el valor anterior.

- **Tipo 3 (SCD):** Toda la información histórica se guarda en el Data Warehouse. En este caso se crean nuevas columnas con los valores antiguos y los actuales son remplazados con los nuevos.

2.2.6.7. Granularidad.

La granularidad es el nivel de detalle en que se almacena la información. Por ejemplo:

- Datos de ventas o compras de una empresa, pueden registrarse día a día
- Datos pertinentes a pagos de sueldos o cuotas de socios, podrán almacenarse a nivel de mes.

A mayor nivel de detalle, mayor posibilidad analítica, ya que los mismos podrán ser resumidos o sumarizados. Los datos con granularidad fina (nivel de detalle) podrán ser resumidos hasta obtener una granularidad media o gruesa. No sucede lo mismo en sentido contrario. La granularidad está centrada en las medidas es decir en la tabla de hecho.

2.2.6.8. Relaciones.

Las relaciones mantienen la integridad de la información del negocio, referenciando datos entre diferentes tablas, en este caso, relacionando las dimensiones con los hechos.

2.2.6.9. Llaves Subrogadas.

Conocidas también como *surrogate keys,* son caracterizadas por formar parte de las dimensión y de los hechos, al ser un campo único o identificador único de la tabla (Dimensión), lo que permite que se mantenga la integridad en el modelo dimensional (con el hecho). Las claves subrogadas son independientes a las claves primarias de los orígenes de datos, por lo general estas son creadas en los procesos ETL. Las claves subrogadas deben ser enteras ya que esto garantiza el rendimiento dentro del modelo dimensional.

2.2.7. Explotación del Data Warehouse.

La explotación de un Data Warehouse, es prácticamente en donde se puede identificar los resultados que puede llegar a brindar una implementación de este tipo, ya que como se ha podido identificar un almacén de datos no es el fin en sí de una solución de BI, y es solamente

el medio para dar como resultado a la Inteligencia de Negocio, objetivo que se lleva a cabo con la adecuada explotación de los datos almacenados en el Data Warehouse. Es decir mediante una interfaz gráfica se puede apreciar y monitorear el Negocio, como por ejemplo revisar cómo van las ventas en diferentes sucursales, o analizar cuál es el producto más vendido, o cantidad de artículos a comprar, realizadas por la Institución. Entre los medios más utilizados para aprovechar los datos de un Data Warehouse, se citarán los siguientes:

2.2.7.1.Reportes y Consultas.

Dentro de la Organización, es de suma importancia los reportes que se pueden obtener acerca de lo que está pasando con el negocio a la cual esté especificada, así como también acerca de consultas que los usuarios y empleados, como también directivos, desean realizar sobre algo en específico lo que es también conocido como Consulta ad-hoc.

Las consultas o informes libres trabajan tanto sobre el detalle como sobre las agregaciones de la información.

Realizar este tipo de explotación en un almacén de datos supone una optimización del tradicional entorno de informes (reporting), dado que el Data Warehouse mantiene una estructura y una tecnología mucho más apropiada para este tipo de solicitudes.

Los sistemas de "Consultas & Reportes", no basados en almacenes de datos se caracterizan por la complejidad de las consultas, los altísimos tiempos de respuesta y la interferencia con otros procesos informáticos que compartan su entorno, esto es debido a la alta normalización de sus datos.

2.2.7.2.Minería de Datos (Data Mining).

De acuerdo a MSDN de “Minería de datos con Microsoft SQL Server 2008” define que (“La minería de datos es el proceso de detectar la información procesable de los conjuntos grandes de datos. Utiliza el análisis matemático para deducir los patrones y tendencias que existen en los datos. Normalmente, estos patrones no se pueden detectar mediante la exploración

tradicional de los datos porque las relaciones son demasiado complejas o porque hay demasiado datos").

Tomando en cuenta el concepto anterior, se puede definir que el data mining es una automatización de la búsqueda de patrones relevantes en una base de datos. Por ejemplo, un patrón podría ser que los hombres casados sin hijos conducen coches deportivos en el doble de ocasiones que los casados con hijos. En este caso, esta información puede ser muy valiosa para un director de marketing de una empresa automotora.

El proceso de minería se integra con el sistema para el almacenamiento de cantidades masivas de datos (Data Warehouse) de la empresa y presenta la información dirigida a usuarios con capacidad de decisión en el negocio. Para que la minería de datos pueda obtener información útil de las bases de datos de la empresa, es necesario que éstas dispongan de datos de calidad (completos, estructurados, relevantes, históricos, actualizados, etcétera), propósito que cumple el Data Warehouse.

Ejemplos de preguntas que podría responder un proceso de Data Mining son:

- ¿Qué consumidores es más probable que anulen su contrato de telefonía móvil?
- ¿Cuál es la probabilidad que un cliente compre por lo menos 60 dólares del último catálogo de venta por correo?

2.2.7.3.Análisis Olap.

OLAP (On-Line Analytical Processing) o Procesamiento Analítico en línea, surge como contraste a OLTP (On-Line Transactional Processing) que define a los sistemas de ambientes transaccionales.

El análisis multidimensional (Análisis OLAP), parte de una visión de la información como dimensiones de negocio, en la que hay que tomar en cuenta que se debe de olvidar lo que son las tablas y campos, dando mayor énfasis a lo que son las dimensiones y medidas.

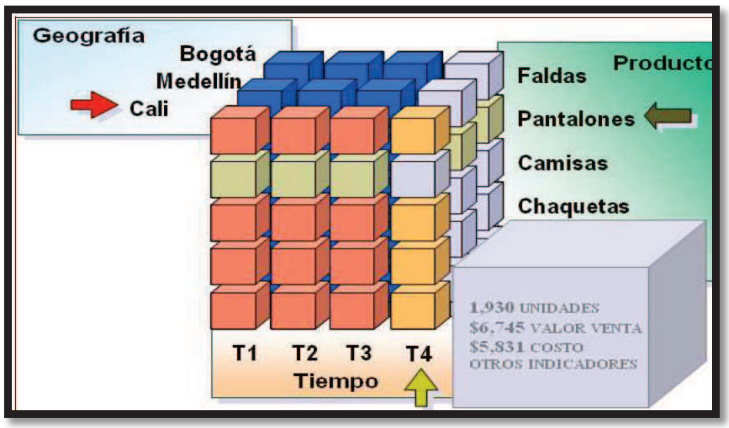


Figura 2.16: OLAP (On-Line Analytical Processing)

Fuente: http://www.monografias.com/trabajos55/mineria-de-datos/mineria-de-datos.shtml.

Las herramientas OLAP ofrecen a los usuarios, acceso a los almacenes de datos (Data Warehouse) para que puedan obtener las funcionalidades de análisis avanzados que requieren los usuarios finales y analistas, por lo que es considerada la herramienta ideal para sacar el mayor provecho a los datos almacenados en un repositorio.

Las herramientas OLAP se caracterizan por subdividirse en 3 tipos de acuerdo a la manera de almacenar los datos, estos son:

2.2.7.3.1. Rolap (Relational Olap).

Arquitectura en la que se almacenan los datos en un motor de base de datos relacional, pero de igual manera se proporciona la funcionalidad analítica. A través de esta implementación se soporta de mejor manera las capacidades OLAP con respecto a las bases de datos relacionales, en el sentido que realiza consultas directas a la base de datos, e igualmente presenta los datos de la manera multidimensional caracterizada por la Arquitectura. Los esquemas más comunes sobre los que se trabaja son estrella o copo de nieve.

El sistema ROLAP utiliza una arquitectura de tres niveles. La base de datos relacional maneja los requerimientos de almacenamiento de datos, y el motor ROLAP proporciona la funcionalidad analítica. El nivel de base de datos usa bases de datos relacionales para el

manejo, acceso y obtención del dato. El nivel de aplicación es el motor que ejecuta las consultas multidimensionales de los usuarios.

2.2.7.3.2. Molap (Multidimentional Olap).

En este tipo de Arquitectura los datos se almacenan de manera dimensional en un servidor de base de datos multidimensional, permitiendo optimizar los tiempos de respuesta en la información, ya que al ser sumarizada y/o agregada ayuda mucho a los datos calculados por adelantado como por ejemplo los totales, lo que aumenta el desempeño de análisis.

2.2.7.3.3. Holap (Hybrid Olap).

La arquitectura OLAP híbrida (HOLAP), se caracteriza por combinar las arquitecturas ROLAP y MOLAP para brindar una solución con las mejores características de ambas: desempeño superior y gran escalabilidad. Un tipo de HOLAP mantiene los registros de detalle (los volúmenes más grandes) en la base de datos relacional, mientras que mantiene las agregaciones en un almacén MOLAP separado. Para tener un idea de cómo funciona la tecnología OLAP se puede hacer una analogía con la representación de un cubo, y es aquí donde nace el término de los Cubos de Información, que no son otra cosa más que el análisis multidimensional de los datos, como se puede apreciar en la Figura 2.17 a continuación.

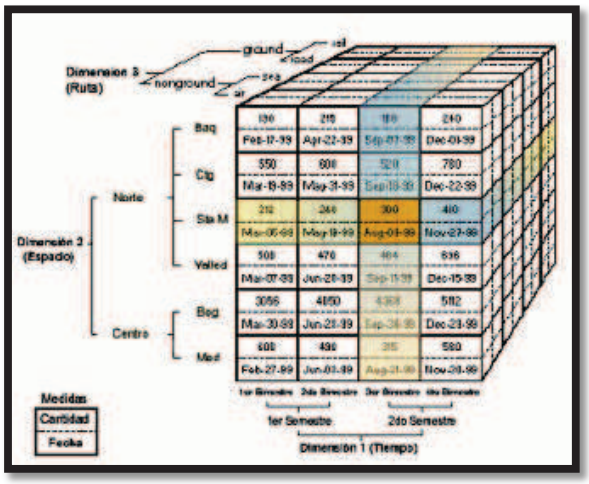


Figura 2.17: Cubo OLAP

Fuente: http://www.monografias.com/trabajos55/mineria-de-datos/mineria-de-datos.shtml.

Con lo anterior tratado se puede analizar desde una mejor perspectiva el concepto de OLAP e identificar en cómo se puede definir la arquitectura OLAP de acuerdo a tipo de almacenamiento de datos.

En un análisis OLAP se implementan funcionalidades las cuales permiten interpretar la información de diferentes maneras, estas funciones u operaciones entre las más importantes son:

- **Drill Down.**

Desglosar una métrica de lo general a lo particular por la jerarquía de sus dimensiones.

- **Roll Up.**

Agregar una métrica de lo particular a lo general por la jerarquía ascendente de sus dimensiones.

Existen dos tipos de operaciones de acuerdo a la cantidad de Dimensión utilizadas para un cubo.

- **Dice.**

Obtener un sub-cubo fijando dos o más de sus dimensiones.

- **Slice.**

Obtener un sub-cubo fijando una de sus dimensiones.

2.3.Metodología Ralph Kimball.

RALPH KIMBALL plantea una metodología para el Desarrollo de soluciones de BI/DW, que garantiza buenas prácticas para la implementación de un proyecto de este tipo, mediante la consecución de una serie de pasos a seguir durante toda la Administración del Proyecto, como se muestra en la figura a continuación.

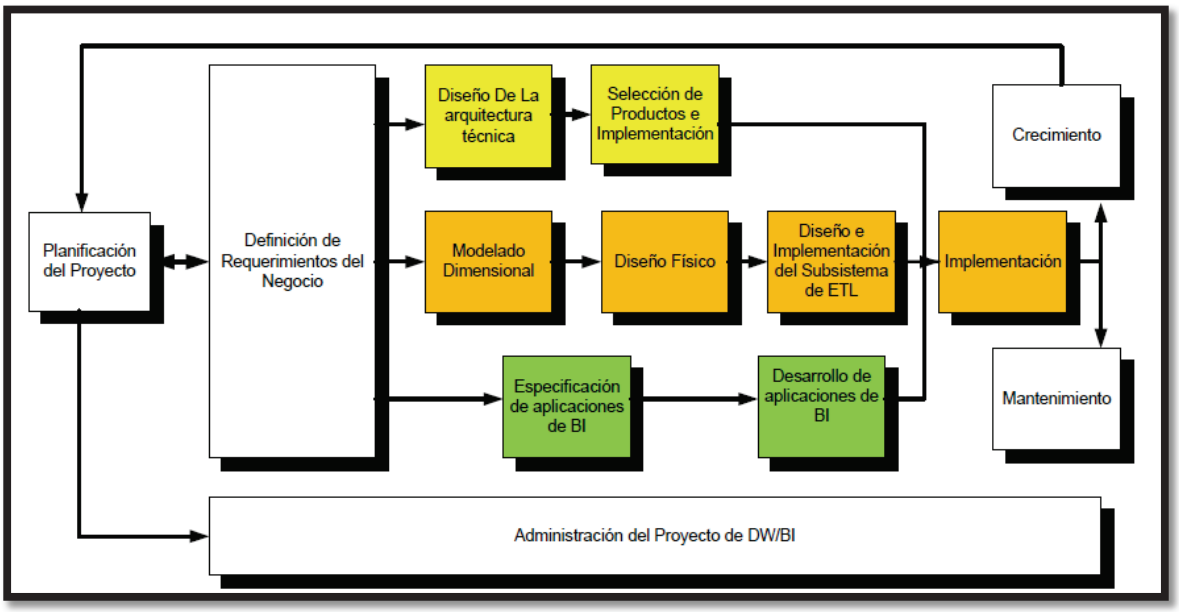


Figura 2.18: The Kimball Lyfecycle Diagram

Fuente: KIMBALL, R. The Data Warehouse Lifecycle Toolkit,. Pag 2.3

En base a esta metodología, se tomará los puntos con mayor importancia para que sean especificados, y se realizará ciertas adecuaciones para este proyecto según las necesidades que se presenten [6].

2.3.1. Planificación del proyecto.

La planificación busca identificar la definición y el alcance del proyecto, incluyendo justificaciones del negocio y evaluaciones de factibilidad.

Definición del proyecto

En la definición se identifica el escenario para poder determinar el alcance del proyecto, en el sentido que se puede apreciar de donde se origina la demanda dentro de la organización.

Los posibles escenarios originados dentro de una empresa son:

- Demanda de un sector del negocio. Un Sector en específico como por ejemplo los directivos del negocio.

- Demasiada demanda de la información. Se caracteriza por provenir de más de un sector del negocio, por ejemplo aparte de los directivos del negocio, los de la empresa y ciertos gerentes relacionados con el negocio.
- En busca de demanda. Escenario en el que se encuentra involucrado el presidente o directivo principal de la empresa.

Determinar preparación de la empresa para un proyecto de Data Warehouse.

Son cinco factores importantes según la metodología de Ralph Kimball, las que se deben tomar en cuenta para poder iniciar un proyecto de Almacén de Datos, estos son:

- Patrocinio de la Gerencia del negocio.
- Motivación del Negocio.
- Acompañamiento del departamento de tecnología y de negocio.
- Presencia de cultura Analítica.
- Factibilidad.

Desarrollo del alcance preliminar.

Se centra en definir de manera preliminar el alcance del proyecto, para poder hacerlo, la base debe ser los requerimientos del negocio. Para la definición del alcance es importante seguir los siguientes parámetros:

Desarrollo de la Justificación del Negocio.

En la Justificación se deben especificar e identificar el costo y beneficio que está relacionado con el proyecto.

Los componentes a ser considerados de inversiones y costos son los siguientes:

- Costos de Hardware. Se toma en cuenta el costo de acuerdo a los equipos que serán utilizados para el proyecto.
- Costos de Software. Tomar en cuenta el costo de Software que se necesitara para el proyecto, como software para ETL, reportería, entre otros.
- Recursos Internos. Recursos con los que la organización cuenta para el proyecto.

- Recursos Externos. En caso de necesitar algún tipo de consultoría o servicio externo a la organización.
- Costos de Mantenimiento. La mayoría de productos de hardware y software necesitan de un mantenimiento.
- Soporte. Todo lo que es el soporte a los usuarios ya sean estos a través de help desk.
- Costos de Crecimiento. Se debe considerar nuevos requerimientos, mejoras tecnológicas de los productos relacionados de software.

Se considera también los retornos y beneficios a obtener, como:

- Incremento de ganancias.
- Incremento en los niveles de servicio a los clientes.
- Aumento de información para la empresa, y su rápido acceso a ella.

Planeación del Proyecto.

Se debe especificar un nombre para el proyecto, así como los roles de la(s) persona(s) que participara en el proyecto, dependiendo la participación de cada una y requerimientos que se hayan especificado. Los principales roles para el proyecto son:

- Patrocinadores de Negocio.
- Gerente o Administrador del proyecto.
- Líder de Negocio del proyecto.
- Analista de negocio.
- Arquitecto de Dato.
- Modelador de datos.
- Administrador de Base de Datos.
- Diseñador de Proceso de ETL.
- Desarrolladores de aplicaciones de BI para usuarios.
- Educador o capacitador del Almacén de datos.

Administración del Proyecto

Aquí se deberá considerar las reuniones con el equipo de trabajo, el monitoreo de estado del proyecto realizado periódicamente, el alcance y estrategias de comunicación ya sea con el equipo de trabajo o con usuarios externos, en estos últimos esencialmente la manera de tratar

por ejemplo cuando llegan nuevos requerimientos a los cuales hay que ver la manera de manejar estas situaciones.

Como entregable de la etapa es el alcance y justificación del Proyecto.

2.3.2. Definición de los Requerimientos del Negocio.

Los requerimientos de los usuarios del negocio ayudan a definir el diseño e implementación del Data Warehouse, es por esto que desde una perspectiva macro los requerimientos son el centro del negocio y por ende factor de éxito para un satisfactorio desarrollo de un Data Warehouse.

Algo importante a tener en cuenta es saber identificar y asociar los requerimientos desde una perspectiva macro y así como también desde pequeñas perspectivas.

Para empezar se debe hablar con los usuarios del negocio, sobre sus trabajos, objetivos, e intentar conocer cómo y en base a que toman decisiones, al contrario lo que no se debe hacer es preguntar a los usuarios que datos quieren que aparezca o conste en el Data Warehouse, ya que al conocer sus objetivos, decisiones, metas tendremos una mayor perspectiva de que desean o de qué tipo de información manejan con más frecuencia.

Preparar para la entrevista.

Se especifican los roles del equipo entrevistador. Es aconsejable que el mismo equipo maneje las entrevistas tanto con los usuarios del negocio como el equipo de tecnologías de la empresa. En el equipo los roles principales son, el del líder de entrevista que su principal función es dirigir el interrogatorio, y el otro rol que es el de una persona que tome nota durante las entrevistas (“Escribano”).

Investigación de la organización.

Antes de poder iniciar el proceso de levantamiento de requerimientos, se deben analizar los reportes anuales de la organización, para poder determinar en base a que se toman las

decisiones y estrategias. Otro factor importante es analizar el negocio y conocer las fortalezas y debilidades.

Selección de los entrevistados.

Para la correcta recolección de requerimientos es importante identificar y seleccionar las personas representativas de cada área de la organización, lo que también es posible identificarlos mediante un organigrama de la organización.

Es fundamental dirigirse hacia los directivos o gerentes de cada negocio, en si a las personas con mayor nivel en la organización debido a que estas personas cuentan con un conocimiento mayor del negocio en el que están involucrados.

Llevar a cabo las entrevistas

Se empieza por las preguntas más generales y seguir con las específicas. La idea es conseguir del personal y directivos, la información suficiente del negocio, objetivos y visión de cada departamento. También es recomendable realizar entrevistas al personal del departamento de tecnologías, para poder tener presente con que se cuenta para el desarrollo del proyecto.

Revisar los informes y los análisis existentes

Algo importante es tomar en cuenta los reportes que están implementados en la organización, ya que pueden existir reportes analíticos o estadísticos que ayudan al análisis de los requerimientos, y servir también de base para las entrevistas en el sentido de que ayudan a tener la idea de cómo se maneja la información en la organización. Vale recalcar que sebe distinguir entre los reportes y el sistema analítico (Data Warehouse) que se va a implementar.

Concluir la entrevista.

Como el resultado de la entrevista es llegar a una conclusión, se pide a cada entrevistado acerca de su o sus criterios de éxito con respecto al proyecto. Por supuesto, cada criterio debe ser medible y cuantificable. Finalmente se debe agradecer a los entrevistados por sus ideas y hacerles saber lo que pasará posteriormente y lo que será su participación.

Revisar el resultado de la entrevista y Análisis.

Es importante tener conocimiento de los sistemas operativos de la empresa, para ayudar a determinar la factibilidad de implementar los requerimientos recolectados. Se debe revisar los reportes claves y analizar la información recogida de las entrevistas, porque generalmente esto proporciona una traducción aproximada al modelo de datos dimensional, ya que por ejemplo las tablas de dimensiones suministrarán los valores de los títulos de informe y los encabezados de columna o fila superior, mientras que los reportes de medidas numéricas se representan como hechos. Esta especificación es útil para examinar los informes fuera de línea (offline) para una mayor comprensión sobre la dimensionalidad del negocio y las necesidades analíticas de los usuarios.

Para finalizar es importante documentar los requerimientos obtenidos para su aprobación, con lo cual esto pasará a ser el entregable de esta etapa.

2.3.3. Diseño Técnico de la Arquitectura.

Se detalla el ambiente de integración, junto con las tecnologías a utilizar, y el cómo va a funcionar, uniendo tres factores que son: los requerimientos, el entorno técnico, y las estrategias de diseño a implementar. Para determinar el diseño técnico se basará en dos divisiones que son:

2.3.3.1.Back Room.

Se implementa la integración de datos mediante el Proceso de ETL, desde el origen de los datos hasta la carga de los mismos en el Data Warehouse que es la base destino.

2.3.3.2.Front Room.

Parte pública y presentable del Data Warehouse, en la cual se implementa las aplicaciones necesarias para poder presentar al usuario final con el fin de que puedan analizar los datos almacenados en el Data Warehouse.

En esta etapa, se realiza el Diseño de la Arquitectura de todo el proceso de integración de los datos y de la presentación de los mismos, con el fin de obtener un esquema gráfico del funcionamiento del Data Warehouse.

2.3.4. Selección e Instalación del Producto.

Mediante las especificaciones tratadas en el diseño técnico de la arquitectura, se visualiza de mejor manera el ambiente en el cual se va a desarrollar el proyecto. Se evalúa las herramientas a utilizar de acuerdo a cada sección o área del proceso de desarrollo. Con lo cual estará dirigido para:

- **Plataforma DBMS.**

Existen dos tipos, motor relacional y motor OLAP, lo que se deberá tener en cuenta cuales son las necesidades, para seleccionar una o ambas opciones.

- **Herramienta ETL.**

Se evalúa la herramienta a utilizar, tomando en cuenta aspectos importantes, como si es funcional, y si cumple las posibles necesidades a presentarse en el desarrollo, como por ejemplo si la herramienta puede generar claves subrogadas y hacer búsquedas de ellas, entre otras; esto ayudará determinar la herramienta a utilizar.

- **Herramienta BI.**

Para seleccionar la mejor opción de una herramienta de Front room o de presentación, se debe tomar en cuenta la funcionalidad, versatilidad y manejo amigable para el usuario final.
En esta etapa como entregable, son las herramientas a utilizar en cada sección de desarrollo.

2.3.5. Modelamiento Dimensional.

El Modelamiento Dimensional es un técnica de diseño lógico para estructurar datos de manera intuitiva, dirigida a los usuarios de negocio, y permitiendo mayor rendimiento en las consultas.

El modelo dimensional es estrictamente diferente a los modelos relacionales, ya que en el primero los datos se caracterizan por ser redundantes mientras que en el segundo se implementa el Modelamiento 3NF (Third Normal Form) la cual busca eliminar los datos redundantes.

En el modelo dimensional se especifican las medidas y las características. Las medidas son usualmente numéricas y son referidas a los *Hechos*, mientras que las características de estas medidas son representadas por las *Dimensiones*, describiendo por ejemplo quien, que, cuando, de la medida.

La estructura característica del modelo dimensional es el de la semejanza a la de una estrella, es por esto que se es llamado también modelo en estrella.

En la etapa de Modelamiento Dimensional, se realiza el Diseño de la Base de Datos del Data Warehouse, es decir el modelo de las tablas, junto con la Matriz de Procesos/Hechos vs. Dimensiones.

2.3.5.1.Cuatro Pasos Para el Proceso de Diseño Dimensional.

2.3.5.1.1. Paso 1 – Seleccionar el Proceso de Negocio.

El primer paso es determinar el proceso que va a ser modelado. Esto es un formalismo ya que la selección del proceso es básicamente referente a los requerimientos que se tomaron, a través de ello ya se puede determinar cuál será el proceso a que se va a modelar. Los procesos de negocio son la parte fundamental del Data Warehouse, y es aconsejable construirlo de manera iterativa es decir un proceso por momento. Vale recalcar que cada proceso de negocio puede producir una o varias tablas de hechos.

2.3.5.1.2. Paso 2 – Declaración de Granularidad.

El segundo paso es definir claramente el nivel de granularidad lo que es lo mismo un registro de la tabla de hecho y decir exactamente la representación de la medida de la tabla de hechos. Un ejemplo es:

- Una medición de inventario realizado cada mes para cada producto en un almacén. Ó
- Medición de cantidad de ventas a un cliente de un producto en una tienda.

El grado de granularidad es fundamental, debido a que de ella depende el volumen de datos que se almacenarán en el Data Warehouse, es decir a menor granularidad será mayor el detalle, y a mayor granularidad será menor el detalle y por ende menor volumen de los datos.

2.3.5.1.3. Paso 3 – Identificación de Dimensiones.

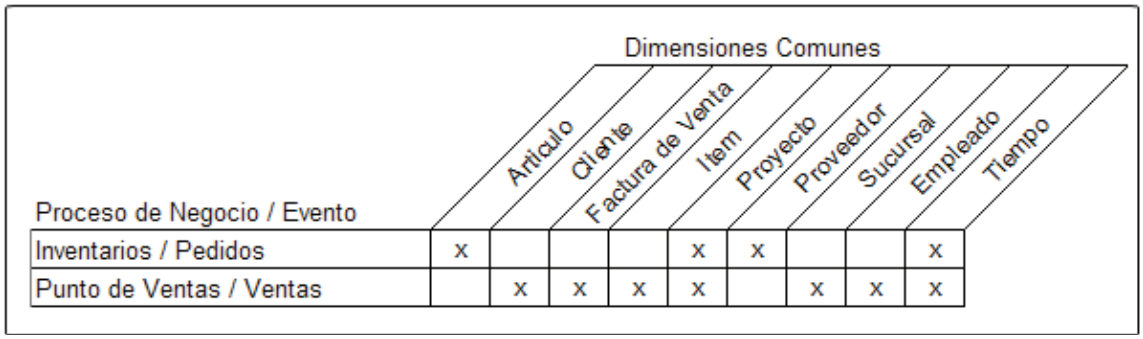
El tercer paso es determinar las dimensiones aplicables a la tabla de hechos en el nivel de granularidad que se detalló. Al agregar nuevas dimensiones los atributos de estas deben cumplir con la misma granularidad que se haya definido.

2.3.5.1.4. Paso 4 – Identificación de Hechos.

El paso final es identificar los hechos o medidas del negocio. Para la mayoría de los procesos orientados a transacciones, solo hay algunos hechos fundamentales, que miden por ejemplo la cantidad o monto.

2.3.5.2. Data Warehouse Bus Matrix

Es una Matriz Procesos versus Dimensiones comunes, la cual permite tener un panorama más claro del Data Warehouse, al mostrar las dimensiones, hechos, de acuerdo al proceso de negocio. [7]



Proceso de Negocio / Evento	Artículo	Cliente	Factura de Venta	Item	Proyecto	Proveedor	Sucursal	Empleado	Tiempo
Inventarios / Pedidos	x				x	x			x
Punto de Ventas / Ventas		x	x	x	x		x	x	x

Figura 2.19: Data Warehouse Bus Matrix

Fuente :http://jpgarcia69.wordpress.com/2008/07/15/10-pasos-para-una-negociacion-efectiva/

2.3.6. Diseño Físico.

Esta etapa se centra en definir la estructura física, como:

- Configuración del entorno de base de datos.
- Creación de espacios de tablas para Datos y Metadatos.
- Creación de tablas, para el Data Warehouse y para las bases origen en caso de ser necesario.

- Creación de llaves primarias y foráneas.
- Creación de índices en las tablas.
- Desarrollo de tablas temporales (Staging Area) en caso de ser necesario.

En esta etapa, se desarrollan los scripts, que son generados al finalizar el diseño y desarrollo físico.

2.3.7. Diseño y Desarrollo del Proceso ETL.

El desarrollo del Sistema ETL es un desafío para un proyecto de este tipo ya que se debe tomar en cuenta aspectos como: estudio y análisis de los datos, integración de los mismos con el objetivo de resolver problemas de calidad, cargas automáticas; por lo que se considera fundamental al Proceso ETL en el marco de desarrollo de un proyecto de BI/DW.
Se plantean pasos para el diseño y desarrollo de un proceso ETL, estos son:

2.3.7.1.Paso 1 – Trazar un Plan de Alto Nivel.

Se diseña un mapa mediante cuadros que especifican la actividad a realizar, con el fin de mostrar el flujo de datos, desde el origen hacia el destino.

2.3.7.2.Paso 2 – Seleccionar la Herramienta ETL.

Se selecciona la herramienta ETL adecuada, tomando en cuenta las funcionalidades que se van a utilizar.

2.3.7.3.Paso 3 – Desarrollar Estrategias por Defecto.

Teniendo la idea de cómo va a funcionar el proceso ETL, se desarrollará estrategias con el fin de garantizar y aumentar el rendimiento del mismo, como por ejemplo:

- Administración de cambios en los atributos de las Dimensiones.
- Carga de Datos.
- Frecuencia de carga.
- Organización del Área Temporal ETL.
- Secuencia de los Procesos de Transformaciones.
- Registro y Manejo de Errores.

2.3.7.4.Paso 4 – Profundizar la Tabla Objetivo.

Se diseña la jerarquía de las tablas objetivo mediante gráficos, es decir, se le da un orden a los niveles de las Dimensiones, con lo que se puede apreciar la estructura desde un nivel macro hasta el más detallado.

Desarrollo de Carga Inicial

2.3.7.5.Paso 5 – Poblar las Tablas de Dimensiones con los Datos Históricos.

Las Dimensiones (tablas objetivo) serán pobladas por primera vez, para lo cual se implementarán actividades como:

- Extracción de Datos
- Transformaciones
- Carga de Datos.

2.3.7.6.Paso 6 – Realizar la Carga Histórica de la Tabla de Hechos.

Al igual que las Dimensiones, se realizará el poblado de los Hechos por primera instancia, tomando en cuenta los datos de las Dimensiones, y asignando adecuadamente los registros de las mismas a la de sus respectivos hechos.

De igual manera se implementará las actividades de extracción, transformación y de carga de datos según la necesidad.

Desarrollo de Carga Incremental.

2.3.7.7.Paso 7 – Procesamiento Incremental de la Dimensiones.

Las Dimensiones serán pobladas de manera incremental, es decir a partir de la segunda carga de datos. Al igual que el poblado inicial, se realizará las transformaciones, y carga de datos, con la diferencia ahora, que la extracción se realizará exclusivamente con datos nuevos o que hayan cambiado en el origen.

2.3.7.8.Paso 8 – Procesamiento Incremental de los Hechos.

Así mismo el poblado incremental se realizará a partir de la segunda carga a la tabla de hecho, y tomando en cuenta únicamente de los datos nuevos o que hayan sufrido cambios; y al igual que la carga inicial se implementa las demás actividades como transformaciones.

2.3.7.9.Paso 9 – Operación y Automatización del Sistema ETL.

Con el fin de garantizar la ejecución periódica del proceso ETL, es necesario automatizar los trabajos y transformaciones. Según la herramienta ETL que se haya seleccionado, esta implementará pasos para configurar tareas programadas, permitiendo que por cierto tiempo como por ejemplo cada día se proceda automáticamente con el poblado de datos hacia el Data Warehouse. Esto es de gran ayuda, en el sentido que no se necesitará la intervención humana para que se realice el proceso ETL.

Como entregable en la etapa de Diseño y Desarrollo ETL es el documento con las especificaciones de acuerdo a cada paso de la misma sobre el Proyecto, lo que muestra cómo funcionará el proceso ETL.

2.3.8. Diseño y Desarrollo de Aplicación de BI.

Lo fundamental aquí, es especificar y proporcionar la interfaz que será mostrada al usuario final, este tipo de aplicaciones permitirá visualizar los datos del Data Warehouse de manera interactiva.

Las aplicaciones de BI son el primer medio de entrega de los valores de un sistema de Data Warehouse dentro de una organización.

Para poder determinar la Aplicación se determinó lo siguiente:

2.3.8.1.Tipo de Aplicación de BI.

Permitirá establecer el tipo de aplicación de BI que se pretende desarrollar para el usuario final. Las aplicaciones por ejemplo pueden ser:

- Herramientas de reporte y consulta de acceso directo.
- Minería de datos (Data Mining).

- Aplicaciones analíticas.
- Paneles y Cuadros de Mando.

2.3.8.2.Preparar el Desarrollo de la Aplicación.

En este punto se prepara la herramienta de BI, para que pueda ejecutar sin problemas las necesidades de los usuarios.

Para el desarrollo de esta actividad se tomará en cuenta los siguientes puntos:

- Instalación de la herramienta.
- Poblar los metadatos de la herramienta de BI.
- Crear los metadatos del Negocio.
- Testear la herramienta de BI.

2.3.8.3.Crear las Aplicaciones

Primero se captura las especificaciones para la aplicación de BI después de completar la definición de los requerimientos. Posterior a esto se realiza el desarrollo de la aplicación de BI tomando en cuenta el tipo de aplicación a implementar.

En esta etapa, se realiza la herramienta de BI con las configuraciones realizadas para poder así pasar al despliegue.

2.3.9. Despliegue

Aquí se unen todas las piezas, como la tecnología a utilizar para los datos, las aplicaciones de BI, la arquitectura, estas convergen en el despliegue. En este instante, se realizará lo que se conoce en término inglés como *deploy* de las aplicaciones para dar como resultado el producto.

En esta etapa se realiza las siguientes actividades:

2.3.9.1.Despliegue ETL

Se realiza el deploy de todo el sistema ETL, como son las transformaciones y trabajos.

2.3.9.2.Despliegue de BI

Se realiza el despliegue de la herramienta de BI conjuntamente con las aplicaciones desarrolladas para la visualización de los datos.

En la etapa de despliegue se entregaran las aplicaciones ya disponibles para la ejecución, tanto del sistema ETL, como también de la herramienta y aplicaciones de BI.

2.3.10. Mantenimiento y Crecimiento

En el mantenimiento se verifica que esté funcionando de manera correcta en cada etapa del desarrollo del Data Warehouse.

Se realiza el Diagnostico de los resultados, las operaciones, en sí, es una pequeña administración para que se encuentre funcionando sin problemas.

En esta etapa no se presentan ninguna documentación debido a que en el caso de mantenimiento, se lo debe realizar constantemente con el fin de verificar el correcto desempeño del Data Warehouse, mientras que con el crecimiento en caso de suceder se debe realizar nuevamente todo el ciclo para adecuar los cambios que se pretenda realizar al Almacén y a todo el proceso BI.

2.4.Herramientas Para el Desarrollo de Data Warehouse.

Para la implementación, administración del Data Warehouse, así como para el análisis de la información y reporte se requiere de las siguientes herramientas.

- SQL Server 2008
- SQL Server Integration Service (SSIS)
- SQL Server Analysis Service (SSAS)
- SQL Server Reporting Service (SSRS)
- Microsoft Office 2010 (Excel 2010).

2.4.1. SQL Server 2008.

Motor de base de datos de SQL Server en su versión 2008, el cual se usara para implementar el Data Warehouse.

Servicio principal para almacenar, procesar y proteger datos, proporciona acceso controlado y procesamiento de transacciones rápido para cumplir con los requisitos de las aplicaciones de datos más exigentes.

2.4.2. SQL Server Integration Services (SSIS).

Es una plataforma que permite generar soluciones de integración de datos de alto rendimiento, entre las que se incluyen paquetes de extracción, transformación y carga (ETL) para el almacenamiento de datos.

Usos De Integration Services (SSIS)

Integration Services proporciona contenedores, tareas y transformaciones para generar Business Intelligence en paquetes de SSIS.

Los contenedores admiten la repetición de flujos de trabajo recorriendo archivos u objetos, y evaluando expresiones. Un paquete puede evaluar datos y repetir flujos de trabajo en función de los resultados.

Las tareas que usan parámetros de entrada también pueden generar Business Intelligence en paquetes. Por ejemplo, el valor de un parámetro de entrada puede filtrar los datos recuperados por una tarea.

También es posible enviar un conjunto de datos a varios destinos y aplicar a continuación distintos conjuntos de transformaciones a los mismos datos. Por ejemplo, un conjunto de transformaciones puede resumir los datos mientras otro conjunto de transformaciones expande los datos buscando valores en tablas de referencia y agregando datos de otros orígenes.

2.4.3. SQL Server Analysis Services (SSAS)

Ofrece funciones de procesamiento analítico en línea (OLAP) y minería de datos mediante una combinación de tecnologías de servidor y cliente, reforzadas por el uso de un entorno de desarrollo y administración especializada al que se une un modelo de objetos bien definido para diseñar, crear, implementar y mantener aplicaciones de Business Intelligence.

Trabajar con Procesamiento Analítico en Línea (Olap)

El procesamiento analítico en línea (OLAP) le permite obtener acceso a datos organizados y agregados de orígenes de datos empresariales, como por ejemplo almacenamientos de datos, en una estructura multidimensional denominada cubo.

Microsoft SQL Server 2008 Analysis Services (SSAS) proporciona herramientas y características para OLAP que puede utilizar para diseñar, implementar y mantener cubos y otros objetos compatibles.

Trabajar en Modo de Proyecto o en Modo en Línea.

Puede definir y crear instancias de objetos OLAP directamente en una base de datos de Analysis Services (se denomina modo en línea o, a veces, modo de conexión directa) o en un proyecto de Analysis Services (se denomina modo de proyecto). Normalmente, desarrollará la solución OLAP en un proyecto de Analysis Services.

Trabajar con Esquemas Relacionales

Por lo general, los objetos OLAP se definen a partir de un origen de datos existente, por ejemplo un almacén de datos o una base de datos de producción.

Definir y Configurar Dimensiones, Atributos y Jerarquías

Las dimensiones, atributos y jerarquías son objetos OLAP que se definen y configuran en el nivel de base de datos de Analysis Services. Estos objetos OLAP son independientes de los cubos OLAP y se pueden usar en uno o varios cubos.

Definir y Configurar Cubos y Propiedades de Cubo

"Los cubos son objetos OLAP que se componen de medidas y dimensiones relacionadas que se configuran en una base de datos de Analysis Services."**[8]**

Se puede definir y configurar varios cubos en una única base de datos; y cada cubo puede usar algunas de las mismas dimensiones o todas ellas. También puede definir un solo cubo que

contenga varios grupos de medida en la misma base de datos en lugar de definir cubos independientes.

Al definir un cubo, también se definen propiedades avanzadas del mismo: cálculos, KPI, acciones, particiones, agregaciones, perspectivas y traducciones.

2.4.4. SQL Server Reporting Service (SSRS)

Reporting Services es una plataforma de elaboración de informes basada en servidor que puede utilizar para crear y administrar informes tabulares, matriciales, de gráficos y de formato libre con datos extraídos de orígenes de datos relacionales y multidimensionales. Los informes que cree se pueden visualizar y administrar mediante una conexión basada en Web. Reporting Services contienelos componentes principales siguientes:

- Un conjunto completo de herramientas que se pueden utilizar para crear, administrar y ver informes.
- Un componente Servidor de informes que aloja y procesa informes en diversos formatos. Los formatos de salida incluyen HTML, PDF, TIFF, Excel, CSV, etc.
- Una API que permite a los programadores integrar o extender procesamientos de datos e informes en aplicaciones personalizadas o crear herramientas personalizadas para generar y administrar informes.

2.4.5. Microsoft Office Excel 2010

Microsoft Office Excel 2010 es una potente herramienta que le permite crear hojas de cálculo, analizar y compartir información con otros usuarios en un entorno manejable.

Mediante una interfaz de usuario completamente renovada y nueva vistas interactivas PivotTable, Excel 2010 es un front-end ideal para su actividad de Inteligencia Empresarial. Excel 2010 incorpora nuevas funciones de hipercubo y un mejor soporte para SQL Server 2008 Analysis Services, lo que le permitirá aumentar el valor de sus inversiones en Data Warehouse.

Ampliar las inversiones en inteligencia empresarial ya que Office Excel 2010 es totalmente compatible con Microsoft SQL Server 2008 Analysis Services.

Aprovecha la flexibilidad y las nuevas funciones de cubo de Office Excel 2010 para crear un informe personalizado a partir de una base de datos OLAP. La conexión a orígenes de datos también es más sencilla mediante la Biblioteca de conexiones de datos.

CAPITULO III: DESARROLLO E IMPLEMENTACIÓN DEL DATA WAREHOUSE PARA EL PADRÓN ELECTORAL DEPARTAMENTAL DE POTOSÍ

Para el desarrollo del Data Warehouse, se utilizará como guía la Metodología de Ralph Kimball tratada en el Capítulo II, sección 2.3. Esto permitirá realizar el proyecto en base a pasos que garantizan la calidad del mismo. Esta metodología será adaptada a las necesidades del Proyecto de Tesis.

3.1.Planificación del Proyecto.

3.1.1. Definición del Proyecto.

Para el desarrollo del Proyecto en base al Anexo 1, se realizó la entrevista a vocales y directores el cual nos ayudó a identificar el alto interés por parte de los niveles ejecutivos y de los diferentes directores su interés por obtener información que les permita tomar decisiones.

En Base a la factibilidad el proyecto resulta factible ya que el Tribunal Electoral Departamental de Potosí posee licencias del motor de base de datos Sql server 2008 y por ende las herramientas para el desarrollo del Data Warehouse.

3.1.2. Alcance y Justificación del Proyecto.

En cuanto a estos puntos, ya fueron definidos en la introducción del proyecto. Pero vale recalcar de manera puntual, el alcance de proyecto, que es el desarrollo un Data Warehouse para generar información analítica. Su justificación, es de contar con información tratada y útil para el soporte de decisiones en función al análisis de la información.

Para la Planeación del Proyecto se definió los roles a cumplir en el desarrollo del proyecto, estos son:

- Desarrollador del Proyecto. Persona única la cual estará encargada de:
    - Desarrollar.
    - Monitorear el Desarrollo.
    - Diseñar y Modelar.
    - Analizar el Desarrollo del Proyecto.
- Jefes Departamentales. Personas con alto rango en la institución a las cuales se les va a ir notificando periódicamente los avances del Proyecto

- Vocales.
- Director Departamento de Informática.

3.2.Definición de los Requerimientos del Negocio.

Para la definición de los requerimientos se realizaron 3 actividades que ayudaron a determinar las necesidades de los usuarios de la institución, estas son:

- Realizar conjuntamente con el Director de Informática la petición de información de cada una de las áreas involucradas, acerca de qué tipo de información necesitan y tratan sobre el Padrón Electoral.
- Reunión con los responsables de las diferentes Direcciones.
- Realizar el estudio de los reportes implementados en los Sistemas actuales de la institución, Sistemas relacionados a los procesos electorales. En este punto vale recalcar que se dio más énfasis a los reportes estadísticos, ya que a través de ellos se puede recopilar información relevante para el negocio a evaluar.

Tomando en cuenta los puntos anteriores se obtuvo los requerimientos que se mencionan a continuación.

3.2.1. Estadísticas.

Las direcciones del Tribunal Electoral Departamental de Potosí supieron dar a conocer la información que interesa y sobre la cual se toma decisiones, estas son:

- Habilitados por Circunscripción.
- Mesas por Circunscripción.
- Habilitados por Asiento Electoral.
- Mesas por Asiento Electoral.
- Cantidad de Notarios por Circunscripción.
- Cantidad de Notarios por Asiento Electoral.
- Habilitados por sexo.
- Habilitados por Provincia.
- Habilitados por edades.
- Inscritos por fechas.
- Cantidad de Mesas, Asientos Electorales, Notarios por Circunscripción.

- Detalle de Inscritos y Habilitados por Asiento Electoral.
- Detalle de Inscritos y Habilitados por Municipio.

  A estos requerimientos, se especificó un espacio de tiempo, que según la toma de requerimientos fueron:
- Anual.
- Cuatrimestral.
- Mensual.
- Diario.

En función a lo planteado, para el Diseño y Desarrollo del data Warehouse se contara con todo lo expuesto anteriormente.

3.3.Diseño Técnico de la Arquitectura.

En el Diseño de la Arquitectura se definió dos ambientes, el primero es el Back Room y el segundo que es Front Room, los mismos que serán detallados a continuación.

3.3.1. Back Room.

En este ambiente se especificará el ambiente del proceso ETL, así como los orígenes de datos. Como se puede apreciar en la Figura 3.1, se detalla el Diseño que se realizó para la implementación del Back Room, que consta desde el Origen de los Datos hasta el almacenamiento en el Data Warehouse

Origen De Datos.

La base de datos Origen se encuentra sobre la plataforma de Sql Server 2005. Como se puede observar en la Figura 3.1 se estableció los orígenes de datos de Geografía Electoral y Padrón Electoral. Vale recalcar que en la base de Geografía Electoral se encuentra todo los datos geográficos del departamento en cambio en la Base de Datos del Padrón Electoral se encuentra almacenada la información demográfica de los ciudadanos inscritos.

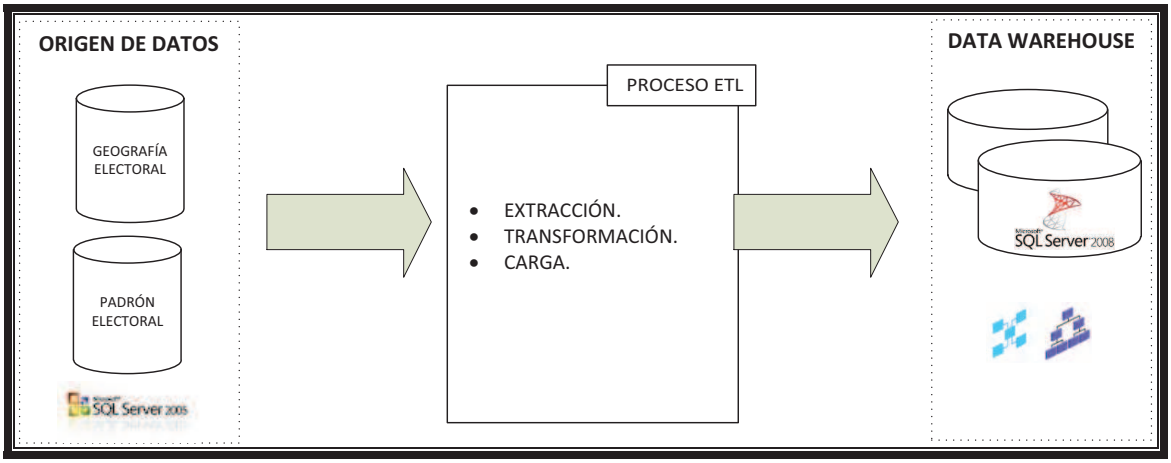


Figura 3.1: Back Room – Proyecto

Fuente: Propia

ETL.

En cuanto al proceso ETL se realizará el Mapeo de los datos Origen, que serán especificados posteriormente en el Modelamiento Dimensional, después se desarrollará el proceso de extracción, seguido de la transformación de los datos Origen y finalmente se procederá con la carga hacia el Data Warehouse.

Se especificará con más detalle el Proceso ETL y los Metadatos del Proyecto en la sección de Diseño y Desarrollo ETL.

Destino de datos (Data Warehouse)

El destino que es el Almacén de datos en Sql Server 2008, mantendrá el modelo dimensional que será especificado en la sección Diseño y Modelamiento Dimensional.

3.3.2. Front Room.

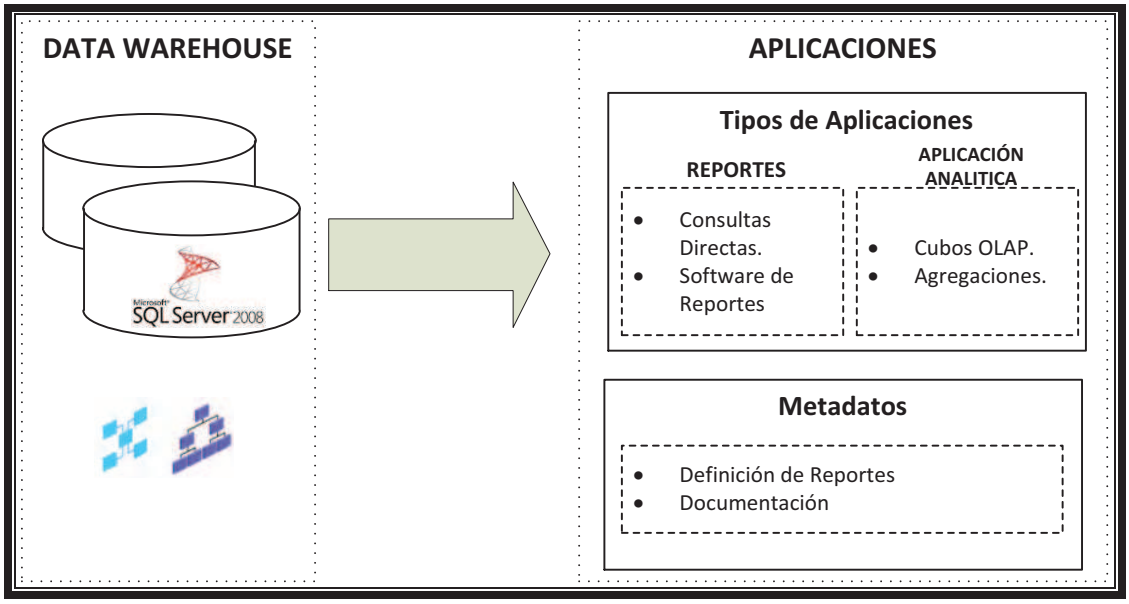


Figura 3.2: Front Room – Proyecto

Fuente: Propia

En este punto se especificó la arquitectura de presentación del Data Warehouse, mediante aplicaciones que le servirán al usuario.

Como se observa en la figura 3.2 en el Data Warehouse donde se encuentra los datos que fueron tratados y almacenados posteriormente se procede a la presentación de acuerdo al tipo de aplicación del usuario final.

3.4.Selección del Producto.

Para la selección de los productos a utilizar se tomó en cuenta las características planteadas anteriormente en el Capítulo II - Sección 2.4 sobre las Herramientas para el Desarrollo de un Data Warehouse, con lo cual se planteó las siguientes herramientas a ser utilizadas en el Proyecto:

3.4.1. Plataforma DBMS.

La plataforma que se utilizo es el Motor de base de datos Sql Server 2008 version 10.0.1600.22 ((SQL_PreRelease).080709-1414) ya que el Tribunal Electoral Departamental cuenta con la licencia y su estructura de base de datos sobre la plataforma Sql Server 2005.

Vale recalcar que este motor tiene incorporado la funcionalidad de la tecnología OLAP, por lo que aparte de ser el motor de base de datos, será también un servidor o base de datos OLAP.

3.4.2. Herramienta ETL.

La herramienta ETL que se utilizó fue SQL Server Integration Services (SSIS) la cual permite generar soluciones de integración de datos de alto rendimiento, entre las que se incluyen paquetes de extracción, transformación y carga (ETL) para el almacenamiento de datos.

3.4.3. Herramienta de BI.

La herramienta de BI que se seleccionó para la presentación de la información mediante cubos OLAP es SQL Server Analysis Services (SSAS) el cual ofrece funciones de procesamiento analítico en línea (OLAP) y minería de datos mediante una combinación de tecnologías cliente servidor.

3.5.Modelamiento Dimensional.

Para poder realizar el Modelamiento del almacén de datos primero se analizó el modelo relacional de los datos origen tanto de la Base de Datos de Geografía Electoral como el Padrón Electoral.

3.5.1. Paso1. Selección del Proceso de Negocio.

Los requerimientos del presente proyecto de negocio están enfocados a:

- Nuevos Inscritos.
- Habilitados.
- Depurados.

3.5.2. Paso2. Definición de la Granularidad.

En cuanto al nivel de la granularidad para el proceso de negocio se definió en las dimensiones de geografía y tiempo.

3.5.3. Paso3. Identificar las Dimensiones y Mapear los Datos Origen.

Las dimensiones definidas soportan los requerimientos establecidos con la granularidad específica para cada negocio. El mapeo de datos está especificado para cada dimensión. Para el presente proyecto se definieron las dimensiones que se detallan a continuación.

3.5.3.1.Dimensión Tiempo.

Contiene la información de tiempo a partir del inicio de la inscripción al Padrón Electoral biométrico hasta la última inscripción a la misma.

Nombre de la Tabla:	DimTiempo
Tipo de Tabla:	Dimensión
Esquema Origen:	Ninguno
Origen:	Generado con programación sql.

DimTiempo

Nombre de columna	Tipo de datos	Permitir valores NULL
TimeID	int	☐
Fecha	date	☑
Anio	varchar(4)	☑
MesID	int	☑
Mes	varchar(10)	☑
Cuarto	int	☑
Semana	int	☑
NumDiaSemana	int	☑
DiaSemana	varchar(10)	☑
		☐

Figura 3.3: Tabla de dimensión Tiempo

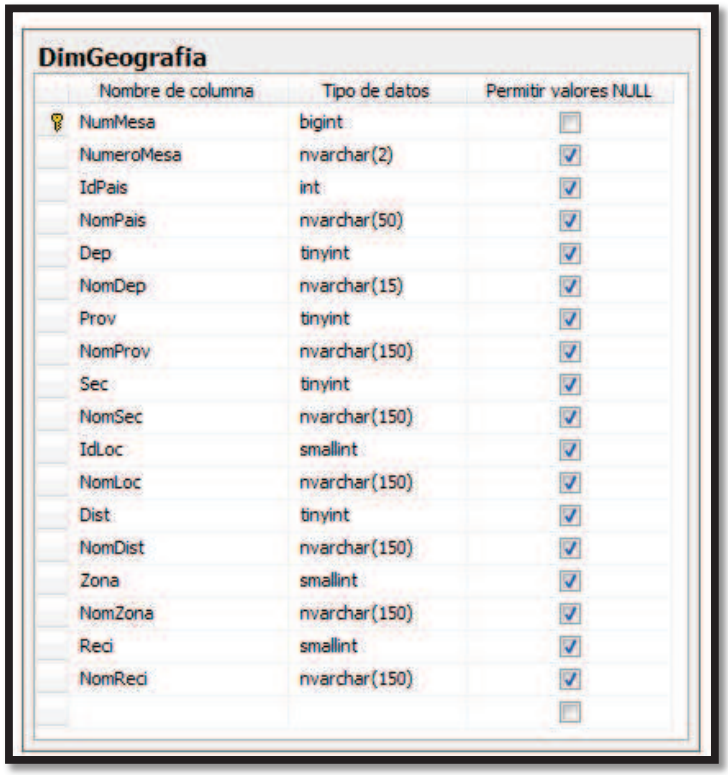
Fuente: Propia

3.5.3.2.Dimensión Geografía.

Contiene la Información Geográfica que maneja el Tribunal Departamental Electoral.

Nombre de la Tabla:	DimGeografia
Tipo de Tabla:	Dimensión
Esquema Origen:	GeografiaElectoral

Tablas de Origen: Dep, Pais, Prov, Sec, Loc, Dist, Zona, Reci, Mesa



DimGeografia

Nombre de columna	Tipo de datos	Permitir valores NULL
NumMesa	bigint	☐
NumeroMesa	nvarchar(2)	☑
IdPais	int	☑
NomPais	nvarchar(50)	☑
Dep	tinyint	☑
NomDep	nvarchar(15)	☑
Prov	tinyint	☑
NomProv	nvarchar(150)	☑
Sec	tinyint	☑
NomSec	nvarchar(150)	☑
IdLoc	smallint	☑
NomLoc	nvarchar(150)	☑
Dist	tinyint	☑
NomDist	nvarchar(150)	☑
Zona	smallint	☑
NomZona	nvarchar(150)	☑
Reci	smallint	☑
NomReci	nvarchar(150)	☑
		☐

Figura 3.4: Tabla de dimensión Geografía

Fuente: Propia

3.5.3.3.Dimensión Grado de Instrucción.

Contiene la Información del grado de instrucción que maneja el Tribunal Departamental Electoral

Nombre de la Tabla:	DimGradoInstruccion
Tipo de Tabla:	Dimensión
Esquema Origen:	PadronElectoral
Tablas de Origen:	GradoInstruccion

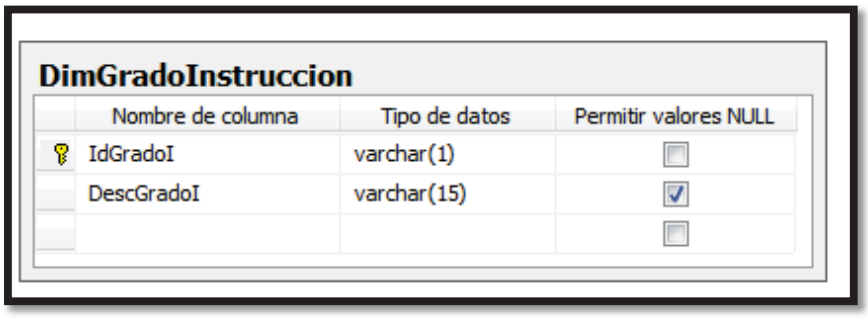


Figura 3.5: Tabla de dimensión Grado de Instrucción

Fuente: Propia

3.5.3.4.Dimensión Estado Civil.

Contiene la Información del estado civil de los ciudadanos inscritos en el Padrón Electoral.

Nombre de la Tabla:	DimEstadoCivil
Tipo de Tabla:	Dimensión
Esquema Origen:	PadronElectoral
Tablas de Origen:	EstadoCivil

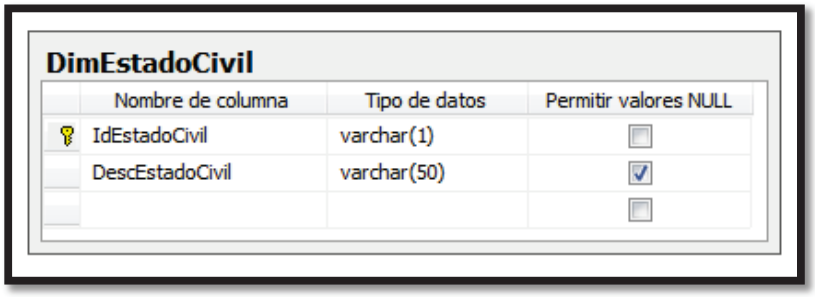


Figura 3.6: Tabla de dimensión Estado Civil

Fuente: Propia

3.5.3.5.Dimensión Sexo.

Contiene la Información del sexo de los ciudadanos inscritos en el Padrón Electoral.

Nombre de la Tabla:	DimSexo
Tipo de Tabla:	Dimensión

Esquema Origen: PadronElectoral

Tablas de Origen: Sexo

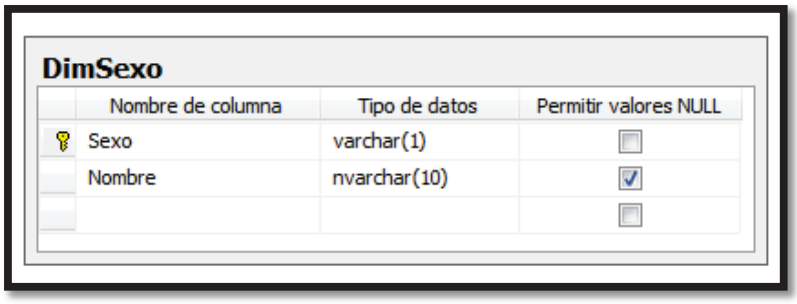


Figura 3.7: Tabla de dimensión Sexo

Fuente: Propia

3.5.3.6.Dimensión Tipo Cedula.

Contiene la Información del tipo de cedulas válidas para la inscripción en el Padrón Electoral.

Nombre de la Tabla: DimTipoCedula

Tipo de Tabla: Dimensión

Esquema Origen: PadronElectoral

Tablas de Origen: TipoCedula

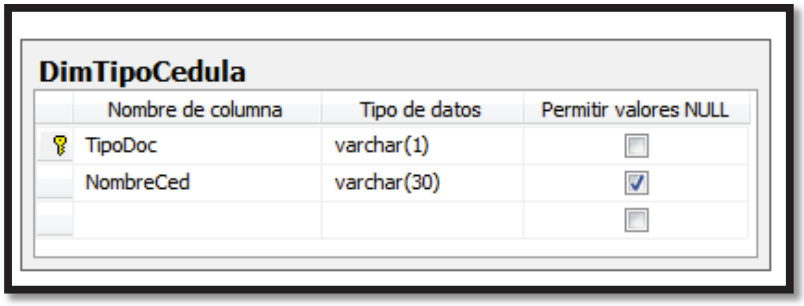


Figura 3.8: Tabla de dimensión Tipo Cedula

Fuente: Propia

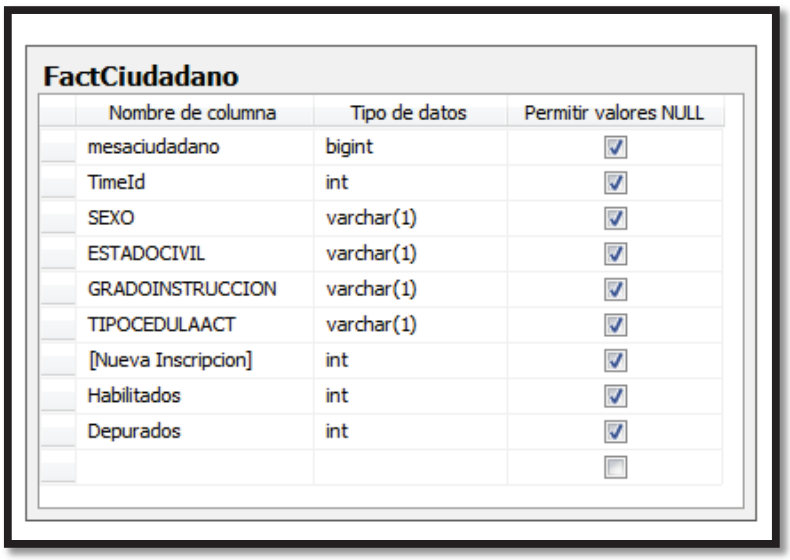
3.5.4. Paso4. Identificar los Hechos y Mapear los Datos Origen.

Los hechos que se definieron soportan los requerimientos establecidos cumpliendo con la granularidad específica. El mapeo de los datos es especificado en la tabla de hechos.

3.5.4.1.Ciudadano.

Contiene las medidas de las transacciones que se realiza en base al proceso de negocio, y las claves foráneas que permite relacionar la tabla de hechos con las dimensiones.

Nombre de la Tabla:	FactCiudadano
Tipo de Tabla:	Hecho
Esquema Origen:	PadronElectoral
Tablas de Origen:	Ciudadano



FactCiudadano

Nombre de columna	Tipo de datos	Permitir valores NULL
mesaciudadano	bigint	☑
TimeId	int	☑
SEXO	varchar(1)	☑
ESTADOCIVIL	varchar(1)	☑
GRADOINSTRUCCION	varchar(1)	☑
TIPOCEDULAACT	varchar(1)	☑
[Nueva Inscripcion]	int	☑
Habilitados	int	☑
Depurados	int	☑
		☐

Figura 3.9: Tabla de Hechos Ciudadano

Fuente: Propia

3.5.5. Data Warehouse Bus Matrix

	Dimensión					
Procesos de Negocio	Estado civil	Geografia	Grado Instrucción	Sexo	Tiempo	Tipo Cedula
Nuevos Inscritos	x	x	x	x	x	x
Habilitados	x	x	x	x	x	x
Depurados	x	x	x	x	x	x

Figura 3.10: Data Warehouse Bus Matrix Detallado

Fuente: Propia

3.6.Diseño Físico.

El diseño físico se realizó en función a la Base de datos Sql Server 2008, los scripts para la creación de la Base de Datos y tablas del presente proyecto se encuentra en el Anexo 2 (Creación Base de Datos PADRONELECTORALDW además del script para el llenado de la tabla tiempo.

3.7.Diseño y Desarrollo del Proceso ETL.

A continuación se elabora un mapa gráfico, el cual permitirá tener una mayor idea de cómo se ejecutara el Proceso ETL.

3.7.1. Paso1. Trazar un Plan de Alto Nivel.

3.7.1.1.Geografía.

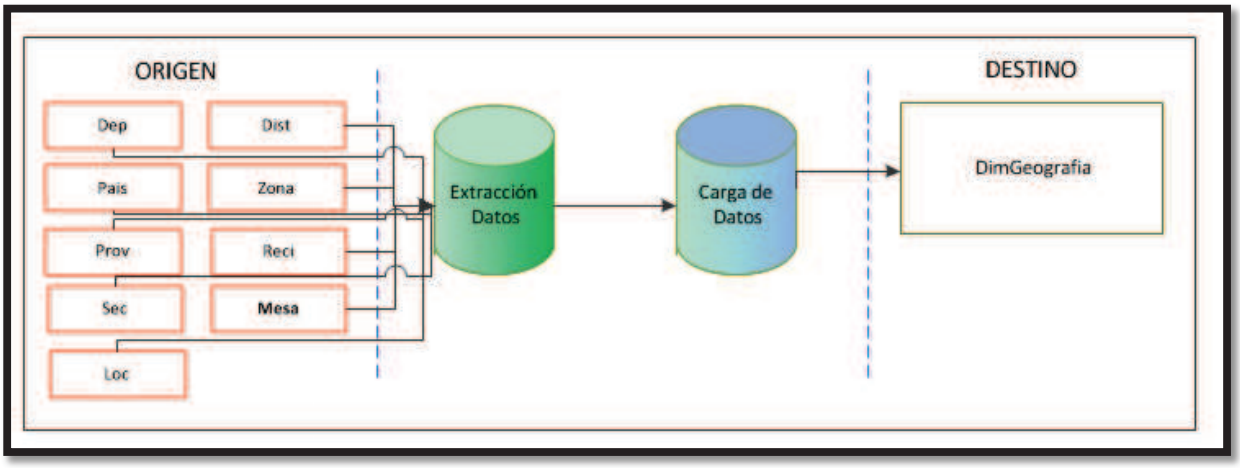


Figura 3.11: Plan ETL Geografía

Fuente: Propia

3.7.1.2.Estado Civil.

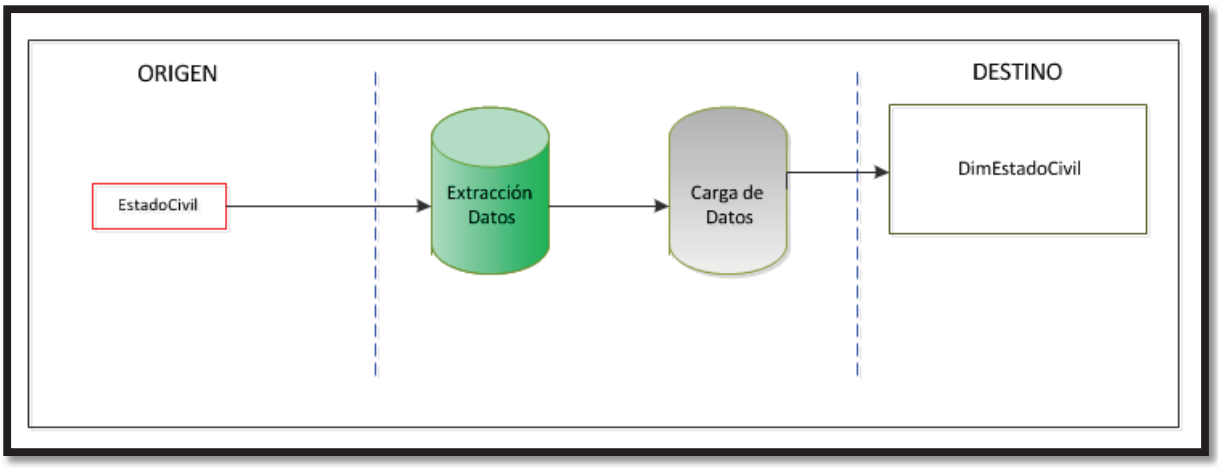


Figura 3.12: Plan ETL Estado Civil

Fuente: Propia

3.7.1.3.Grado Instrucción.

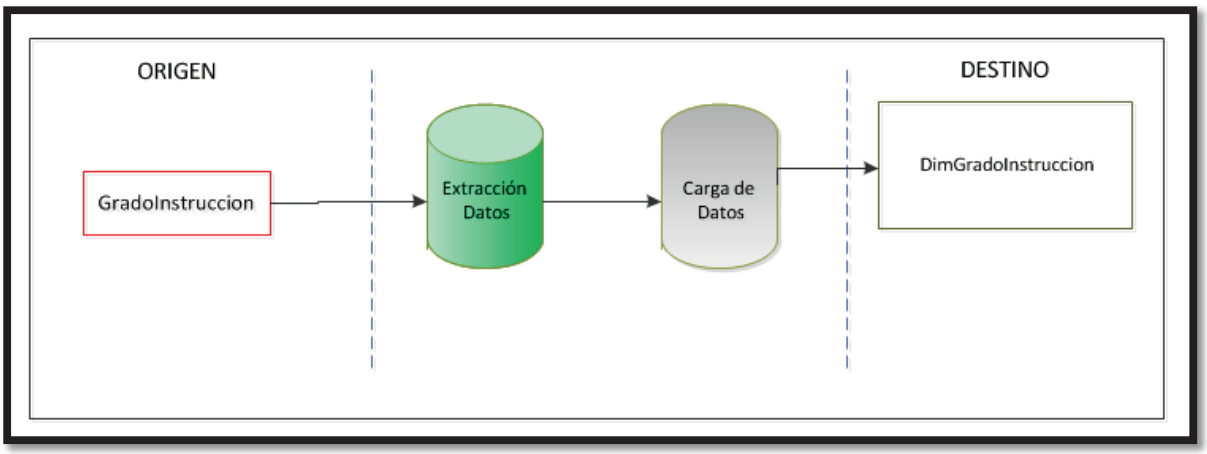


Figura 3.13: Plan ETL Grado de Instrucción

Fuente: Propia

3.7.1.4.Sexo.

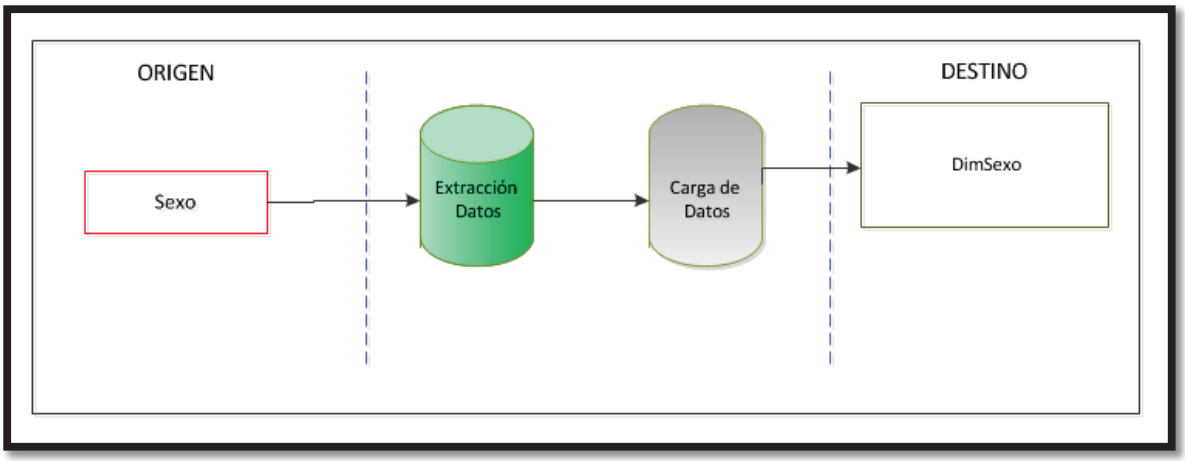


Figura 3.14: Plan ETL Sexo

Fuente: Propia

3.7.1.5.Tipo Cedula.

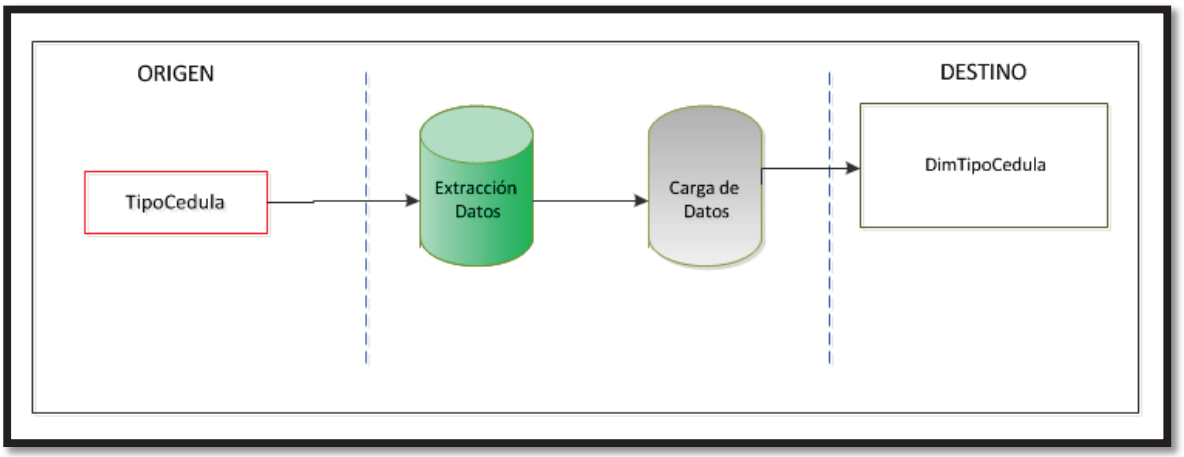


Figura 3.15: Plan ETL Tipo de Cedula

Fuente: Propia

3.7.1.6.Ciudadano.

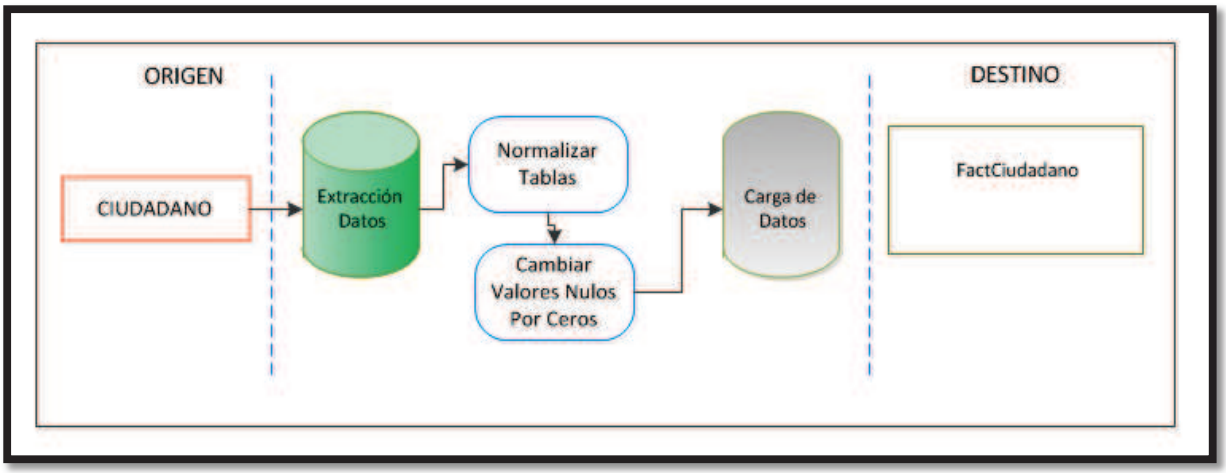


Figura 3.16: Plan ETL Ciudadano

Fuente: Propia

3.7.2. Paso2. Selección de la Herramienta ETL.

La herramienta específica que se utilizó para la elaboración de los procesos ETL fue SQL Server Integration Services.

3.7.3. Paso3. Desarrollar Estrategias por Defecto.

Se definió las siguientes estrategias:

- La extracción de datos se lo realiza desde las Bases de Datos de Padrón Electoral y Geografía Electoral, las cuales fueron adjuntados en una máquina virtual, esto debido a que la información del Padrón Electoral solo realiza cambios en los procesos de inscripción o depuración del Padrón Electoral. Esto ayudara en la carga de los datos en la Base de Datos de PadronElectoraDW.
- Tanto datos de las tablas de dimensiones como de hechos serán borrados y cargados con los datos actuales esto ejecutando el agente de SQL server el mismo que llama a los paquetes de SQL Server Integration Server.
- En cuanto a la Carga de datos no se tiene fechas específicas, debido a que no se sabe con precisión cuando se realizara un proceso de inscripción o depuración en el Padrón Electoral es por esta razón, que la tarea puede ser ejecutada cuando el administrador del sistema así lo requiera.

3.7.4. Paso4. Profundizar la Tabla Objetivo.

A continuación se muestra de manera específica las relaciones entre tabla de dimensiones y de hechos, como también las jerarquías que se definieron en algunas tablas de dimensiones.

Relación entre tabla de dimensiones y de hechos

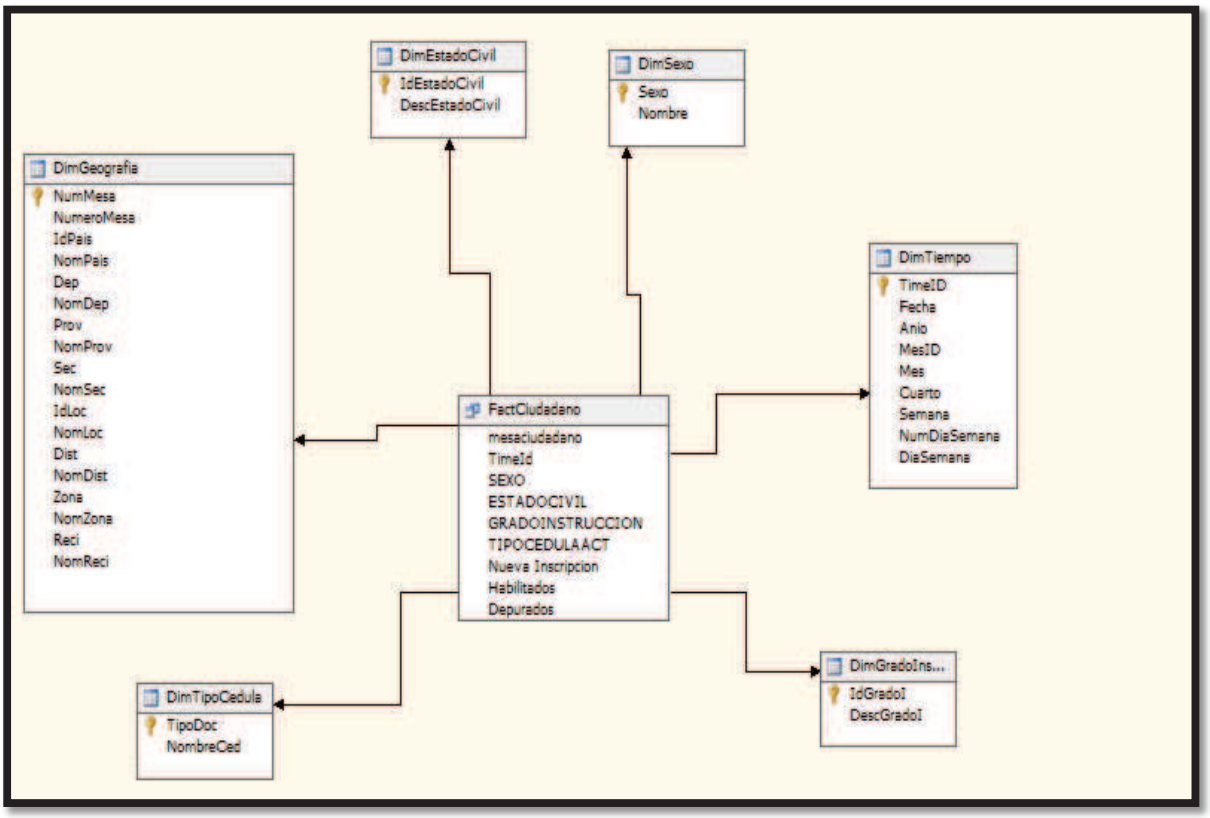


Figura 3.17: Relación entre la Tabla de Dimensiones y de Hechos

Fuente: Propia

Jerarquía Geográfica.



Figura 3.18: Jerarquía Tabla de Dimensión de Geografía

Fuente: Propia

Jerarquía Tiempo.

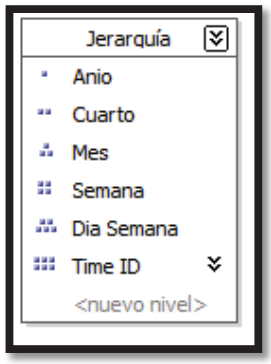


Figura 3.19: Jerarquía Tabla de Dimensión de Tiempo

Fuente: Propia

3.7.5. Paso5. Poblar las Tablas de Dimensiones Con los Datos Históricos.

La extracción de los datos se lo realizo mediante sentencias SQL a las bases de datos de Origen, como resultado devolverá todos los registros de acuerdo a las tablas consultadas.

Las consultas desarrolladas para la extracción de los datos se encuentran descritas en el Anexo 3 (Extracción de Datos Origen)

3.7.6. Paso6. Realizar la Carga Histórica de la Tabla de Hechos.

La carga de datos en la tabla de hechos se la realizo con SQL Server Integration Service. La extracción de datos de la tabla de hechos se encuentra descrita en el Anexo 3 (Extracción de Datos Origen sección ciudadano).

La transformación de los datos antes de ser cargados fueron transformados según los procesos del negocio.

La Carga de los datos se lo realizo en base a la transformación de los datos con cambios en los valores nulos por ceros.

3.7.7. Paso7. Procesamiento Incremental de las Dimensiones.

La carga incremental en las dimensiones se realiza con el borrado de las tablas de dimensiones y la extracción, transformación de las bases de datos del Padrón Electoral y Geografía Electoral para luego concluir con la carga a la Base de Datos del Data Wherehouse.

3.7.8. Paso8. Procesamiento Incremental de los Hechos.

La carga incremental en la tabla de hecho se realiza con el borrado de la tabla de hecho y la extracción, transformación de la base de datos del Padrón Electoral para luego concluir con la carga a la Base de Datos del Data Wherehouse.

3.7.9. Paso9. Operación y Automatización del Sistema ETL.

Se realizó dos actividades para la automatización del sistema ETL.

Programación de Paquetes.

Con la ayuda de SQL Server Integration Service se programó los paquetes para la tabla de hechos y las de dimensiones.

Programación de Trabajos

Se programó trabajos en la sección del agente de Sql Server el mismo que puede ser ejecutado cuando así lo requiera el administrador del sistema.

3.8.Diseño y Desarrollo de la Aplicación de BI.

3.8.1. Tipo de Aplicación de BI.

La aplicación de BI está enfocada a los reportes estándar y cubos de información, este tipo de aplicación está enfocado para presentar la información mediante reportes diseñados, o los cubos de información para visualizar los datos (medidas e indicadores) de manera multidimensional permitiendo la navegabilidad entre dimensiones.

3.8.2. Preparar el Desarrollo de la Aplicación.

Se configuro Sql Server Analysis Services y se procesó el cubo para poder acceder a los datos del Data Warehouse, para que la herramienta presente los datos en una capa amigable para el usuario.

3.8.3. Crear la Aplicación.

Luego de configurar Sql Server Analysis Service, se procedió a la generación de reportes de acuerdo a los requerimientos del negocio, esto se detalla de manera gráfica a continuación.

3.8.3.1.Desarrollo de la Aplicación BI con SSAS.

Para la creación del Data Warehouse para el Padrón Electoral de Potosí se tomó en cuenta los siguientes puntos**:**

- **Orígenes de Datos.**

Se creó la conexión hacia el origen de datos PadronElectoralDW

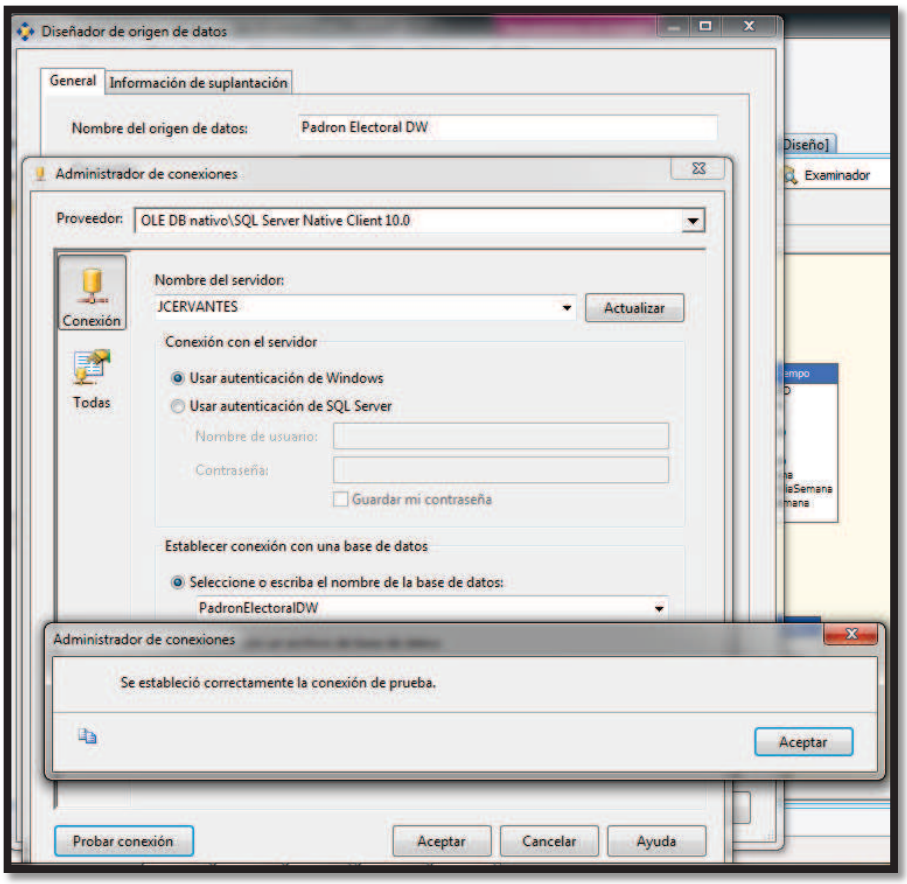


Figura 3.20: Configuración de Origen de Datos

Fuente: Propia

- **Vista de Origen de Datos.**

Se diseñó la vista de origen de datos con las tablas de dimensiones y hechos el mismo que se muestra a continuación.

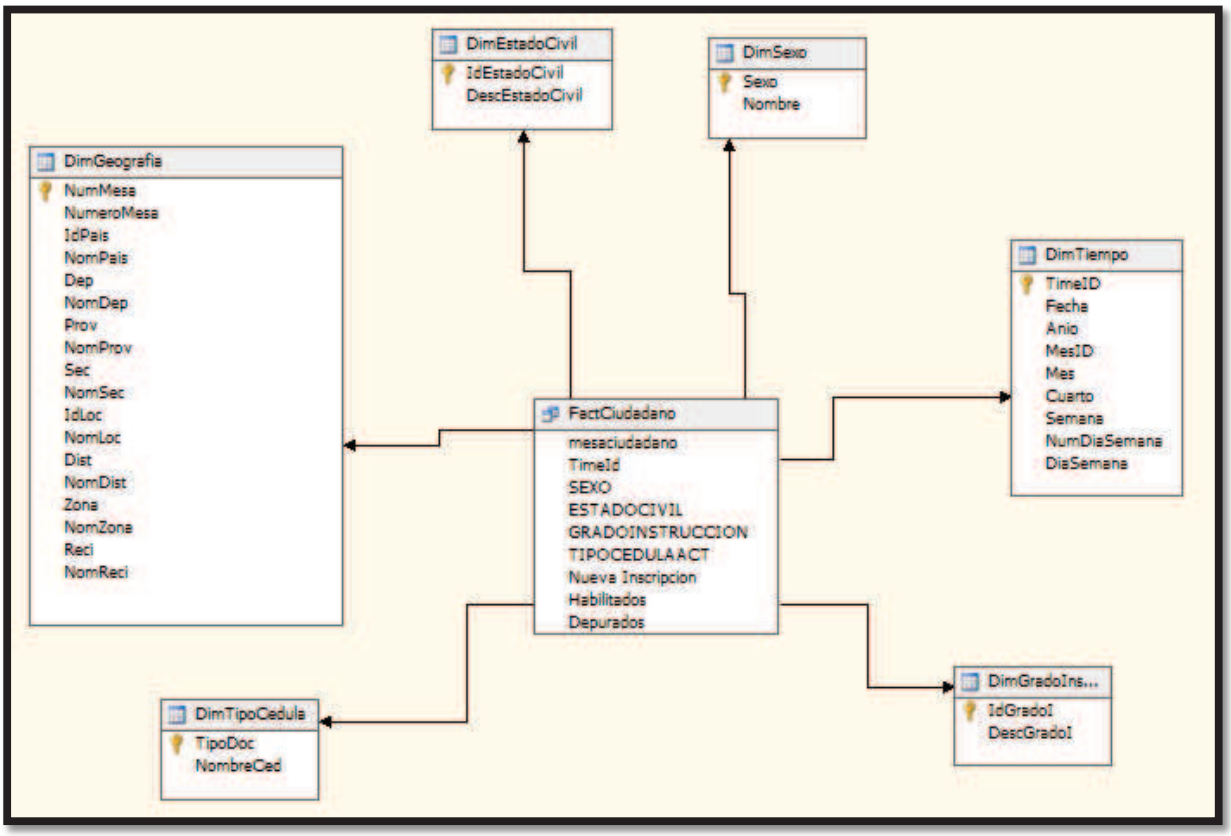


Figura 3.21: Vista de Orígenes de Datos Creados

Fuente: Propia

- **Dimensiones.**

Se agregaron las dimensiones

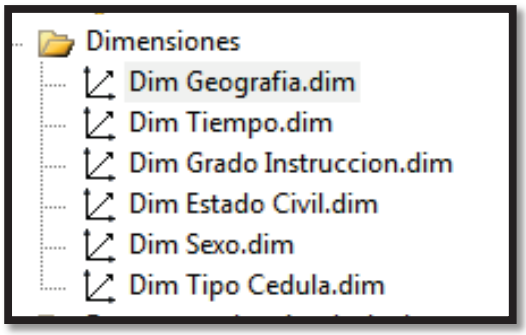


Figura 3.22: Dimensiones Agregadas

Fuente: Propia

Con las siguientes jerarquías.

- **Jerarquía de geografía**.

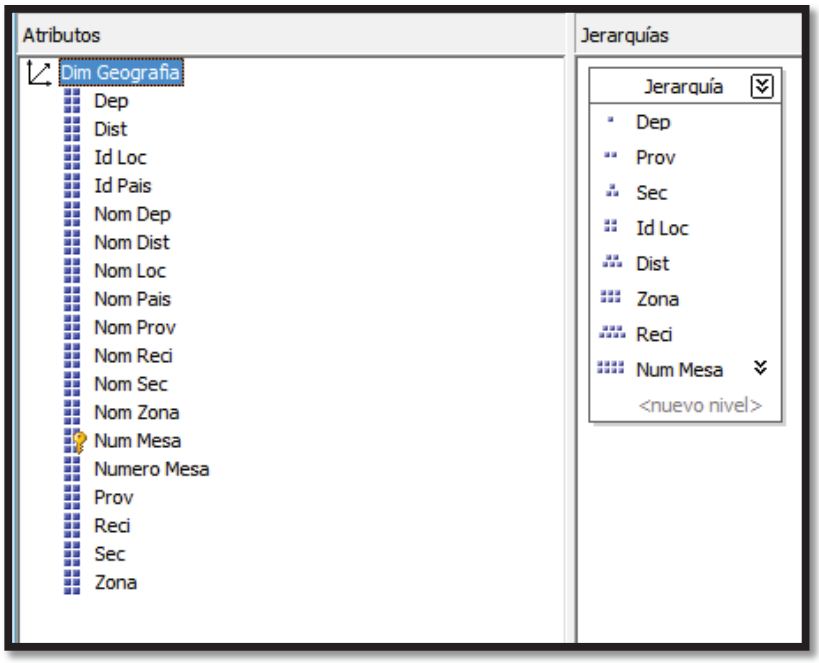


Figura 3.23: Dimensión de Geografía con su jerarquía específica

Fuente: Propia

- **Jerarquía Tiempo.**

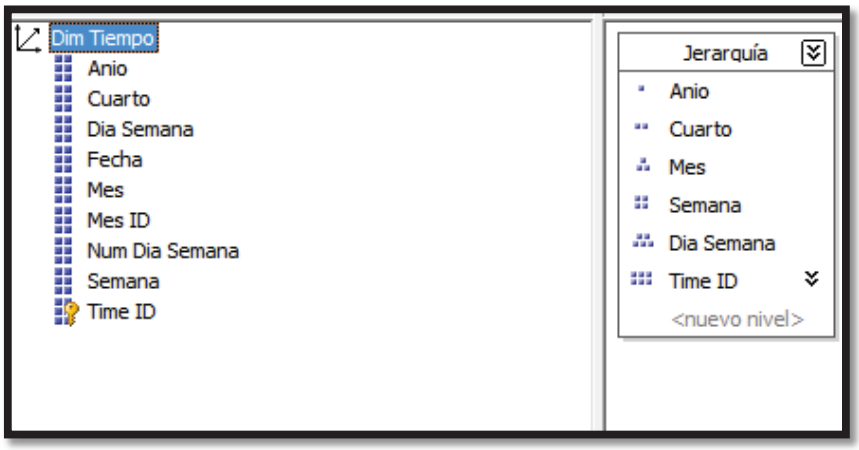


Figura 3.24: Dimensión de Tiempo con su jerarquía específica

Fuente: Propia

- **Cubos.**

Se creó el cubo en base a las tablas de dimensiones y hechos el mismo que presenta las medidas de acuerdo al requerimiento del negocio.

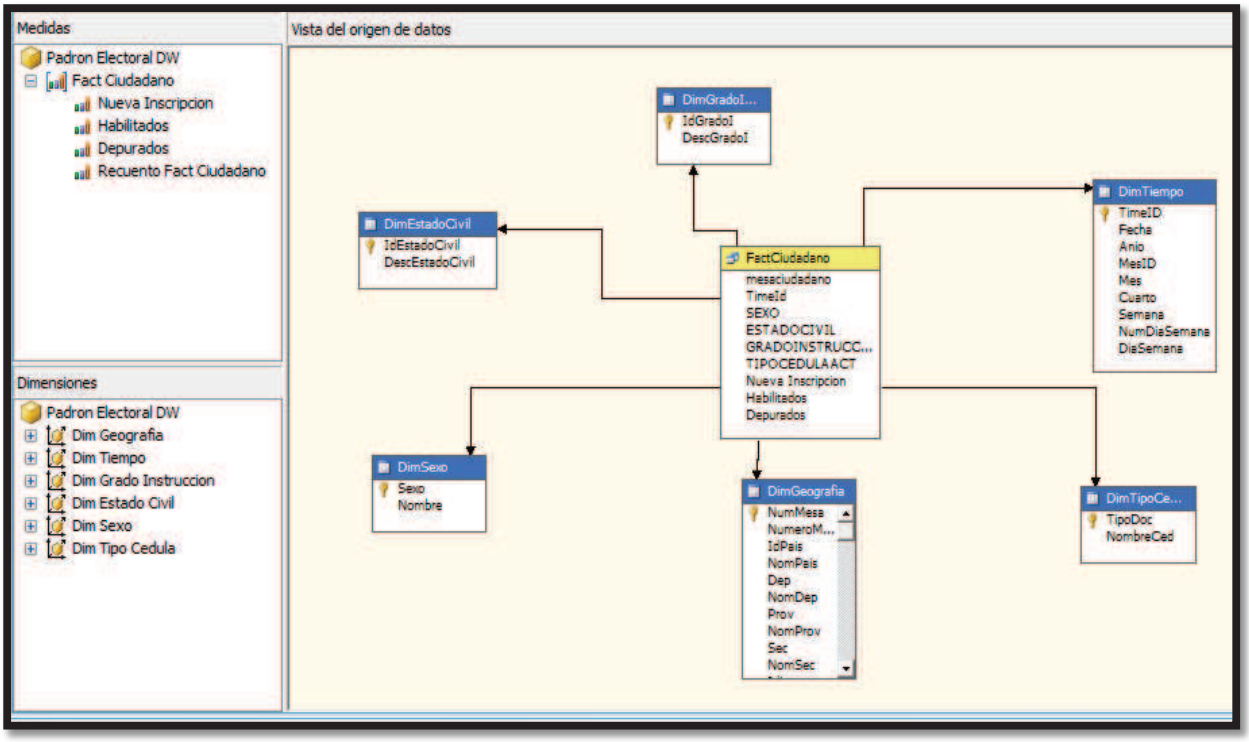


Figura 3.25: Cubo del Negocio Generado

Fuente: Propia

3.9.Despliegue.

Al momento de realizar el despliegue ETL como de BI se realizó las actividades en la Base de Datos del Padrón Electoral y Geografía Electoral como también en la Base de Datos destino utilizado para el Data Warehouse.

Adicionalmente y con el fin de poner en producción el Data Warehouse se debe tomar en cuenta:

- **Recurso Humanos**
    - Administrador de Base de Datos.

- Personal capaz de recolectar requerimientos a niveles gerenciales.
- Personal técnico capacitado en el desarrollo y administración de sistemas Data Warehouse.

- **Recursos De Hardware.**
  - Servidor de Base de Datos SQL Server el mismo que permitirá soportar diferentes transacciones, con almacenamiento superior a 500 debido a que la cantidad de datos a almacenar ira en crecimiento**.**

- **Recursos De Software.**
  - Manejador de Base de Datos.
  - Herramientas SSIS, SSAS, SSRS para el proceso de manipulación de los datos.
  - Herramientas para presentación de información o reportes a usuarios finales.

3.9.1. Despliegue ETL.

Creación del Repositorio.

Se realizó la creación tanto de la Base de Datos como de las tablas del Data Warehouse con los script que se encuentran en el anexo 2 (Creación De Base De Datos Data Warehouse PadronElectoralDW)

Desplegar ETL.

Se programaron los ETL en SSIS, se crearon para este cometido un paquete para la tabla de hechos y cinco paquetes para las tablas de dimensiones.

Verificación de ETL

Para la verificación de la Extracción, transformación y Carga puede ver el monitor de actividad de trabajo del Agente de SQL Server, ya que los paquetes son llamados a partir del trabajo PadronElectoralDWLoad en el Agente de SQL Server.

3.9.2. Despliegue BI.

Esta fase está orientada al correcto funcionamiento de las diferentes partes del Data Warehouse. En ella existen la tecnología, los datos y las aplicaciones de usuarios finales con la máxima eficacia y eficiencia posible. Es esencial que exista un buen soporte técnico que tenga una gran capacidad de respuesta ante los retos que puedan surgir una vez que se ha instalado el sistema, en forma de incidencias o dudas que puedan tener los usuarios del mismo. Puesto que este punto se refiere a una fase en la que el sistema está completamente terminado. Se exhibe como la presentación al usuario, alguno de los reportes desarrollados en Excel y en Reporting Service de acuerdo a los requerimientos que se recolectaron.

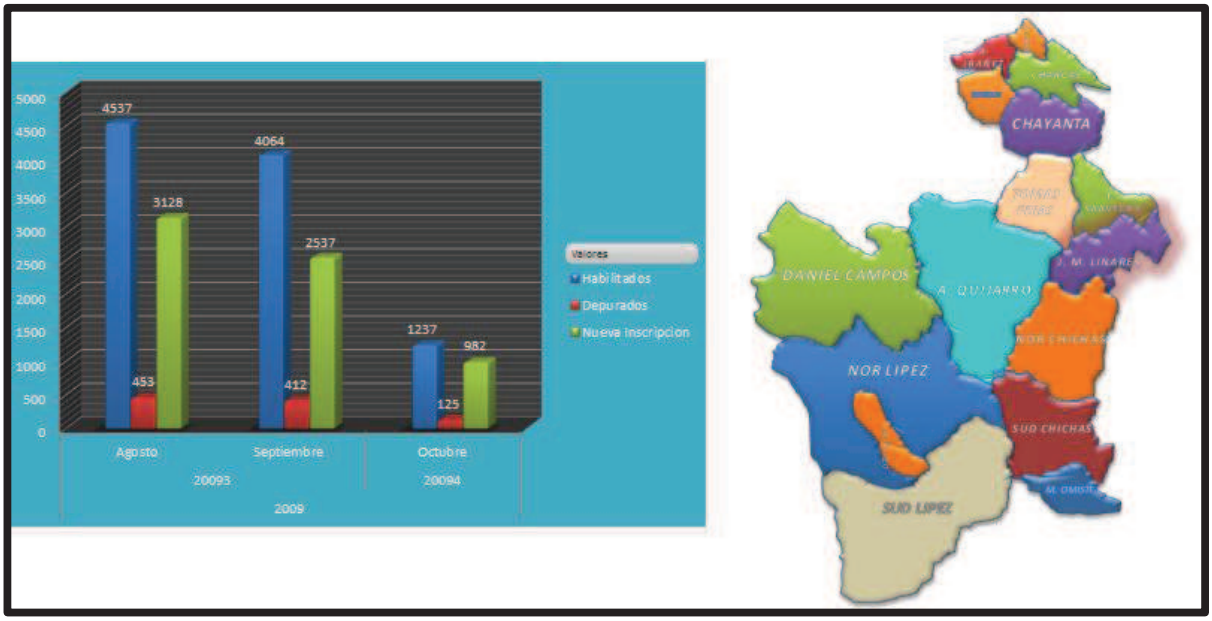


Figura 3.26: Reporte Generado con Reporting Service

Fuente: Propia

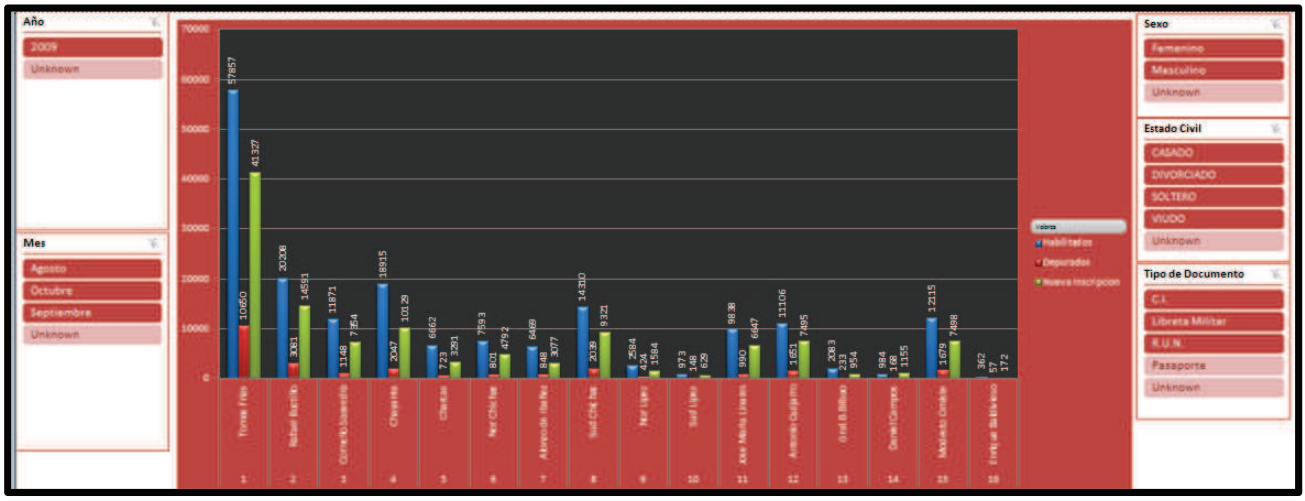


Figura 3.27: Reporte Generado con Excel 2010

Fuente: Propia

3.10. Mantenimiento Y Crecimiento.

Al igual que ocurría en el apartado anterior esta fase se refiere a la operativa del sistema una vez desplegado y, por tanto, este proyecto no abarca de manera explícita su mantenimiento y crecimiento. La creación de un Data Warehouse es un proceso que acompaña a la evolución de la organización durante toda su historia. Se necesita continuar con las actualizaciones de forma constante para poder seguir la evolución de las metas por conseguir. Sin embargo, al contrario de los sistemas tradicionales, los cambios en el desarrollo deben ser vistos como signos de éxito. Es importante establecer unas prioridades para poder manejar los nuevos requerimientos de los usuarios y de esa forma poder evolucionar y crecer.

.

CAPITULO IV: RESULTADOS

Una vez desarrollada y aplicada el Data Warehouse para el Padrón Electoral, se realiza un análisis de los resultados obtenidos en base a 2 variables las cuales son:

- Tiempo de elaboración de Informes.
- Satisfacción del usuario estratégico.

4.1.Tiempo de Elaboración de Informes.

Muestra.

De una población de 7 personas comprendidas entre vocales y directores del Tribunal Electoral Departamental de Potosí, la muestra tomada en consideración para analizar la variable tiempo de elaboración de informes fue una. Para la obtención de la muestra se siguió un muestreo no probabilístico llamado muestreo causal o incidental, cuyo muestreo trata de un proceso en el que el investigador selecciona directa o intencionadamente los individuos de la población. Se realizó la entrevista al jefe del departamento de informática del Tribunal Electoral Departamental de Potosí Lic. Jhonny Ecos Huayhuata, con quien se clasifico en base a la obtención de informes por su complejidad (baja, media y alta) y por el tiempo de obtención de los mismos (baja, media y alta).

Los puntos a evaluar en este criterio fueron:

- **Análisis de requerimientos**: El tiempo que se tomó en analizar los requerimientos de cada uno de las unidades del Tribunal Electoral Departamental de Potosí.
- **Construcción:** Se consideró el tiempo de elaboración del informe, el mismo que dependerá si es un informe de baja, media o alta complejidad.
- **Pruebas:** Tiempo que se utilizó en comprobar el informe elaborado con los usuarios de las unidades del Tribunal Electoral Departamental de Potosí.

La figura 4.1 muestra los criterios de elaboración de informes/reportes. Las filas, columnas representan los tiempos de elaboración de informes del Sistema Transaccional vs. El Data Warehouse. Los mismos que tienen un tiempo de elaboración de bajo, medio y alto, los

valores de las celdas constituyen los puntajes asignados a cada uno de ellos de acuerdo a los criterios de jefe de informática y tesista.

CRITERIOS DE ELABORACIÓN DE INFORMES/REPORTES	TIEMPO DE ELABORACIÓN (min)					
	Sistema Transaccional			Data Warehouse		
	baja	media	alta	baja	media	alta
Análisis de requerimientos	2	4	4	2	4	4
Construcción	4	15	25	1	2	5
Pruebas	2	4	8	1	2	4
Total	8	23	37	4	8	13

Figura 4.1: Comparación del Tiempo de Elaboración de Informes Sistema Transaccional vs Data Warehouse

Fuente: Propia

A continuación se muestra la figura 4.2 en el cual se visualiza el tiempo de elaboración de informes entre el Sistema Transaccional y el Data Warehouse.

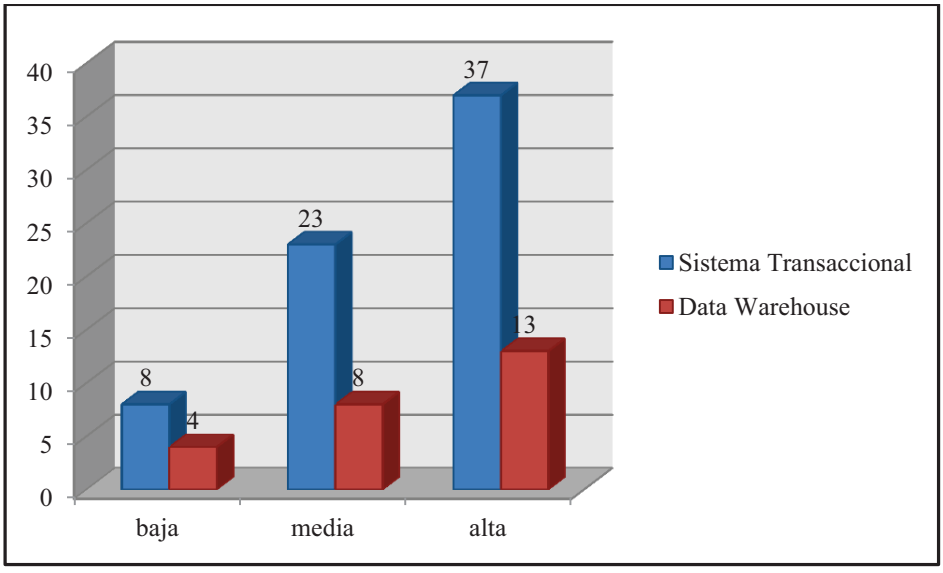


Figura 4.2: Tiempo de Elaboración de Informes Sistema Transaccional vs Data Warehouse

Fuente: Propia

Análisis.

Como se puede ver en la figura anterior los tiempos para la elaboración de los informes del Sistema Transaccional lo realiza en mayor tiempo en cambio la elaboración de los informes en el Data Warehouse son realizados en un tiempo menor. Tiempo reducido a menos de la mitad del tiempo con respecto a los informes del Sistema Transaccional. Estos informes permiten a los jefes de unidad contar con información rápida y oportuna.

4.2. Satisfacción del Usuario Estratégico.

Muestra.

De una población de 6 personas se tomó en consideración para analizar y obtener resultados 4 técnicos administrativos del departamento de informática del Tribunal Electoral Departamental de Potosí, para la obtención de la muestra se siguió un muestreo no probabilístico llamado muestreo causal o incidental.

Se evaluó los siguientes puntos:

4.2.1. En una Escala de 1 a 5 cuál es el nivel de satisfacción con los reportes y vistas OLAP.

Escala		N(i)	Frecuencia Relativa	Frecuencia Relativa Acumulada	Frecuencia Relativa%	Frecuencia Relativa Acumulada%
1	Regular	0	0,00	0,00	0,00%	0,00%
2	Malo	0	0,00	0,00	0,00%	0,00%
3	Bueno	1	0,25	0,25	25,00%	25,00%
4	Muy Bueno	2	0,50	0,75	50,00%	75,00%
5	Excelente	1	0,25	1,00	25,00%	100,00%
Total		4	1,00		100,00%	

Figura 4.3: satisfacción con los reportes y vistas OLAP

Fuente: Propia

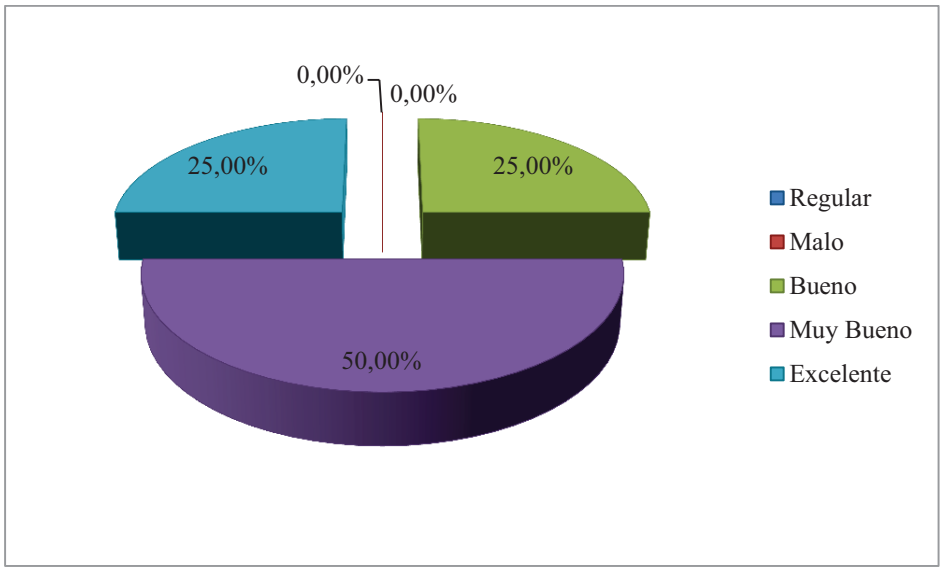


Figura 4.4: Porcentaje de satisfacción con los reportes y vistas OLAP

Fuente: Propia

Análisis.

2 personas de las 4 a las que se realizó el cuestionario informan que su nivel de satisfacción con los reportes y vistas OLAP es muy Bueno cuyo valor representa al 50% de la población y una persona informa que es excelente que corresponde al 25 % de la población.

4.2.2. ¿Cómo considera el tiempo de respuesta en la obtención de las consultas de los informes (Bajo, Medio y Alto).

Tiempo de Respuesta		N(i)	Frecuencia Relativa	Frecuencia Relativa Acumulada	Frecuencia Relativa%	Frecuencia Relativa Acumulada%
1	Bajo	0	0,00	0,00	0,00%	0,00%
2	Medio	1	0,25	0,25	25,00%	25,00%
3	Alto	3	0,75	1,00	75,00%	100,00%
Total		4	1,00		100,00%	

Figura 4.5: Tiempo de respuesta en la obtención de las consultas

Fuente: Propia

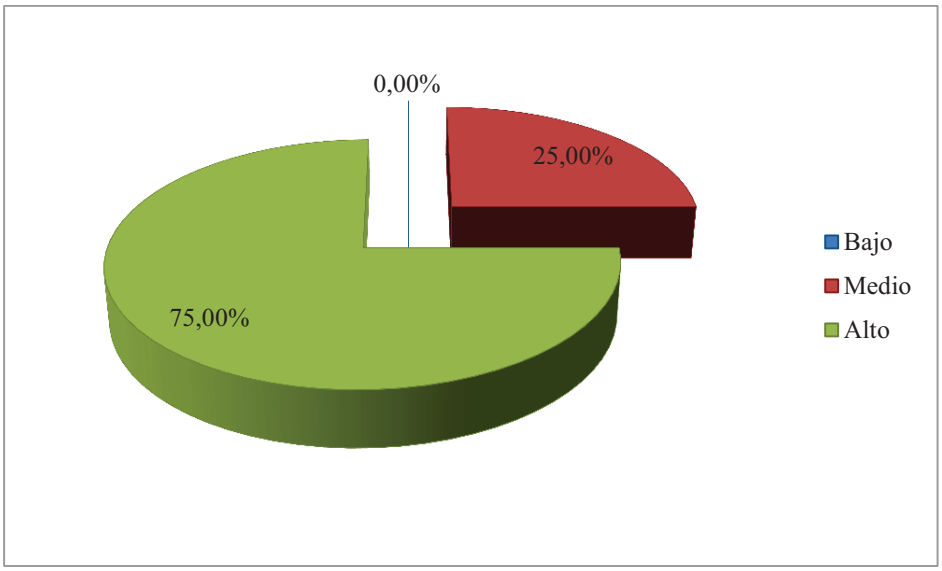


Figura 4.6: Porcentaje Tiempo de respuesta en la obtención de las consultas

Fuente: Propia

Análisis.

3 personas de las 4 que respondieron al cuestionario, respondieron que el tiempo de respuesta en la obtención de las consultas de los informes es alto, es decir la consulta se ejecuta rápidamente porcentaje que representa el 75%.

4.2.3. ¿La información que obtiene al consultar los informes son los correctos?

Datos		N(i)	Frecuencia Relativa	Frecuencia Relativa Acumulada	Frecuencia Relativa%	Frecuencia Relativa Acumulada%
1	Si	4	1,00	1,00	100,00%	100,00%
2	No	0	0,00	1,00	0,00%	100,00%
Total		4	1,00		100,00%	

Figura 4.7: Los informes son los correctos

Fuente: Propia

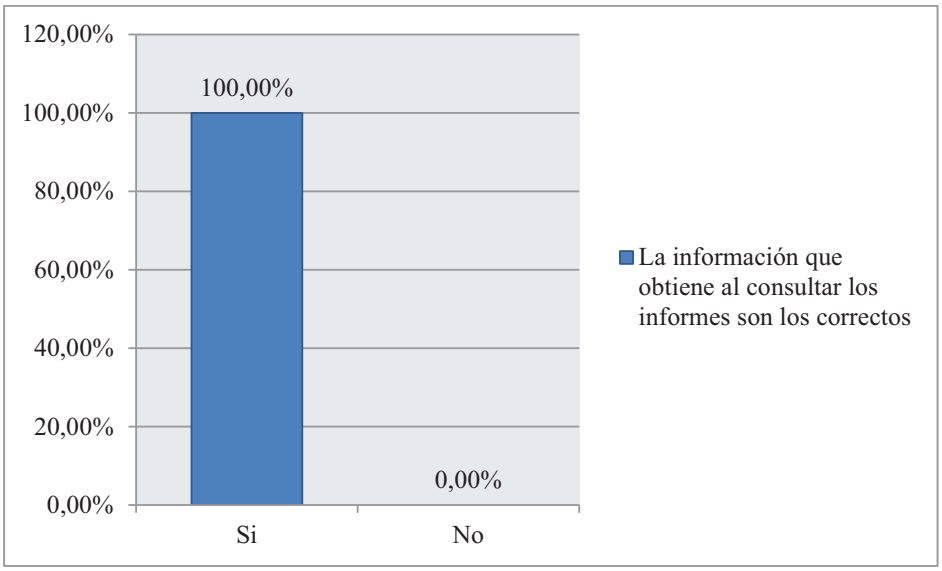


Figura 4.8: Que porcentaje de los informes son los correctos

Fuente: Propia

Análisis.

La muestra completa respondió que los datos que se obtienen en los informes es la correcta, el porcentaje corresponde al 100%.

4.2.4. ¿Considera fácil el uso del Data Warehouse?

Datos		N(i)	Frecuencia Relativa	Frecuencia Relativa Acumulada	Frecuencia Relativa%	Frecuencia Relativa Acumulada%
1	Si	4	1,00	1,00	100,00%	100,00%
2	No	0	0,00	1,00	0,00%	100,00%
Total		4	1,00		100,00%	

Figura 4.9: Considera fácil el uso del Data Warehouse

Fuente: Propia

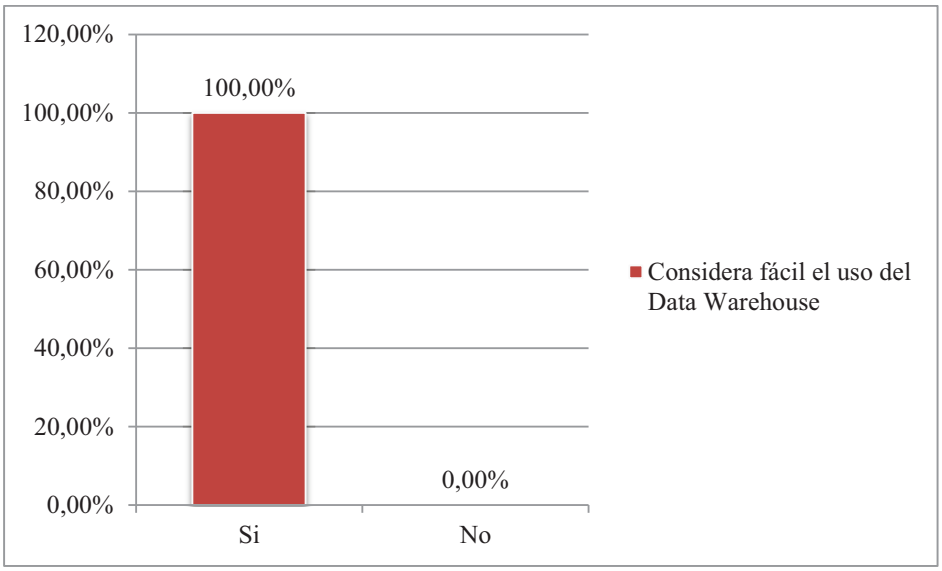


Figura 4.10: El porcentaje de la información que considera fácil el uso del Data Warehouse

Fuente: Propia

Análisis.

La muestra completa respondió que considera fácil el uso del Data Warehouse, el porcentaje corresponde al 100%.

4.2.5. ¿Para obtener informes usted prefiere realizar consultas o utilizar el Data Warehouse?

	Datos	N(i)	Frecuencia Relativa	Frecuencia Relativa Acumulada	Frecuencia Relativa%	Frecuencia Relativa Acumulada%
1	Sistema Transaccional	0	0,00	0,00	0,00%	0,00%
2	Data Warehouse	4	1,00	1,00	100,00%	100,00%
	Total	4	1,00		100,00%	

Figura 4.11: Para obtener informes usted prefiere realizar consultas o utilizar el Data Warehouse

Fuente: Propia

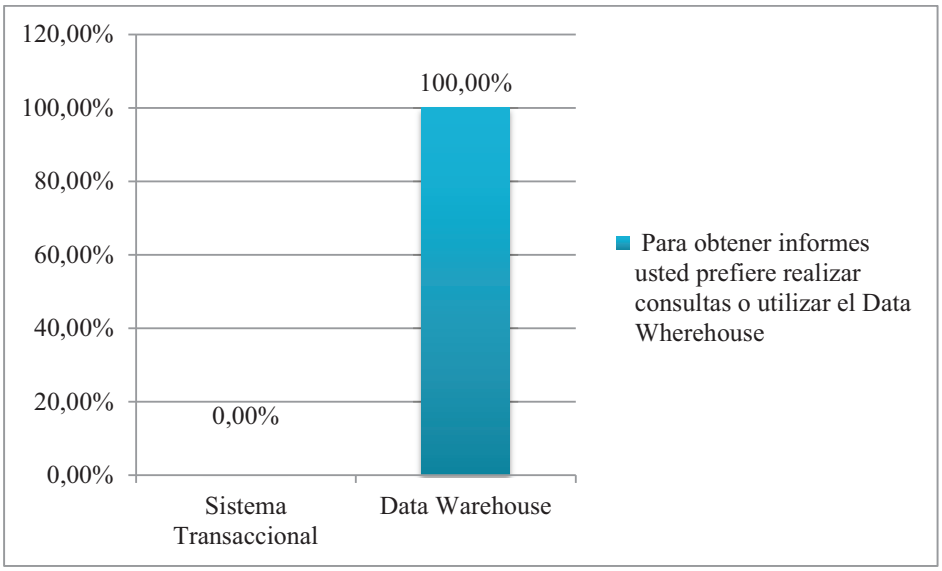


Figura 4.12: Porcentaje que prefiere utilizar el Data Warehouse

Fuente: Propia

Análisis.

El 100% respondió que prefiere utilizar el Data Warehouse al sistema transaccional.

CONCLUSIONES Y RECOMENDACIONES

En este apartado se establecen las conclusiones y recomendaciones finales del presente trabajo y los posteriores trabajos que puedan realizarse.

CONCLUSIONES

- Se implementó el Data Warehouse, con la generación de información útil y analítica dentro de un nivel organizacional táctico – operativo, producto de los requerimientos del negocio.
- El modelo Dimensional del Data Warehouse permite una mejor funcionalidad en las consultas, almacenando datos relevantes para su análisis.
- Es importante implementar técnicas de extracción de datos, esto ayuda al desempeño del proceso ETL y al ahorro de tiempo y recursos
- El Data Warehouse permite desarrollar aplicaciones creando un resumen Bases de Datos operativas con la información importante, logrando la obtención de resultados de una forma eficaz y eficiente.
- El proceso de construcción de un Data Warehouse es un proceso muy extenso y toma tiempo considerable para que la información este bien estructurado y sea valiosa para el negocio.
- Se utilizó la metodología propuesta por Ralph Kimball como guía para el proyecto, la misma que ayudó para planificar por etapas el desarrollo, permitiendo la administración y mantenimiento durante todo el ciclo de vida.
- Se utilizó las herramientas de Base de Datos SQL Server, el motor de Base de Datos para el almacenamiento de los datos, SSIS para la extracción, transformación y carga de los datos, SSAS para la creación y generación del cubo, y Excel y Reporting Service para la generación de los reportes.

RECOMENDACIONES.

- En este trabajo solo se trabajó con las mediciones de nuevos inscritos, habilitados y depurados con diferentes dimensiones. Sería conveniente ampliar muchas más mediciones, pudiendo ser cantidad de mesas, localidades y municipios.
- Investigar sobre las técnicas de minería de datos ya que esto ayudaría a la institución a tener un proyecto más completo.
- Se recomienda la implementación y presentación de indicadores con herramientas que visualicen gráficos que ayuden a interpretar el estado de los mismos para un mejor análisis. Pudiendo ser estas por medio de medidores (dashboard).

BIBLIOGRAFIA

- BERNABEU, R. Data Warehousing Investigación Y Sistematización De Conceptos – Hefesto Metodología Propia Para La Construcción De Un Data Warehouse. Córdoba, Argentina: 2009
- BUIGUES, A. Arquitectura de un Data Warehouse [En línea]. [fecha de consulta: julio de 2011], disponible: http://anabuigues.com/2010/03/05/arquitectura-de-un-data-warehouse/
- Business Intelligence Facil. Serie sobre cómo construir un Data Warehouse [En línea] , [fecha de consulta: marzo de 2011], disponible: http://www.businessintelligence.info/serie-dwh/como-construir-datawarehouse.html
- CANALES INOCENCIO, B. Inteligencia de negocios, en qué ayuda a tu negocio y cuál es su proceso [En línea]. [fecha de consulta: junio de 2011], disponible: http://www.gestiopolis.com/canales2/gerencia/1/busint.htm
- FABIAN, A. Implementación De Un Datawarehouse Para El Instituto Geográfico Militar. Trabajo de Titulación (Ingeniero en Sistemas e Informática). Escuela politécnica del ejército. Sangolqui:2010.
- GONZALES, E. Técnicas de Mineria de Datos [En línea]. [fecha de consulta: abril de 2011], disponible: http://www.monografias.com/trabajos55/mineria-de-datos/mineria-de-datos.shtml.
- INMON, W. Building the Data warehouse, Third Edition. Wiley Publishing, 2002.
- KIMBALL, R y CASERTA, J. The Data Warehouse ETL Tokit .Indianapolis: Wiley Publishing, 2004.
- KIMBALL, R.; REEVES, L.; ROSS, M. y THORNTHAILE The Data Warehouse Lifecycle Toolkit, Second Edition. Wiley.
- LANGIT, L. Foundations of Sql Server 2005 Business Intelligense. Verlag New York: Apress 2007.
- LOCKSLEY, L. 10 pasos para una negocioación efectiva[En línea], [fecha de consulta: marzo de 2011], disponible: http://jpgarcia69.wordpress.com/2008/07/15/10-pasos-para-una-negociacion-efectiva/
- MACLENNAN, J; TANG, Z. y CRIVAT, B. Data Mining with Microsoft Sql Server 2008. Indianapolis: Wiley Publishing, 2009.
- MSDN, SQL Server Integration Services [En línea], [fecha de consulta: agosto de 2011], disponible:http://msdn.microsoft.com/es-es/library/ms141026%28v=SQL.100%29.aspx
- MUNDY, J.; THORNTHWAITE, W. y KIMBALL, R. The Microsoft Data Warehouse Toolkit, Second Edition. Indianapolis: Wiley Publishing, 2011.

- PONNIAH, P. Data Modeling Fundamentals A Practical Guide For It Professionals. New Jersey: Wiley 2007
- PONNIAH, P. Data Warehousing Fundamentals A Comprehensive Guide For It Professionals. New York: Wiley 2001
- RAINARDI VINCENT. Building a Data Warehouse with Examples in Sql Server. United States: Apress 2008.
- RAMOS, S. Microsoft Business Intelligence: Vea el Cubo Medio Lleno, SolidQ. España: SolidQ Press, 2011
- UTLEY, G. Business Intelligense with Microsoft Office PerformancePoint Server 2007. New York: Mc Graw Hill, 2008.
- WANG, J. Encyclopedia of Data Warehousing and Mining. United States: Idea Group Reference.
- WIKIPEDIA, Almacén de Datos [En línea]. [fecha de consulta: junio de 2011], disponible: http://es.wikipedia.org/wiki/Almac%C3%A9n_de_datos
- WITHEE, K. Microsoft Business Intelligense for Dummies. Indianapolis: Wiley Publishing, 2010.

REFERENCIAS BIBLIOGRÁFICAS

1. RAMOS, S. Microsoft Business Intelligence: Vea el Cubo Medio Lleno, SolidQ. España: SolidQ Press, 2011, Pag. 9
2. CANALES INOCENCIO, B. Inteligencia de negocios, en qué ayuda a tu negocio y cuál es su proceso [En línea]. [fecha de consulta: junio de 2011], disponible: http://www.gestiopolis.com/canales2/gerencia/1/busint.htm.
3. WIKIPEDIA, Almacén de Datos [En línea]. [fecha de consulta: junio de 2011], disponible: http://es.wikipedia.org/wiki/Almac%C3%A9n_de_datos.
4. RAMOS, S. Microsoft Business Intelligence: Vea el Cubo Medio Lleno, SolidQ. España: SolidQ Press, 2011. Pag. 11
5. BUIGUES, A. Arquitectura de un Data Warehouse [En línea]. [fecha de consulta: julio de 2011], disponible: http://anabuigues.com/2010/03/05/arquitectura-de-un-data-warehouse/
6. KIMBALL, R.; REEVES, L.; ROSS, M. y THORNTHAILE The Data Warehouse Lifecycle Toolkit, Second Edition. Wiley. Pag 2.3
7. KIMBALL, R.; REEVES, L.; ROSS, M. y THORNTHAILE The Data Warehouse Lifecycle Toolkit, Second Edition. Wiley. Pag. 90.
8. MSDN, SQL Server Integration Services, [fecha de consulta: agosto de 2011], disponible:http://msdn.microsoft.com/es-es/library/ms141026%28v=SQL.100%29.aspx

ANEXO 1

GUÍA DE ENTREVISTA A VOCALES Y DIRECTORES

Fecha:……..

Entrevistado:…….

Objetivo: Identificar la necesidad de contar que tipo de reportes contribuirá con la toma de decisiones.

Preguntas:

1. A lo largo del trabajo en la institución ¿Considera que necesita reportes al padrón electoral que usted mismo podría obtenerlos, sin necesidad de pedir a otras áreas la elaboración de los mismos?

SI ☐ NO ☐

Por qué……..

……..

2. ¿Cuenta usted con todo el material necesario para llevar a cabo los procesos de empadronamiento en el Tribunal Electoral Departamental de Potosí?

SI ☐ NO ☐

Por qué……..

……..

3. ¿Qué problemas encuentra en los reportes que se le proporcionan?

……
……

4. ¿Qué reportes le interesa para llevar a cabo un proceso de empadronamiento?

……
……

5. ¿Está de acuerdo con la elaboración de un Data Warehouse que permita elaborar los diferentes reportes que usted necesita?

SI ☐ NO ☐

ANEXO 2

CREACIÓN DE BASE DE DATOS DATA WAREHOUSE PADRONELECTORALDW

Creación de Base de Datos

```
USE [master]

GO

CREATE DATABASE [PadronElectoralDW] ON  PRIMARY

( NAME = N'PadronElectoralDW', FILENAME = N'C:\Program Files\Microsoft SQL
Server\MSSQL10.MSSQLSERVER\MSSQL\DATA\PadronElectoralDW.mdf' , SIZE = 160768KB ,
MAXSIZE = UNLIMITED, FILEGROWTH = 1024KB )

 LOG ON

( NAME = N'PadronElectoralDW_log', FILENAME = N'C:\Program Files\Microsoft SQL
Server\MSSQL10.MSSQLSERVER\MSSQL\DATA\PadronElectoralDW_log.ldf' , SIZE = 199296KB ,
MAXSIZE = 2048GB , FILEGROWTH = 10%)
```

Creación de tablas

```
USE [PadronElectoralDW]
GO
```

DimEstadoCivil

```
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
SET ANSI_PADDING ON
GO
CREATE TABLE [dbo].[DimEstadoCivil](
	[IdEstadoCivil] [varchar](1) NOT NULL,
	[DescEstadoCivil] [varchar](50) NULL,
 CONSTRAINT [PK_DimEstadoCivil] PRIMARY KEY CLUSTERED
(
	[IdEstadoCivil] ASC
)WITH (PAD_INDEX  = OFF, STATISTICS_NORECOUSE [PadronElectoralDW]
GO
```

DimGeografia

```
CREATE TABLE [dbo].[DimGeografia](
	[NumMesa] [bigint] NOT NULL,
	[NumeroMesa] [nvarchar](2) NULL,
	[IdPais] [int] NULL,
	[NomPais] [nvarchar](50) NULL,
	[Dep] [tinyint] NULL,
	[NomDep] [nvarchar](15) NULL,
	[Prov] [tinyint] NULL,
	[NomProv] [nvarchar](150) NULL,
	[Sec] [tinyint] NULL,
	[NomSec] [nvarchar](150) NULL,
	[IdLoc] [smallint] NULL,
	[NomLoc] [nvarchar](150) NULL,
	[Dist] [tinyint] NULL,
	[NomDist] [nvarchar](150) NULL,
	[Zona] [smallint] NULL,
	[NomZona] [nvarchar](150) NULL,
	[Reci] [smallint] NULL,
	[NomReci] [nvarchar](150) NULL,
 CONSTRAINT [PK_DimGeografia] PRIMARY KEY CLUSTERED
(
	[NumMesa] ASC
)WITH (PAD_INDEX  = OFF, STATISTICS_NORECOMPUTE  = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS  = ON, ALLOW_PAGE_LOCKS  = ON) ON [PRIMARY]
) ON [PRIMARY]

GO
MPUTE  = OFF, IGNORE_DUP_KEY = OFF, ALLOW_ROW_LOCKS  = ON, ALLOW_PAGE_LOCKS  =
ON) ON [PRIMARY]
) ON [PRIMARY]
GO
```

DimGradoInstruccion

```
CREATE TABLE [dbo].[DimGradoInstruccion](
	[IdGradoI] [varchar](1) NOT NULL,
	[DescGradoI] [varchar](15) NULL,
 CONSTRAINT [PK_DimGradoInstruccion] PRIMARY KEY CLUSTERED
(
	[IdGradoI] ASC
)WITH (PAD_INDEX  = OFF, STATISTICS_NORECOMPUTE  = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS  = ON, ALLOW_PAGE_LOCKS  = ON) ON [PRIMARY]
) ON [PRIMARY]
GO
```

DimSexo

```
CREATE TABLE [dbo].[DimSexo](
	[Sexo] [varchar](1) NOT NULL,
	[Nombre] [nvarchar](10) NULL,
 CONSTRAINT [PK_DimSexo] PRIMARY KEY CLUSTERED
(
```

```
	[Sexo] ASC
)WITH (PAD_INDEX  = OFF, STATISTICS_NORECOMPUTE  = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS  = ON, ALLOW_PAGE_LOCKS  = ON) ON [PRIMARY]
) ON [PRIMARY]

GO
```

DimTiempo

```
 CREATE TABLE [dbo].[DimTiempo](
	[TimeID] [int] NOT NULL,
	[Fecha] [date] NULL,
	[Anio] [varchar](4) NULL,
	[MesID] [int] NULL,
	[Mes] [varchar](10) NULL,
	[Cuarto] [int] NULL,
	[Semana] [int] NULL,
	[NumDiaSemana] [int] NULL,
	[DiaSemana] [varchar](10) NULL,
 CONSTRAINT [PK_DimTiempo] PRIMARY KEY CLUSTERED
(
	[TimeID] ASC
)WITH (PAD_INDEX  = OFF, STATISTICS_NORECOMPUTE  = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS  = ON, ALLOW_PAGE_LOCKS  = ON) ON [PRIMARY]
) ON [PRIMARY]

GO
```

DimTipoCedula

```
CREATE TABLE [dbo].[DimTipoCedula](
	[TipoDoc] [varchar](1) NOT NULL,
	[NombreCed] [varchar](30) NULL,
 CONSTRAINT [PK_DimTipoCedula] PRIMARY KEY CLUSTERED
(
	[TipoDoc] ASC
)WITH (PAD_INDEX  = OFF, STATISTICS_NORECOMPUTE  = OFF, IGNORE_DUP_KEY = OFF,
ALLOW_ROW_LOCKS  = ON, ALLOW_PAGE_LOCKS  = ON) ON [PRIMARY]
) ON [PRIMARY]

GO
```

FactCiudadano

```
CREATE TABLE [dbo].[FactCiudadano](
	[mesaciudadano] [bigint] NULL,
	[TimeId] [int] NULL,
	[SEXO] [varchar](1) NULL,
	[ESTADOCIVIL] [varchar](1) NULL,
	[GRADOINSTRUCCION] [varchar](1) NULL,
	[TIPOCEDULAACT] [varchar](1) NULL,
	[Nueva Inscripcion] [int] NULL,
	[Habilitados] [int] NULL,
```

```
        [Depurados] [int] NULL
) ON [PRIMARY]
GO
```

Llenado de la tabla Tiempo

```
SET LANGUAGE spanish
/*Ponemos que el primer dia de la semana
1 es Lunes. Para el Domingo usar 7*/
SET DATEFIRST 1
declare @fi datetime, @ff datetime
set @fi = '2009-01-08'
Set @ff = '2009-16-10'
while @fi  <= @ff
begin
insert into DimTiempo (TimeID, Fecha, Anio, MesID, Mes,
Cuarto, Semana, NumDiaSemana, DiaSemana)
select year(@fi)*10000+month(@fi)*100+day(@fi) as TimeId,
@fi as Fecha,
year(@fi) as Anio,
year(@fi)*100+month(@fi) as MesID,

datename(MONTH ,@fi) as Mes,
year(@fi)*10+datepart(q,@fi) as Cuarto,
datename(year,@fi)+datename(ww,@fi) as Semana,
datepart( dw, @fi) as NumDiaSemana,
datename( dw, @fi) as DiaSemana
set @fi = @fi + 1
end
```

ANEXO 3

EXTRACCIÓN DE DATOS ORIGEN

Geografía

```
SELECT       m.NumMesa, SUBSTRING(CONVERT(nvarchar, m.NumMesa), CONVERT(int,
LEN(CONVERT(nvarchar, m.NumMesa))) - 1, 2) AS NumeroMesa, pa.IdPais, pa.NomPais,
                  de.Dep, de.NomDep, p.Prov, ISNULL(p.NomProv, N'') AS NomProv, s.Sec,
ISNULL(s.NomSec, N'') AS NomSec, l.IdLoc, ISNULL(l.NomLoc, N'') AS NomLoc, d.Dist,
                  ISNULL(d.NomDist, N'') AS NomDist, z.Zona, ISNULL(z.NomZona, N'') AS NomZona,
r.Reci, ISNULL(r.NomReci, N'') AS NomReci
FROM          Dep AS de INNER JOIN
                  Pais AS pa ON de.IdPais = pa.IdPais INNER JOIN
                  Prov AS p ON de.Dep = p.DepProv INNER JOIN
                  Municipio AS s ON p.Prov = s.ProvSec AND p.DepProv = s.DepSec INNER JOIN
                  Loc AS l ON s.DepSec = l.DepLoc AND s.ProvSec = l.ProvLoc AND s.Sec = l.SecLoc
INNER JOIN
                  Dist AS d ON l.IdLoc = d.IdLocDist INNER JOIN
                  Zona AS z ON l.IdLoc = z.IdLocZona AND d.Dist = z.DistZona INNER JOIN
                  Reci AS r ON l.IdLoc = r.IdLocReci AND z.Zona = r.ZonaReci INNER JOIN
                  Mesa AS m ON pa.IdPais = m.PaisMesa AND de.Dep = m.DepMesa AND r.Reci =
m.ReciMesa AND l.IdLoc = m.IdLocMesa AND d.Dist = m.DistMesa AND
                  z.Zona = m.ZonaMesa
```

Grado de Instrucción

```
Select * from GradoInstruccion.
```

Sexo

```
Select * from Sexo
```

Estado Civil

```
Select * from EstadoCivil.
```

.

Tipo Cedula

```
Select * from TipoCedula.
```

Ciudadano

```
SELECT     dbo.DW_ciudadano0.*,  dbo.DW_ciudadano1.Habilitados, dbo.DW_ciudadano2.Depurados
FROM       dbo.DW_ciudadano2 INNER JOIN
                  dbo.DW_ciudadano1 ON dbo.DW_ciudadano2.mesaciudadano =
dbo.DW_ciudadano1.mesaciudadano AND
                  dbo.DW_ciudadano2.TimeId = dbo.DW_ciudadano1.TimeId AND dbo.DW_ciudadano2.SEXO
= dbo.DW_ciudadano1.SEXO AND
                  dbo.DW_ciudadano2.ESTADOCIVIL = dbo.DW_ciudadano1.ESTADOCIVIL AND
                  dbo.DW_ciudadano2.GRADOINSTRUCCION = dbo.DW_ciudadano1.GRADOINSTRUCCION
AND
                  dbo.DW_ciudadano2.TIPOCEDULAACT = dbo.DW_ciudadano1.TIPOCEDULAACT FULL
OUTER JOIN
                  dbo.DW_ciudadano0 ON dbo.DW_ciudadano1.mesaciudadano =
dbo.DW_ciudadano0.mesaciudadano AND
                  dbo.DW_ciudadano1.TimeId = dbo.DW_ciudadano0.TimeId AND dbo.DW_ciudadano1.SEXO
= dbo.DW_ciudadano0.SEXO AND
                  dbo.DW_ciudadano1.ESTADOCIVIL = dbo.DW_ciudadano0.ESTADOCIVIL AND
                  dbo.DW_ciudadano1.GRADOINSTRUCCION = dbo.DW_ciudadano0.GRADOINSTRUCCION
AND
                  dbo.DW_ciudadano1.TIPOCEDULAACT = dbo.DW_ciudadano0.TIPOCEDULAACT
```

Donde:

DW_ciudadano0

```
SELECT     TOP (100) PERCENT mesaciudadano, EstadoRegistro, YEAR(FECHAINS) * 10000 +
MONTH(FECHAINS) * 100 + DAY(FECHAINS) AS TimeId, SEXO, ESTADOCIVIL,
                  GRADOINSTRUCCION, TIPOCEDULAACT, COUNT(*) AS [Nueva Inscripcion]
FROM       dbo.ciudadano
WHERE     (EstadoRegistro = 0)
GROUP BY mesaciudadano, EstadoRegistro, YEAR(FECHAINS) * 10000 + MONTH(FECHAINS) * 100 +
DAY(FECHAINS), SEXO, ESTADOCIVIL, GRADOINSTRUCCION,
                  TIPOCEDULAACT
```

DW_ciudadano1

```
SELECT     TOP (100) PERCENT mesaciudadano, EstadoRegistro, YEAR(FECHAINS) * 10000 +
MONTH(FECHAINS) * 100 + DAY(FECHAINS) AS TimeId, SEXO, ESTADOCIVIL,
                  GRADOINSTRUCCION, TIPOCEDULAACT, COUNT(*) AS [Habilitados]
FROM       dbo.ciudadano
WHERE     (EstadoRegistro = 1)
GROUP BY mesaciudadano, EstadoRegistro, YEAR(FECHAINS) * 10000 + MONTH(FECHAINS) * 100 +
DAY(FECHAINS), SEXO, ESTADOCIVIL, GRADOINSTRUCCION,
                  TIPOCEDULAACT
```

DW_ciudadano2

```
SELECT     TOP (100) PERCENT mesaciudadano, EstadoRegistro, YEAR(FECHAINS) * 10000 +
MONTH(FECHAINS) * 100 + DAY(FECHAINS) AS TimeId, SEXO, ESTADOCIVIL,
                      GRADOINSTRUCCION, TIPOCEDULAACT, COUNT(*) AS [Depurados]
FROM         dbo.ciudadano
WHERE     (EstadoRegistro = 2)
GROUP BY mesaciudadano, EstadoRegistro, YEAR(FECHAINS) * 10000 + MONTH(FECHAINS) * 100 +
DAY(FECHAINS), SEXO, ESTADOCIVIL, GRADOINSTRUCCION,
                      TIPOCEDULAACT
```

ANEXO 4

CUESTIONARIO PARA DEMOSTRAR EL GRADO DE SATISFACCIÓN DE LOS USUARIOS PARA EL MANEJO DEL DATA WAREHOUSE

Fecha:……..

Entrevistado:…….

Objetivo: demostrar el grado de satisfacción de los usuarios por el manejo del Data Warehouse.

CUESTIONARIO:

1. ¿En una escala de 1 a 5 cuál es su nivel de satisfacción con los reportes y vistas OLAP?

   1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

2. ¿Cómo considera el tiempo de respuesta en la obtención de las consultas de los informes?

   Bajo ☐ Medio ☐ Alto ☐

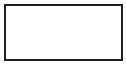
3. ¿La información que obtiene al consultar los informes son los correctos?

   SI ☐ NO ☐

4. Considera fácil el uso del Data Warehouse?

   SI ☐ NO ☐

5. Para obtener informes usted prefiere realizar consultas o utilizar el Data Warehouse

Sistema Transaccional  Data Warehouse 