



UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS
FACULTAD 1

*SOLUCIÓN DE INTELIGENCIA INSTITUCIONAL PARA EL SISTEMA
ÚNICO DE IDENTIFICACIÓN NACIONAL*

TRABAJO DE DIPLOMA PARA OPTAR POR EL TÍTULO DE INGENIERO EN
CIENCIAS INFORMÁTICAS

*Autores: Ariel Antonio Fundora López
Roberto Carlos Cardero Pérez*

Tutor: Ing. Elizabetha Rojas Santiesteban.

La Habana, 14 de junio del 2013



Declaración de autoría

Declaramos ser autores del presente trabajo de diploma para optar por el título de Ingeniero en Ciencias Informáticas, que fue desarrollado en el transcurso de los años 2012-2013. Declaramos que todo lo expuesto se ajusta a la verdad, y asumimos la responsabilidad moral y jurídica que se derive de este juramento profesional. Autorizamos a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo.

Para que así conste firmamos la declaración jurada de autoría en La Habana a los ____ días del mes de _____ del año 2013.

Autor: Ariel Antonio Fundora López

Autor: Roberto Carlos Cardero Pérez



Dedico esta investigación:

A mi familia que ha sido la inspiración para mis logros como persona y estudiante. En especial a mi bisabuela Tita que aunque hoy no esté con nosotros siempre estará conmigo, y a mi madre por su infinito amor incondicional, por ser mi fuerza en todo momento y mi eterno apoyo.

Ariel

Dedico este trabajo:

A mi hija, por responder con alegría al amor y dedicación de sus padres; a mi amiga especial, mi esposa, por ser mi ayuda idónea, por su amor, su servicio, y el privilegio de ser el compañero de su vida; a mi madre, por su abnegada dedicación, la confianza depositada, y la firme convicción de que llegaría a ser alguien útil para la sociedad; a mi abuela Milagro, por contribuir desde mi niñez a la formación de un carácter aprobado por el Cielo, y por la fe en el bien común; a mi abuelo Roberto, que como combatiente puso en riesgo su vida por un país donde los jóvenes tuvieran un acceso digno a la educación, una realidad que es comprobable más de medio siglo después en la formación académica de su nieto.

Roberto



Quiero agradecer primeramente a mis padres, quienes han sido mi impulso y orgullo profesional, esta tesis es para que se sientan orgullosos.

A todas las personas que a lo largo de estos años en la UCI han estado conmigo y me han brindado su apoyo: a mi novia Tandita por estar a mi lado, darme su amor y hacerme reír durante estos años; a mi tutora Elizabeta por haberme guiado con sabiduría durante todo el trayecto del trabajo de diploma, a mi compañero de tesis Robertico por todo su esfuerzo y energía, y a Erick por toda su ayuda brindada; a todos los profesores que contribuyeron a hacer de mí un hombre de futuro, sobre todo a Yassef y a Joel pues gracias a ellos soy un mejor profesional.

A mis amigos, mis compañeros de aula, los socios del ocio y a todos los que de una forma u otra hicieron mi estancia más placentera en la universidad; especialmente a mi amigo Sael por compartir tantos desvelos, competencias y jodederas, y a las chicas del 121 por tantos recuerdos imborrables.

A toda mi familia: a mi padre por darme su ejemplo, sus consejos y apoyarme en las dificultades; a mi hermanos pequeños Marcos y Luis Daniel por traer siempre alegría a mi vida, a mi abuela Edith por enseñarme a ser soñador y espiritual, y a mi abuelo Mateo por enseñarme a ser realista; a Adanay por cuidar de mí durante todo este tiempo, a Luis por enseñarme perseverancia, a mi hermano Luis Javier por ser mi socio, mi amigo y compañero de descarga. A mi mis tíos Ivet y Marchena, mi abuela Vicenta, mi tia Marlén y mi prima Mayla por darme su amor. En fin, a todas las maravillosas personas que he conocido a lo largo de estos cinco años de carrera, gracias a ustedes seré cada día mejor.

Ariel



Dedico la presente como agradecimiento al apoyo brindado durante estos años de estudio y como reconocimiento de gratitud por haber finalizado la carrera.

A mi esposa Yanieska, y a mi hija Jasmine, como una demostración de gratitud por la inspiración que necesitaba para concluir mi carrera profesional.

A mis padres Alicia Pérez Fuentes y Juan Cardero Hernández; a mi padrastro Jorge Luis Purniel y a mi madrastra Lusmila Saiz que se han portado tan bien conmigo que le tengo el afecto de un hijo. Por el ánimo y el apoyo que recibí de ustedes, también es suyo el objetivo logrado.

A mi abuela Milagro por enseñarme a amar a Dios, por su apoyo incondicional, por sus regaños, por sus sí y sus no. Porque ha sido de Dios de quién emanó la fuerza que recibiera en medio de la debilidad mientras he estado en la universidad.

A mi tío Manuel y a mi abuelo Roberto porque han hecho todo lo que ha estado a su alcance para que lograra terminar los estudios en la universidad.

A mis hermanas Gabriela, Gretel y Loraine, a mi hermanito Alejandro que está por nacer; y a mi tía Damaris, a mis primos Yanet, Jorgito, y Lisy por sentirse orgullosos de mí.

Agradezco y reconozco también por este trabajo a la labor de mi compañero de tesis Ariel Antonio Fundora López, y a la Ing. Elizabeta Rojas Santiesteban por su asesoría y dirección en el proceso de investigación.

Reconocimiento para todos los profesores y especialistas que contribuyeron a nuestra formación profesional durante toda la carrera. A la Universidad de las Ciencias Informáticas, por la gestión universitaria y el soporte institucional para la realización de este trabajo.



Agradecimientos

A mis amigos y compañeros de apartamento por los años de convivencia y porque también pudieron llegar hasta aquí: Yoevis, Herbert, Nilbert, José, Daniel, Isidro, y Tayler. A los integrantes de los grupos 1505 y 1502, con quienes tuve la oportunidad de compartir horas de docencia y de estudio. A mis amigos y amigas de siempre; a Adrián Rivera, a quien considero mi hermano y que me ayudó en momentos difíciles de la carrera y quien también aportó a la mejora de este trabajo de tesis.

En resumen, a todas aquellas personas que de una u otra forma, colaboraron en la realización de esta investigación, hago extensivo mi más sincero agradecimiento.

Roberto



Resumen

Cuba se encuentra envuelta en un proceso de informatización que pretende elevar la calidad del servicio prestado a la población. En este contexto, el Ministerio del Interior (MININT) juega un papel esencial y en colaboración con el Centro de Identificación y Seguridad Digital (CISED) adjunto a la Universidad de las Ciencias Informáticas (UCI) desarrollan el Sistema Único de Identificación Nacional (SUIN). Resultado de las actividades diarias de este sistema se genera un gran cúmulo de datos, donde la toma adecuada de decisiones y el análisis de la información es primordial.

Actualmente el MININT no cuenta con una herramienta informática disponible para apoyar tales análisis. Por tanto, el presente trabajo de diploma tiene como objetivo general desarrollar una solución de Inteligencia Institucional que permita mejorar el proceso de toma de decisiones en el Sistema Único de Identificación Nacional. El desarrollo de la solución se realiza mediante la implementación de un Almacén de Datos que está guiado por la metodología HEFESTO, y de una aplicación web para que los funcionarios puedan visualizar la información clave, mediante reportes estadísticos que los apoyen en el proceso de toma de decisiones. La solución de Inteligencia Institucional tiene especial importancia para el país por el aporte a la prestación de servicios que supone su inclusión en las soluciones informáticas que complementan al SUIN y particularmente por la digitalización del Cuadro de Mando Integral del MININT.

Palabras claves:

Almacén de Datos, Inteligencia Institucional, SUIN, toma de decisiones.



Índice de contenido

Introducción	1
Estructura del documento	4
Capítulo 1 Fundamentos teóricos	5
1.1 Introducción.....	5
1.2 Inteligencia de Negocio	5
1.3 Inteligencia Institucional.....	7
1.3.1 ¿Por qué usar una Solución de Inteligencia Institucional en una organización?	8
1.3.2 Ventajas de una solución de Inteligencia Institucional.....	8
1.4 Data Warehouse	9
1.4.1 Características de un Data Warehouse.....	10
1.4.2 Razones que justifican la creación de un Data Warehouse	11
1.4.3 Ventajas y Desventajas	12
1.4.4 Data Marts	13
1.5 Arquitectura del Data Warehouse	14
1.5.1 OLTP	15
1.5.2 Integración de Datos – ETL	16
1.5.3 Almacén de datos físico y su estructura multidimensional	17
1.5.4 Herramientas de consulta y análisis	23
1.6 Ambiente de desarrollo	24
1.6.1 Justificación de la metodología	24
1.6.2 Herramienta de diseño y modelado de Almacenes de Datos: Embarcadero ER/Studio 8.0	25
1.6.3 Herramienta de Almacenamiento de Datos: Oracle Database.....	25
1.6.4 Herramienta de reportes: Oracle Business Intelligence Enterprise Edition.....	26
1.6.5 Herramienta para la realización de los ETL: Oracle Warehouse Builder	27
1.6.6 Herramienta para la gestión de Base de Datos: PL/SQL Developer.....	27
1.7 Conclusiones parciales.....	27
Capítulo 2 Análisis y diseño del Almacén de Datos.....	29
2.1 Introducción.....	29
2.2 Análisis de los requerimientos.....	29



2.2.1	Identificar preguntas.....	29
2.2.2	Identificar indicadores y perspectivas	31
2.2.3	Modelo Conceptual	32
2.3	Análisis de los OLTP	33
2.3.1	Conformar indicadores	33
2.3.2	Establecer correspondencias	34
2.3.3	Nivel de granularidad.....	37
2.3.4	Modelo Conceptual ampliado	40
2.4	Modelo Lógico de Almacén de Datos	41
2.4.1	Tipo de Modelo Lógico del Almacén de Datos	41
2.4.2	Tablas de dimensiones.....	41
2.4.3	Tablas de hechos.....	42
2.4.4	Uniones.....	43
2.5	Modelo de Arquitectura del Almacén de Datos	44
2.6	Conclusiones parciales.....	44
Capítulo 3 Implementación y pruebas		45
3.1	Introducción.....	45
3.2	Implementación del almacén	45
3.3	Implementación de la aplicación web	53
3.4	Despliegue de la solución	58
3.5	Pruebas.....	59
3.5.1	Lista de chequeo	59
3.5.2	Análisis de los tiempos de respuesta de los reportes.....	62
3.5.3	Prueba de integridad de datos	63
3.5.4	Prueba de Aceptación.....	64
3.6	Conclusiones parciales.....	66
Conclusiones		67
Recomendaciones		68
Referencias bibliográficas		69
Anexos		72



Índice de figuras

Figura 1.1 Pirámide del conocimiento en Business Intelligence.....	7
Figura 1.2 Inteligencia Institucional.....	8
Figura 1.3 Arquitectura de un almacén de datos.....	15
Figura 1.4 Proceso ETL.....	17
Figura 1.5 Estructura de un cubo multidimensional.....	19
Figura 1.6 Esquema Estrella.....	20
Figura 1.7 Esquema Copo de Nieve.....	20
Figura 2.1 Modelo Conceptual.....	32
Figura 2.2 Correspondencia entre la perspectiva Cambio de Dirección con las tablas del Esquema Identidad.....	34
Figura 2.3 Correspondencia entre la perspectiva Trámite con las tablas del Esquema Identidad.....	35
Figura 2.4 Correspondencia entre la perspectiva Fecha con las tablas del Esquema Identidad.....	35
Figura 2.5 Correspondencia entre la perspectiva Persona con las tablas del Esquema Identidad.....	36
Figura 2.6 Correspondencia entre la perspectiva Oficina con las tablas del Esquema ADM.....	36
Figura 2.7 Modelo Conceptual Ampliado.....	40
Figura 2.8 Dimensiones.....	41
Figura 2.9 Hechos.....	42
Figura 2.10 Modelo Lógico.....	43
Figura 2.11 Arquitectura de Solución de Inteligencia Institucional para el SUIN.....	44
Figura 3.1 Correspondencia de la tabla temporal de persona.....	46
Figura 3.2 Correspondencia de la tabla temporal de cambio de dirección.....	47
Figura 3.3 Correspondencia de la tabla temporal de trámite.....	48
Figura 3.4 Correspondencia de la dimensión oficina.....	49
Figura 3.5 Flujo de proceso general de carga.....	52
Figura 3.6 Flujo del proceso ETL de Persona.....	53
Figura 3.7 Jerarquía de la dimensión Cambio de Dirección.....	54
Figura 3.8 Jerarquía de la dimensión Fecha.....	54
Figura 3.9 Reporte de grupos etarios en Cuba.....	55
Figura 3.10 Reporte de densidad poblacional.....	56



Índice de figuras

Figura 3.11 Reporte del indicador nivel de fiabilidad del servicio del CMI	57
Figura 3.12 Diagrama de despliegue	58
Figura 3.13 Comportamiento de los Indicadores	62
Figura 3.14 Resultado de la consulta efectuada sobre la Base de Datos del SUIN.....	63
Figura 3.15 Reporte mostrado en la aplicación web.	64
Figura 3.16 Acta de aceptación del producto	65



Índice de tablas

Tabla 1.1 Ventajas y desventajas de un Almacén de Datos	13
Tabla 2.1 Indicadores y perspectivas	32
Tabla 2.2 Descripción de los campos de la perspectiva Cambio de Dirección	38
Tabla 2.3 Descripción de los campos de la perspectiva Trámite.....	39
Tabla 2.4 Descripción de los campos de la perspectiva Fecha	39
Tabla 2.5 Descripción de los campos de la perspectiva Oficina.....	39
Tabla 3.1 Tablas del Almacén de Datos.....	49
Tabla 3.2 Secuencias del Almacén de Datos	50
Tabla 3.3 Índices del Almacén de Datos	51
Tabla 3.4 Lista de chequeo de la Solución de Inteligencia Institucional.....	61
Tabla 3.5 Comparación de los tiempos de respuesta en el acceso a la información.....	62



Introducción

El ambiente del actual entorno empresarial, exige un uso cada vez más eficiente de la información disponible. Los negocios son una competición, donde sobrevivir depende de la capacidad para actuar rápidamente en un ambiente cambiante, pues la información es la clave para obtener una ventaja competitiva, y de acuerdo al uso que cada institución le dé a este valioso recurso se podrá incrementar los niveles de eficiencia y productividad.

En este contexto, juegan un papel esencial las soluciones de Inteligencia de Negocio (término conocido en inglés como *Business Intelligence*), definidas en 1989 por Howard Dresner (1) como "... un conjunto de conceptos y métodos para mejorar el proceso de decisión utilizando un sistema de soporte basado en hechos...". Estas ofrecen a las organizaciones un modo de alinear las acciones con la estrategia empresarial, pues proporcionan capacidades de reporte, análisis e integración de datos con el fin de optimizar la rapidez y calidad de la toma de decisiones en todos los niveles de la organización.

Las soluciones de Inteligencia de Negocios aunque son variadas, comparten un propósito común: el proceso de apoyo a la toma de decisiones. Además, están enfocadas al óptimo aprovechamiento de la información que está al alcance de la organización. La aplicación de este tipo de soluciones se ha extendido por todo el mundo, pero actualmente la mayor parte se encuentra concentrada en los países desarrollados; debido a que estos poseen altos niveles de desarrollo de su economía, de la competitividad y de la cultura estratégica de los directivos de las altas empresas. Encabezando este movimiento se identifican a Europa y Estados Unidos de América (2).

La Inteligencia de Negocios no queda restringida al entorno empresarial, pues sus principios y técnicas también se aplican a la Administración Pública, que gana en precisión al advertir en su desarrollo las particularidades de la tarea gubernamental, surgiendo el término de Inteligencia Institucional. El objetivo fundamental de la Inteligencia Institucional es incrementar la satisfacción ciudadana a partir mejorar el servicio prestado a la población.

Desde el año 2009 como parte del proceso de informatización en la sociedad cubana, el Ministerio del Interior en colaboración con el Departamento de Soluciones Integrales (DSI) del Centro de Identificación y Seguridad Digital, adjunto a la Universidad de las Ciencias Informáticas, desarrollan el Sistema Único de Identificación Nacional.



El despliegue de este sistema por todo el país ha permitido a las oficinas de trámites de la Dirección de Identificación y Registro (DIR) procesar a diario de 8000 a 9000 trámites, los cuales son almacenados en una Base de Datos Relacional, que se ha comprobado que tiene un ritmo de crecimiento mensual de más de 3 Gb, principalmente como resultado de las trazas de los trámites.

Además de los procesos transaccionales que se llevan a cabo diariamente, el SUIN cuenta con un módulo de reportes que sirve de apoyo para la toma de decisiones. Al ser implantado el sistema las solicitudes de estos reportes demoraban un promedio de seis minutos para ser procesadas y mostradas, lo cual no obedece a la necesidad de inmediatez de la información (3).

Por otro lado, el proceso de recolección y análisis de información del Cuadro de Mando Integral (CMI) del MININT, el cual establece los indicadores organizacionales y objetivos estratégicos, se realiza manualmente con lentitud y sujeto a errores humanos, lo que afecta la toma de decisiones en la organización con la inmediatez y calidad que se precisa.

Vale señalar, que los funcionarios del MININT presentan la necesidad de obtener informaciones precisas sobre los datos almacenados, y en este proceso la Base de Datos del SUIN recibe una carga adicional pues al mismo tiempo que se ejecutan las consultas del ambiente operacional (como la inserción de la información resultante de diversos trámites), se efectúan las consultas requeridas por los reportes.

En el año 2011 se desarrolló una aplicación de Inteligencia de Negocio, de la cual se perdió toda la implementación de los reportes debido a que durante el movimiento del ambiente de desarrollo hacia el MININT fueron formateados los servidores donde se encontraba instalada la aplicación. Además, la misma no se implementó de acorde a los requisitos de software que requiere el cliente.

Teniendo en cuenta la situación problemática referida anteriormente, se define como **problema de investigación**: ¿Cómo mejorar el proceso de toma de decisiones para el Sistema Único de Identificación Nacional?

A partir del problema de investigación planteado se define como **objeto de estudio**: La Inteligencia de Negocios especificándose como **campo de acción**: la solución de Inteligencia Institucional para el Sistema Único Identificación Nacional.

Se propone entonces como **objetivo general**: Desarrollar una solución de Inteligencia Institucional que permita mejorar el proceso de toma de decisiones en el Sistema Único de Identificación Nacional.



Del análisis del objetivo general se derivan los siguientes **objetivos específicos**:

1. Desarrollar el marco teórico de la investigación.
2. Identificar los requisitos de información para la solución de Inteligencia Institucional del SUIN.
3. Realizar el diseño del Almacén de Datos para la solución de Inteligencia Institucional del SUIN.
4. Realizar la implementación del Almacén de Datos para la solución de Inteligencia Institucional del SUIN.
5. Desarrollar la aplicación web para la solución de Inteligencia Institucional del SUIN.
6. Validar la solución de Inteligencia Institucional para el SUIN.

Para dar cumplimiento a los objetivos trazados, se realizan las siguientes **tareas de investigación**:

1. Investigación de los principales conceptos asociados a las soluciones de Inteligencia de Negocio y específicamente la Inteligencia Institucional.
2. Revisión bibliográfica para la definición de la metodología a seguir para el desarrollo de la solución de Inteligencia Institucional para el SUIN.
3. Análisis de las herramientas de modelado, desarrollo, gestión de datos y reportes.
4. Análisis de la Arquitectura de un Data Warehouse.
5. Entrevista con funcionarios y especialistas del MININT para la identificación de los requisitos de información del SUIN, así como los indicadores organizacionales del MININT.
6. Análisis del modelo de datos del SUIN.
7. Entrevista a desarrolladores para el análisis del negocio de los distintos módulos del SUIN.
8. Identificación y descripción de las medidas, dimensiones y hechos.
9. Estructuración del modelo conceptual y el modelo lógico de datos.
10. Definición de la arquitectura del Almacén de Datos.
11. Diseño de los procesos de integración de los datos.
12. Implementación del modelo de datos.
13. Implementación de los cubos multidimensionales.
14. Implementación de los flujos de procesos.
15. Implementación de los reportes candidatos.
16. Realización de listas de chequeo para validar la solución de Inteligencia Institucional del SUIN.
17. Realización de las pruebas de integridad de datos al almacén.
18. Realización de las pruebas de aceptación del software para validar la solución de Inteligencia Institucional del SUIN.



Los **métodos de investigación** científica empleados son:

Métodos Teóricos:

- **Histórico-Lógico:** Durante el estudio de la evolución de la Inteligencia de Negocio como ayuda en el proceso de toma de decisiones, así como en la situación actual de su aplicación a las instituciones públicas.
- **Analítico-Sintético:** Durante el análisis del objeto de estudio a partir de diferentes fuentes bibliográficas, donde se sintetiza la información examinada que servirá de fundamento para el diseño e implementación de la solución de Inteligencia Institucional.
- **Modelación:** Para demostrar, verificar y validar la solución de Inteligencia Institucional por medio de modelos que permitan una comprensión adicional de las características del proceso de Inteligencia Institucional y la visualización de la posible solución al problema planteado.

Método Empírico:

- **Entrevista:** Para realizar entrevistas a funcionarios del MININT con el objetivo de determinar las características y necesidades de la organización.

Estructura del documento

El presente documento se encuentra dividido en tres capítulos estructurados de la siguiente forma:

Capítulo 1 Fundamentos teóricos: se describen los principales conceptos relacionados con el dominio del problema, se identifican y detallan las etapas de la arquitectura de una solución basada en almacenes de datos, y se analizan las principales metodologías y herramientas para darle solución al problema planteado.

Capítulo 2 Análisis y Diseño del Almacén de Datos: se muestran los resultados y pasos a seguir para el análisis y el diseño del Almacén de Datos según propone la metodología HEFESTO. Se diseña el modelo lógico del almacén y se propone la arquitectura de la solución.

Capítulo 3 Implementación y prueba: se describe la implementación de la solución de Inteligencia Institucional para el SUIN, se exponen los requisitos necesarios para el despliegue de la solución y se detallan las pruebas aplicadas al software.



Capítulo 1 Fundamentos teóricos

1.1 Introducción

En este capítulo se describen los fundamentos teóricos y conceptuales acerca de la Inteligencia de Negocios tanto para las organizaciones con retorno monetario como las que no. Se introducen conceptos sobre Almacenes de Datos, y se justifica su aplicación a partir de sus contribuciones y alcances. Se conceptualiza la Arquitectura de una solución basada en Almacenes de Datos y se exponen los elementos que la definen y la caracterizan. Por último se justifica tanto la metodología que guiará el proceso de desarrollo del Almacén de Datos como las herramientas que se utilizarán para su diseño e implementación.

1.2 Inteligencia de Negocio

En la actualidad, la capacidad para tomar decisiones ágiles y acertadas en los negocios se ha convertido en una de las claves que permiten a las organizaciones ser exitosas y permanecer en el tiempo. Pues prácticamente todas las empresas disponen de bases de datos que almacenan datos sobre sus actividades y colaboradores, por ende resulta lógico pensar que estos datos podrían ser agrupados, refinados y analizados para intentar extraer información que permita ayudar en la toma de decisiones de la empresa.

Encontrar patrones de conducta en los clientes, presentar información en tiempo real sobre las distintas sucursales de la empresa, optimizar procesos internos de la organización como la prestación de servicios, son ejemplos que muestran qué se podría obtener a partir de los datos de la empresa. Este hecho, la conversión de los datos operativos en información que dé soporte a la toma de decisiones, es lo que se conoce como Inteligencia de Negocio.

Varios autores han conceptualizado la Inteligencia de Negocios, por lo que a continuación se mostrarán conceptos relativos a este tema:

La Inteligencia de Negocios o *Business Intelligence* es un conjunto de técnicas y herramientas tanto de gestión empresarial como de aplicación tecnológica, que permiten a partir de la formulación estratégica y teniendo como objetivo dar soporte a los procesos de planeación y control en las organizaciones, la extracción e integración de los datos, que son generados como resultado de la operación de las diferentes áreas funcionales en una organización, su posterior procesamiento y distribución en forma de información (4).



Capítulo 1: Fundamentos teóricos

La Inteligencia de Negocios es mucho más que una actitud empresarial o una tecnología a disposición de las organizaciones; es un marco de referencia para la gestión del rendimiento empresarial, un ciclo continuo por el cual las compañías definen sus objetivos, analizan sus progresos, adquieren conocimiento, toman decisiones, miden sus éxitos y comienzan el ciclo nuevamente (5).

Se entiende por *Business Intelligence* al conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización (6).

Esta tecnología, hace posible la gestión de información tanto organizada como no estructurada, resultante de la sinergia del procesamiento de datos y el uso de herramientas analíticas que fusionan la gestión de la información y del conocimiento, convirtiéndose en una ventaja altamente competitiva a nivel organizacional (7).

Según Dresner (1) quien fue uno de los que primero conceptualizó esta expresión, la Inteligencia de Negocio: *“Es un paraguas bajo el que se incluye un conjunto de conceptos y metodologías cuya misión consiste en mejorar el proceso de toma de decisiones en los negocios basándose en hechos y sistemas que trabajan con hechos”*.

Teniendo en cuenta las perspectivas antes expuestas se puede decir entonces que la Inteligencia de Negocio, es el conjunto de tecnologías y herramientas enfocadas a la creación y administración del conocimiento mediante el análisis de datos existente en una organización, cuyo objetivo fundamental es contribuir a tomar decisiones que mejoren el desempeño de la organización y promueva su ventaja competitiva en el mercado. La definición antes expuesta se representa a través de la siguiente imagen:



Figura 1.1 Pirámide del conocimiento en Business Intelligence.

En otras palabras, a partir de la combinación de tecnologías, herramientas y procesos, se podrán transformar los grandes volúmenes de datos acumulados de las empresas en información útil para generar conocimiento, destinado a optimizar el proceso de toma de decisiones en los negocios y satisfacer las necesidades de una gran variedad de usuarios finales.

1.3 Inteligencia Institucional

La Inteligencia Institucional es un concepto que surge teniendo en cuenta los principios y técnicas de la Inteligencia de Negocio, pero aplicados a instituciones públicas o gubernamentales, las cuales cuentan con procesos, actividades y tareas de complejidad particular. La principal diferencia entre ambos términos reside en el enfoque, pues mientras que la Inteligencia de Negocios posee un enfoque comercial donde debe primar el retorno monetario que se mide mediante el retorno de inversión (ROI por sus siglas en inglés), la Inteligencia Institucional persigue un objeto social, donde busca mejorar la calidad del servicio prestado a la población y disminuir el tiempo y coste del mismo, para así incrementar la satisfacción ciudadana.

La Inteligencia Institucional no difiere de la Inteligencia de Negocios en cuanto a principios y técnicas de diseño (8); pero su enfoque orientado a la satisfacción ciudadana, ayuda a las entidades de la administración pública a detectar deficiencias y a mejorar sus procesos.



Figura 1.2 Inteligencia Institucional

1.3.1 ¿Por qué usar una Solución de Inteligencia Institucional en una organización?

Para que una organización alcance el éxito debe ser capaz de tomar decisiones precisas y de forma rápida, pero en muchas ocasiones ocurre lo siguiente (9):

Se poseen los datos pero se carece de la información respectiva para tomar las decisiones. Por ejemplo, se almacenan los datos de los clientes pero no es posible distinguir patrones de comportamientos entre ellos, faltando información para plantear una estrategia que permita maximizar el rendimiento de la empresa.

Falta integración. Muchas organizaciones poseen múltiples sistemas transaccionales en los diversos departamentos, pero carecen de una visión global de la empresa. Sus bases de datos no suelen estar integradas, lo que implica la existencia de islas de información, lo cual dificulta la toma de decisiones.

Existen largos tiempos de respuesta y poca agilidad. A medida que las preguntas que se buscan responder se vuelven más complejas, y el volumen de datos aumenta, los tiempos de respuesta se traducen en una incómoda espera que obstaculiza la fluidez del trabajo.

Escasa flexibilidad. En muchas ocasiones el usuario tiene que amoldarse a los informes predefinidos que se configuraron en el momento de la implementación de un sistema, los cuales no siempre responden a sus dudas reales.

1.3.2 Ventajas de una solución de Inteligencia Institucional

Es válido señalar que son múltiples las ventajas (2) (10) que las soluciones de Inteligencia Institucional proporcionan a las empresas, al permitir que los datos se conviertan en un centro de beneficios, facilitando el análisis de información útil a las organizaciones, entre las que se destacan:



- Brinda a las empresas la capacidad de analizar de una manera rápida por qué pasan las cosas, por lo que se pueden identificar y corregir situaciones antes de que se conviertan en problemas; como consecuencia se minimiza el riesgo, la incertidumbre y se pueden conseguir nuevas oportunidades o readaptarse frente a la ocurrencia de sucesos inesperados.
- Puede generar escenarios, pronósticos y reportes que apoyen a la toma de decisiones, cerrando el círculo que hace pasar de la decisión a la acción.
- Al estar más integrada la información de la organización en una fuente única de fácil acceso, se reducirá el tiempo mínimo requerido para recoger toda la información relevante del negocio y resultará más sencillo e intuitivo para cualquiera de sus trabajadores obtener y analizar la información que necesita, posibilitando que estén más capacitados para tomar una mejor decisión.
- Automatiza la asimilación de la información, debido a que la extracción y carga de los datos necesarios se realiza a través de procesos predefinidos.
- Proporciona herramientas de análisis para establecer comparaciones y tomar decisiones.
- Impulsa un mejor posicionamiento de la organización en el ámbito de la prestación de los servicios, mejora su desempeño y eleva el prestigio de la misma.

El ambiente del actual entorno empresarial, exige un uso cada vez más eficiente de la información disponible. Pues lograr la satisfacción ciudadana es una competencia, donde permanecer arriba depende de la capacidad para actuar rápidamente en un ambiente cambiante, por tanto, las soluciones de Inteligencia Institucional resultan idóneas ya que brindan a las organizaciones la capacidad de analizar de una forma rápida por qué pasan las cosas, pues ponen a disposición de los usuarios la información exacta en el lugar y momento adecuados, posibilitando así optimizar la toma de decisiones e incrementar la efectividad de cualquier empresa. La Inteligencia Institucional ayuda a rastrear lo que en realidad funciona y lo que no.

1.4 Data Warehouse

Como en las soluciones de Inteligencia Institucional resulta necesario gestionar datos que se encuentran guardados en diversos formatos y fuentes, para luego depurarlos, integrarlos y almacenarlos en un único destino que permita su posterior análisis y exploración, resulta necesario contar con una herramienta que satisfaga estas necesidades, esto se logra con el *Data Warehouse* (también conocido como almacén o factoría de datos).

El almacén provee un ambiente que concentra toda la información de interés para la organización. Por consiguiente, como refleja Méndez (11) constituye una tecnología para el manejo de la información



construido sobre la base de optimizar el uso y análisis de la misma, siendo utilizado por las organizaciones para adaptarse a los vertiginosos cambios en los mercados. Su función esencial es ser la base de un sistema de información gerencial, es decir, debe cumplir el rol de integrador de información proveniente de fuentes funcionalmente distintas (Bases Corporativas, Bases propias, de Sistemas Externos, etc.) y brindar una visión integrada de dicha información, especialmente enfocada hacia la toma de decisiones por parte del personal jerárquico de la organización.

Una de las definiciones más famosas sobre Almacenes de Datos, es la de W. H. Inmon (12), reconocido como el padre del *Data Warehouse*, quien define: “*Un Data Warehouse es una colección de datos orientada a temas, integrada, variante en el tiempo y no volátil usados en la estrategia de toma de decisiones administrativas*”.

1.4.1 Características de un Data Warehouse

Según Inmon (10) (13) (14) las características de los Almacenes de Datos son las siguientes:

- Orientado al tema: pues la información se clasifica en base a los aspectos que son de interés para la empresa, para facilitar su acceso y entendimiento por parte de los usuarios finales.
- Integrado: cuando los datos, tanto internos como externos, se mueven al *Data Warehouse* desde las aplicaciones orientadas al ambiente operacional, deben ser consolidados en una instancia antes de ser agregados al almacén. A este proceso se lo conoce como Extracción, Transformación y Carga de Datos (ETL, por sus siglas en inglés).
- De tiempo variante o histórico: toda la información del almacén es requerida en algún momento. Al realizar algún cambio en la información, el almacén puede guardar los estados anteriores de la misma.
- No volátil: la información es útil para el análisis y la toma de decisiones solo cuando es estable. Mientras que en el ambiente operacional los datos cambian constantemente (se actualizan a través del insertar, actualizar y eliminar), en el almacén la manipulación básica de los datos es mucho más simple, pues existen dos únicos tipos de operaciones: la carga inicial de datos y el acceso a los mismos, no hay actualización de datos (en el sentido general de actualización), por tanto, en el Almacén de Datos la información podrá ser leída pero no modificada.

De acuerdo con Kimball (15) (16), algunos de los objetivos fundamentales de un Almacén de Datos son: lograr que la información de la organización sea accesible ya que los contenidos deben ser obvios y



navegables, lograr que la información de la organización sea consistente pues en el almacén se eliminarán las inconsistencias en los procesos de transformación y dar soporte a la toma de decisiones pues luego de que el almacén de datos posea la información cargada y actualizada, permitirá mejorar la toma de decisiones.

Es válido señalar que un factor fundamental en la necesidad de implementar un Data Warehouse es la excesiva complejidad de los modelos relacionales que soportan los sistemas operacionales, lo cual dificulta en extremo la extracción de información para su análisis. Pues como dijera Inmon (12) “todos los datos pueden ser de gran valor si son administrados correctamente”, y el objetivo principal de un proyecto *Data Warehouse* es que el personal de la empresa lo utilice para satisfacer sus propias necesidades de información, mejorando en consecuencia la competitividad de la empresa.

1.4.2 Razones que justifican la creación de un Data Warehouse

A continuación se muestran varias razones que justifican la creación de un Data Warehouse (17), entre las que se encuentran:

- Rendimiento: se tarda mucho menos en acceder a los datos del repositorio del Data Warehouse que en hacer una consulta a varias bases de datos distintas. Además hacer consultas complicadas a las bases de datos de los sistemas operacionales puede empeorar el tiempo de respuesta de estos sistemas para otros usuarios.
- Múltiples orígenes de datos: combinar los datos de distintas fuentes suele ser una tarea bastante complicada para las personas encargadas de tomar decisiones con esa información. Normalmente hay que homogenizar los datos de una forma u otra, pues es probable que no se utilicen los mismos criterios de almacenamiento (nombres de las entidades, atributos considerados, tipos, etc.) en las bases de datos departamentales. Sin embargo, en el almacén los datos se homogenizan durante el proceso de carga.
- Limpieza de los datos: las empresas no siempre cuentan con aplicaciones únicas para cada parte operativa del negocio, sino que pueden poseer replicaciones y distintos sistemas para atender un mismo conjunto de operaciones, y en esos casos es probable que las bases de datos de los sistemas operacionales contengan datos duplicados, a veces erróneos, superfluos o incompletos. Estos datos se corrigen durante el proceso de carga al Data Warehouse.



Capítulo 1: Fundamentos teóricos

- **Ajustes:** en ocasiones se hace necesario un ajuste de los datos para posibles comparaciones. Por ejemplo: si se está combinando información financiera de distintos países habrá que ajustar toda esta información conforme a una única norma contable para hacerla comparable. Esos ajustes se realizan en el almacén durante el proceso de carga.
- **Periodicidad:** La periodicidad de los datos en las distintas bases transaccionales de datos puede ser distinta: diaria, semanal o mensual, y es necesario entonces realizar homogeneización de los datos.
- **Datos históricos:** Los datos históricos no se suelen guardar en los sistemas operacionales, pero son un elemento esencial de cualquier análisis. El Almacén de Datos es el lugar adecuado para estos datos.
- **Agregados:** Muchas veces para tomar decisiones, no es necesario entrar en la línea de más detalle durante el análisis, en este sentido, en el almacén se suelen guardar sólo los agregados necesarios (por ejemplo el número de clientes trimestralmente que acceden a cada oficina, el tipo de trámite más solicitado en función de la edad del público objetivo, etc.).

1.4.3 Ventajas y Desventajas

Algunas de las ventajas y desventajas (10) más relevantes que trae aparejada la implementación de un Almacén de Datos, que muestran de mejor modo sus características son:

Ventajas	Desventajas
Supone una optimización tecnológica para el análisis de la información y la generación de reportes.	Requiere de una continua actualización y mantenimiento.
Permite planear estrategias en la organización de forma más efectiva.	Puede haber resistencia al cambio por parte de los usuarios.
La información se encuentra disponible todo el tiempo, aun si los sistemas fuentes no lo están.	Una vez implementado puede ser complicado añadir nuevas fuentes de datos.
Mejora las relaciones con los clientes.	Los almacenes de datos se pueden quedar obsoletos relativamente pronto.
Permite realizar consultas propias del ámbito del Data Warehouse, que fuera de él sería imposible de conseguir, por restricciones de recursos y tiempo.	Los beneficios del Almacén de Datos son apreciados en el mediano y largo plazo.
Reduce los tiempos de respuesta.	Se subestima el esfuerzo necesario para su diseño y creación.



No afecta el procesamiento en los sistemas fuentes, pues todos los análisis necesarios para el soporte a la toma de decisiones se trasladan hacia el Almacén de Datos, el cual se actualiza cuando los sistemas operacionales están fuera de línea, eludiendo así interrupciones innecesarias.	Requiere una gran inversión, debido a que su correcta construcción no es tarea sencilla y consume muchos recursos, además, su implementación implica desde la adquisición de herramientas de consulta y análisis, hasta la capacitación de los usuarios.
Facilita la aplicación de técnicas estadísticas de análisis y modelado para encontrar relaciones ocultas entre los datos del almacén.	
Transforma datos orientados a las aplicaciones en información orientada a la toma de decisiones.	
Existe información extra en el Data Warehouse, pues muchos datos son pre-calculados. Como consecuencia la información que más se necesita ya está lista cuando se requiere.	
Reduce la generación excesiva de información, pues sólo utiliza los datos requeridos y desecha cualquier otra información innecesaria.	
Permite predecir el comportamiento futuro con un alto porcentaje de certeza, basado en el entendimiento del pasado.	

Tabla 1.1 Ventajas y desventajas de un Almacén de Datos

Como el esfuerzo de las organizaciones debe estar en convertir la enorme cantidad de datos que poseen en sus bases de datos corporativas en información útil, para finalmente extraer el mayor conocimiento posible; reunir la información apropiada desde diversas fuentes de aplicación en un ambiente integral centralizado, simplificará el problema de acceso a la información y en consecuencia, acelerará el proceso de análisis, consultas y el menor tiempo de uso de la información. El Almacén de Datos organiza y almacena los datos que se necesitan para el procesamiento analítico e informático, con el propósito de responder a preguntas del negocio, brindarles a los usuarios finales una interfaz amigable, comprensible y fácil de utilizar, donde los mismos puedan tomar decisiones sobre los datos sin tener que poseer demasiados conocimientos informáticos.

1.4.4 Data Marts

En el contexto de los almacenes de datos es preciso mencionar el término de *Data Mart*, el cual representa una Base de Datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Además, se caracteriza por disponer la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento. Puede ser



alimentado desde los datos de un *Data Warehouse*, o integrar por sí mismo un compendio de distintas fuentes de información. (57)

En este sentido, el *Data Mart* es como un pequeño *Data Warehouse*, con un volumen de datos más limitados, para un determinado número de usuarios y un área funcional específica de la organización. En otras palabras, es una Base de Datos multidimensional orientada a un tema específico que permite dar soporte a una empresa pequeña, departamento o área de negocio de una empresa grande.

Según Bernabéu(10) los *Data Marts* ofrecen varios beneficios, por ejemplo: son fáciles de implementar, conllevan poco tiempo de construcción y puesta en marcha, permiten manejar información confidencial, reflejan rápidamente sus beneficios y cualidades, y reducen la demanda del depósito de datos.

Es importante señalar que el uso efectivo de los *Data Marts* en un ambiente de almacenes de datos, representa un factor fundamental para la efectividad del Almacén de Datos.

1.5 Arquitectura del Data Warehouse

Para comprender íntegramente la finalidad del almacén es importante conocer cómo es el proceso empleado para su construcción, el cual es conocido como *Data Warehousing*, para entender este proceso y como se relacionan todos los componentes involucrados en el almacén, es esencial contar con una arquitectura fiable.

A continuación se definirán los componentes que integran la Arquitectura básica de *Data Warehousing*, pero primero se dará una breve definición de este término. “Una Arquitectura *Data Warehousing* es una forma de representar la estructura total de los datos, la comunicación, el procesamiento y la presentación de los mismos, ante los usuarios finales que disponen de una computadora dentro de la empresa.” (19) En la siguiente figura se muestra las etapas de la arquitectura:

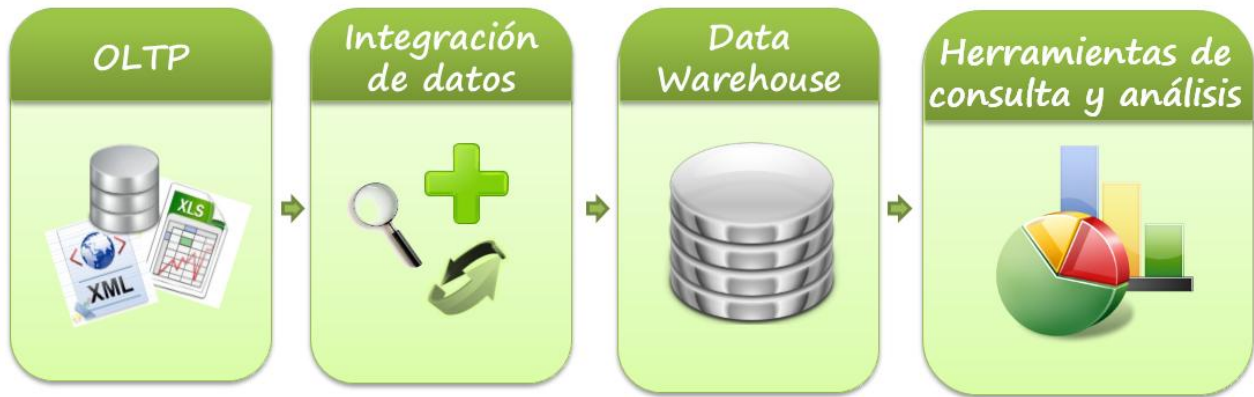


Figura 1.3 Arquitectura de un almacén de datos.

1.5.1 OLTP

Los sistemas OLTP (*On Line Transaction Processing*) son fuentes de datos orientadas al procesamiento de transacciones. Una transacción genera un proceso atómico y que puede involucrar operaciones de inserción, modificación y borrado de datos. Teniendo en cuenta esto OLTP representa toda aquella información transaccional que genera la empresa en su accionar diario.

OLTP tiene como objetivo soportar el procesamiento de información diaria de las empresas, por ende el énfasis recae en maximizar la capacidad transaccional de sus datos. Su estructura es altamente normalizada, para brindar mayor eficiencia a sistemas con muchas transacciones que acceden a un pequeño número de registros y están fuertemente condicionadas por los procesos operacionales que deben soportar, para la óptima actualización de sus datos (10).

Esta estructura es ideal para brindar consultas sobre los datos cargados, pero OLTP no ha sido diseñado para proporcionar funciones potentes de síntesis, análisis y consolidación de los datos, en cambio los esquemas de Data Warehouse están diseñados para poder llevar a cabo procesos de consulta y análisis, aligerar la carga de las bases de datos transaccionales y luego tomar decisiones estratégicas y tácticas de alto nivel.

Actualmente, podemos encontrar distintos tipos de OLTP entre los cuales podemos mencionar:

- Las bases de datos relacionales.
- Archivos de texto, XML y hojas de cálculos.
- Informes semanales, mensuales, anuales, etc.



En el [Anexo 1](#) se muestra una tabla comparativa que explicita las principales diferencias entre los dos ambientes.

1.5.2 Integración de Datos – ETL

Como parte de la arquitectura del Almacén de datos uno de los componentes esenciales y quizás el más crítico, es el proceso de integración de datos en un repositorio que permita almacenar la información ya consolidada para ser explotada por herramientas de análisis. Pues, los datos al existir en distintas bases de datos o ser administrados por diferentes gestores de bases de datos presentan formatos inconsistentes y/o son codificados de distintas formas. Consiguientemente, estas inconsistencias deben resolverse para que los datos sean almacenados en el Data Warehouse.

Curto Díaz (6), define la integración de datos como el conjunto de aplicaciones, productos, técnicas y tecnologías que permiten una visión única y consistente de nuestros datos de negocio, donde las aplicaciones y productos capacitan la integración mediante el uso de las tecnologías y técnicas. En este contexto, el proceso de integración de datos agrupa una serie de técnicas y subprocesos que se encargan de llevar a cabo todas las tareas relacionadas con la extracción, manipulación, control, integración, depuración de datos, carga y actualización del *Data Warehouse*. Es decir, todas las tareas que se realizan desde que se toman los datos de los diferentes OLTP hasta que se cargan en el Almacén de datos. Una de estas técnicas son los procesos de Extracción, Transformación y Carga de Datos (ETL, por sus siglas en inglés).

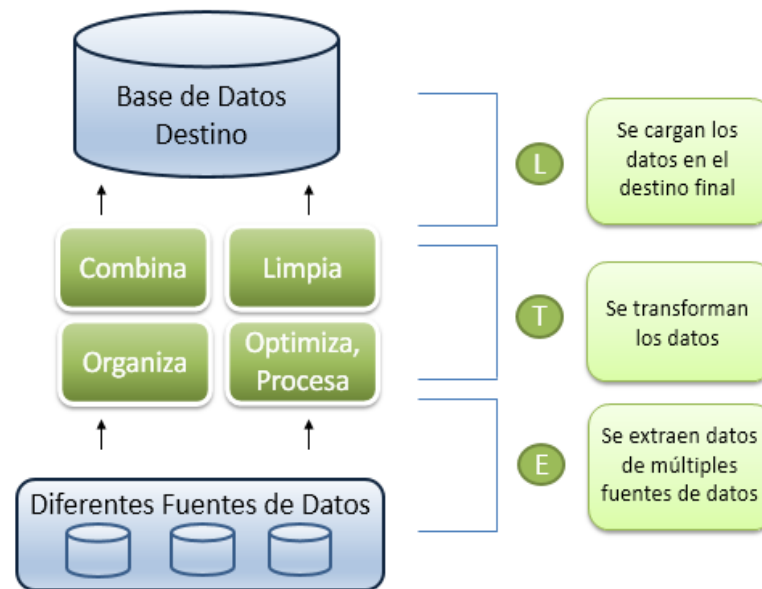


Figura 1.4 Proceso ETL

En el proceso de Extracción se tiene un grupo de técnicas enfocadas en tomar solo los datos requeridos desde los OLTP y mantenerlos en un almacenamiento intermedio; en el proceso de Transformación se analizan los datos para verificar que son correctos y válidos; en el proceso de Carga de Datos se aplican técnicas propias de la carga y actualización del almacén.

Un desarrollo reciente en el software para el desarrollo de las técnicas ETL es la aplicación de procesamiento paralelo. Esto ha permitido desarrollar una serie de métodos para mejorar el rendimiento general de los procesos ETL cuando se trabaja con grandes volúmenes de datos (20).

1.5.3 Almacén de datos físico y su estructura multidimensional

La tecnología *Data Warehousing* tiene una orientación analítica, ya que su objetivo es permitir el análisis de la información para la toma de decisiones, y esto impone un procesamiento y pensamiento distinto, el cual se sustenta por un modelado de Bases de Datos propio, conocido como modelado multidimensional, donde la idea fundamental es que el usuario visualice fácilmente la relación que existe entre los distintos componentes del modelo.

Para mejor comprensión del modelo multidimensional es preciso destacar conceptos esenciales como: hechos, dimensiones, medidas, atributos y jerarquías (9) (21) (22) (23).



Capítulo 1: Fundamentos teóricos

- Hecho: es una operación que se realiza dentro de la organización en un tiempo determinado, que es objeto de análisis para la toma de decisiones y se utiliza para crear medidas o indicadores, a través de acumulaciones preestablecidas. Los hechos están implícitamente definidos por la combinación de los valores de las dimensiones, pues los registros de los hechos poseen una llave primaria compuesta, que está definida por las llaves primarias de las dimensiones. Generalmente los hechos son casi siempre valores numéricos, enteros o reales, a los cuales se les puede aplicar alguna función matemática determinada, como la agregación.
- Medida o indicador del desempeño: Son cuantificadores del desempeño de un hecho que se obtienen realizando acumulaciones sobre el mismo. La información que brinda una medida es usada por los usuarios en sus consultas para evaluar el desempeño de un área temática.
Dimensión: Una dimensión es una característica de un hecho y constituye un calificador mediante el cual el usuario podrá filtrar y manipular la información almacenada, es decir, podrá analizar los hechos. Las dimensiones significan el quién, qué, cuándo y dónde de los datos representados en las tablas de hechos. Los componentes de una dimensión se denominan niveles y se organizan en jerarquías.
- Atributo: Una dimensión se identifica por poseer más de un atributo, donde un atributo es una característica de una dimensión. Los atributos dimensionales son un rol determinante en un almacén de datos, pues son la fuente de todas las necesidades que debieran cubrirse. Esto significa que mientras más descriptivos, manejables y de buena calidad sean los atributos, mejor será el *Data Warehouse*.
- Jerarquías: Entre los atributos de una misma dimensión pueden existir jerarquías, que se definen por la posición relativa que toma un atributo con respecto a los otros miembros de la misma dimensión, formando en su totalidad una estructura de árbol. Puede darse también que una dimensión no necesite jerarquizarse debido a que ninguno de sus atributos posee una posición relativa con respecto a los otros. Las jerarquías son una opción que se ofrece a los usuarios para que puedan navegar según las características de cada dimensión, propiciando una mayor flexibilidad en el análisis de los datos, cada escalón en la jerarquía proporciona un nivel de análisis.

De acuerdo con Kimball (15) el modelo multidimensional “contiene la misma información que un modelo normalizado pero empaqueta los datos de acuerdo con un formato cuyas metas de diseño son la comprensibilidad del usuario final, el rendimiento ante las consultas y la resistencia al cambio”.



Capítulo 1: Fundamentos teóricos

Las bases de datos multidimensionales están orientadas a la realización de consultas complejas y ofrecen alto rendimiento. Dichas bases de datos pretenden superar las limitaciones de los sistemas de gestión de bases de datos relacionales en los procesos de análisis.

Las capacidades de las Base de Datos relacionales agregación (cláusulas *sum*, *avg*, *group by*, etc., de SQL) resultan muy limitadas si se quieren obtener resultados globales y la normalización dificulta una visión general de los datos, forzando a generar más y más tablas virtuales para proceder a un análisis medianamente complejo (9). Por tanto, el modelado multidimensional permite efectuar distintos tipos de análisis de maneras más detallada y eficaz. La forma de representar como están organizados los datos en un modelo multidimensional es a través de un cubo, el cual no necesariamente debe tener tres dimensiones (ver figura 1.5).

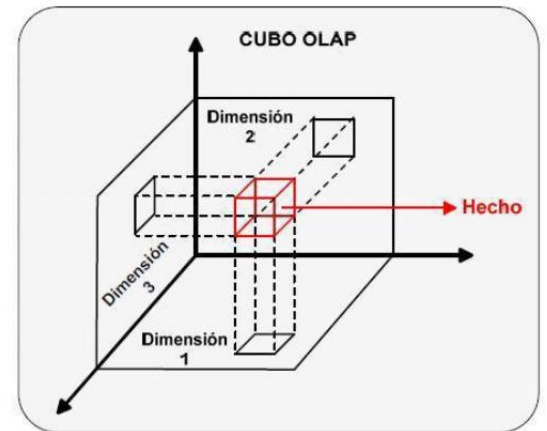


Figura 1.5 Estructura de un cubo multidimensional

Esquemas de diseño

Las bases de datos multidimensionales presentan tres variantes posibles de modelado conocidos como esquemas, los cuales constituyen la colección de tablas del *Data Warehouse*.

Esquema Estrella: Este esquema es denominado estrella por la forma que adopta, pues está compuesto por una tabla central, que es la de “Hechos”, y un conjunto de tablas adyacentes que son las de “Dimensiones”, las que se relacionan con la tabla de hechos a través de las claves primarias. En el modelo estrella las dimensiones no se normalizan, con esto se logra minimizar el número de uniones e incrementar el rendimiento de las consultas.

Vale destacar que esta manera de modelar los datos agiliza las consultas y reportes, pero no es apropiada para los procesos transaccionales porque la técnica de modelado bajaría el rendimiento y complicaría el proceso transaccional ya que la información está desnormalizada.

Un ejemplo de este modelo es el siguiente:

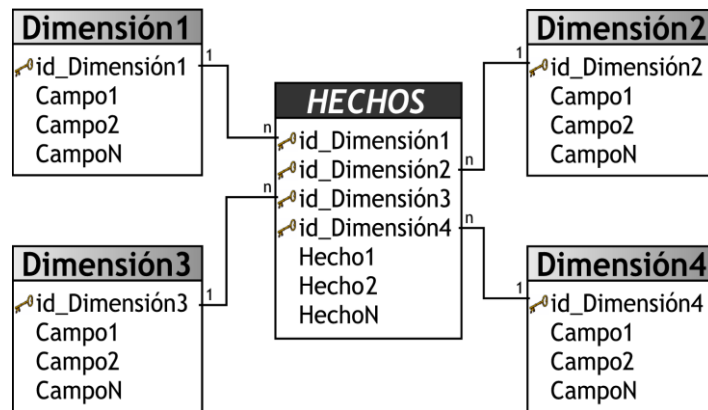


Figura 1.6 Esquema Estrella

Esquema Copo de Nieve: Es una variante del esquema estrella, pero la diferencia reside en que la estructura de las tablas de dimensiones están normalizadas. Cada tabla dimensión contiene sólo el nivel que es clave primaria en la tabla de hechos, junto con la clave foránea de su parentesco al nivel más cercano del diagrama. Por ello, si se implementa un modelo copo de nieve (ver figura 1.7) se pueden desarrollar niveles fuera de las dimensiones (datos normalizados) que permitan efectuar análisis de lo general a lo particular y viceversa. Vale añadir que este modelo es más cercano a un modelo de entidad relación, que al modelo en estrella.

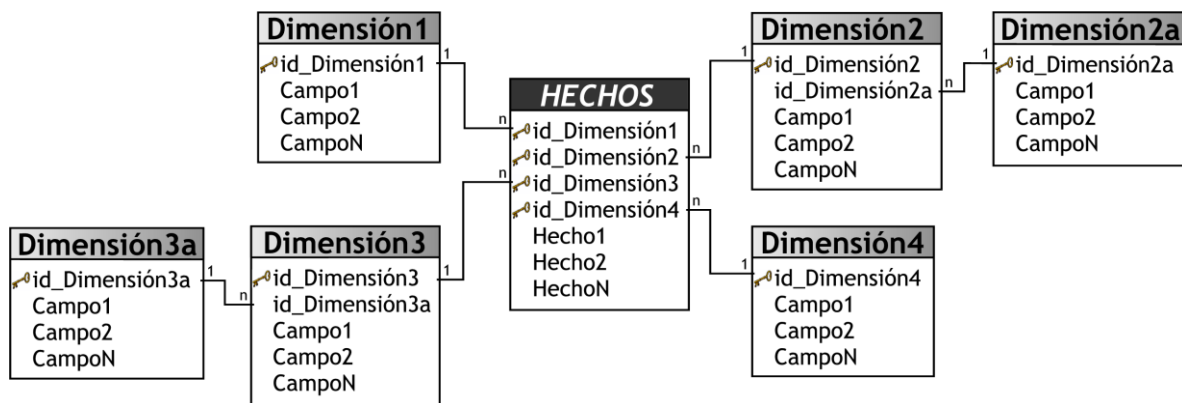


Figura 1.7 Esquema Copo de Nieve

La principal razón para usar modelos copos de nieve es que permite separar los datos de las dimensiones y que se elimina redundancia de datos y con ello la reducción del espacio de almacenamiento, pero se debe tener en cuenta que muchas uniones en las consultas puede sacrificar el desempeño del sistema al



generar las consultas. Esto se debe a que el gestor de consultas debe crear más tablas de dimensiones y relaciones entre tablas (JOINS) lo que provoca una caída significativa en la velocidad de las respuestas.

Esquema Constelación: El modelo constelación (ver [Anexo 2](#)) está compuesto por una serie de modelos estrella, donde existe una tabla de hechos principal y una o más tablas de hechos auxiliares. Estas tablas de hechos auxiliares no necesariamente están relacionadas con las mismas dimensiones de la principal, por lo que puede ocurrir que estén relacionadas con nuevas dimensiones o con un subconjunto de las dimensiones que referencian a la tabla de hechos principal.

Según Bernabéu (10) este modelo cuenta con varias características:

- Permite tener más de una tabla de hechos, por lo cual se podrán analizar más aspectos claves del negocio con un mínimo esfuerzo adicional de diseño.
- Contribuye a la reutilización de las tablas de dimensiones, ya que una misma tabla de dimensión puede utilizarse para varias tablas de hechos.
- No es soportado por todas las herramientas de consulta y análisis.

Teniendo en cuenta lo antes señalado, vale destacar que se decide usar el esquema estrella, pues es el más simple de interpretar y optimiza los tiempos de respuesta ante las consultas de los usuarios. Es soportado por casi todas las herramientas de consulta y análisis, facilitando la interacción con estas. Aunque presenta como punto en contra que se genera un cierto grado de redundancia al estar la información desnormalizada, pero el aumento de la información no es significativo. Resultando este modelo el más viable para la implementación de la solución.

Tipos de implementación de un DW

Los esquemas mencionados pueden ser implementados de varias maneras, mediante el Proceso Analítico en Línea Relacional (ROLAP, por sus siglas en inglés), mediante el Proceso Analítico en Línea Multidimensional (MOLAP, por sus siglas en inglés) y mediante el Proceso Analítico en Línea Híbrido (HOLAP, por sus siglas en inglés).

ROLAP: Es una arquitectura de modelado multidimensional que se caracteriza por almacenar los datos en una Base de Datos relacional (24). Cuando los usuarios finales ejecutan sus análisis multidimensionales a través del motor ROLAP, este transforma dinámicamente las consultas de los usuarios en consultas SQL,



Capítulo 1: Fundamentos teóricos

las cuales son ejecutadas sobre la Base de Datos relacional y finalmente se devuelven el resultado a los usuarios finales.

ROLAP brinda mucha flexibilidad, pues los cubos son generados dinámicamente al momento de ejecutar las consultas, posibilitando la obtención de consultas *ad hoc*¹. La principal desventaja de los sistemas ROLAP, es que los datos de los cubos generalmente se deben calcular cada vez que se ejecuta una consulta sobre ellos (excepto que se guarde alguna consulta precalculada). Esto provoca que ROLAP no sea muy eficiente en cuanto a la rapidez de respuesta ante las consultas de los usuarios.

MOLAP: En este modelo de almacenamiento, los datos se encuentran almacenados en estructuras multidimensionales. Para que estas estructuras puedan ser consultadas requiere que en una instancia previa se generen y calculen los cubos multidimensionales, posibilitando rapidez en la respuesta a las consultas de los usuarios, pues al no tener que calcular los cubos en tiempo de ejecución se mejora el rendimiento (9).

Este modelo de almacenamiento consigue la rapidez en las consultas a costa de mayores necesidades de almacenamiento, retardos en las modificaciones (que deberían ocurrir sólo ocasionalmente, por ejemplo, la adición de una dimensión) y largos procesos de carga y cálculo de acumulados. Vale destacar la disminución que presenta este tipo de implementación en cuanto a flexibilidad, pues cada vez que se requiere o es necesario realizar cambios sobre algún cubo, se debe recalcularse totalmente para que se reflejen las modificaciones llevadas a cabo.

HOLAP: es un sistema híbrido que combina el uso de las arquitecturas ROLAP y MOLAP. En una solución HOLAP, los volúmenes más grandes de datos (es decir, los registros detallados) se mantienen en una Base de Datos relacional, mientras que los datos agregados y precalculados se guardan en un almacén MOLAP independiente (9). Una desventaja que presenta HOLAP, es que resulta necesario realizar un análisis profundo de los distintos tipos de datos, a fin de definir donde se almacenarán.

La selección del modelo de almacenamiento depende de cuán importante sea el rendimiento de las consultas para los usuarios y de la tecnología disponible a utilizar. Por tanto, a continuación se mencionan diferencias entre los modelos para decidir cuál es el más adecuado a utilizar (24) (25):

¹ Consultas *ad hoc*: consultas que personaliza el usuario en tiempo real, en vez de estar atado a las consultas prediseñadas para informes.



- Los sistemas con alta volatilidad de los datos (aquellos en los que cambian las reglas de agregación y consolidación), requieren una arquitectura que pueda realizar esta consolidación *ad-hoc*. Los sistemas ROLAP soportan bien esta consolidación dinámica.
- Los ROLAP pueden crecer hasta un gran número de dimensiones, mientras que los MOLAP generalmente son adecuados para diez o menos dimensiones.
- Los ROLAP soportan análisis OLAP contra grandes volúmenes de datos elementales, mientras que los MOLAP se comportan razonablemente en volúmenes más reducidos (menos de 5 Gb)
- En el modelo ROLAP la respuesta a las consultas y el tiempo de procesamiento suelen ser más lentos que con los modos de almacenamiento MOLAP o HOLAP. Sin embargo, ROLAP permite a los usuarios ver los datos en tiempo real y ahorrar espacio de almacenamiento al trabajar con conjuntos de datos grandes a los que no se suele consultar con frecuencia, como datos puramente históricos.

Considerando lo señalado anteriormente, MOLAP es una solución adecuada para soluciones con unos volúmenes de información y número de dimensiones más modestos con lo que llegan a ser menos eficientes conforme se incrementa la cantidad de datos, mientras que ROLAP es una arquitectura flexible que crece para dar soporte a amplios requerimientos. Por tanto, el almacén de datos se implementará usando el modo de almacenamiento relacional, que resulta el más idóneo de acuerdo a los volúmenes de información histórica a tratar.

1.5.4 Herramientas de consulta y análisis

Las herramientas de consulta y análisis permiten a los usuarios realizar la exploración de los datos del Almacén de Datos. Con estas herramientas mediante una interfaz gráfica los usuarios generan consultas que son enviadas al almacén, donde se realiza la extracción de información y se devuelven los resultados obtenidos. Luego, estos resultados son expuestos ante los usuarios en formatos intuitivos para su posterior análisis.

Una de las principales ventajas de utilizar estas herramientas, es que los usuarios no necesitan conocer cuáles son las características y funcionalidades de las estructuras de datos utilizadas, ni por saber emplear el lenguaje SQL para realizar las consultas, solo se deben enfocar en el análisis de la información.



1.6 Ambiente de desarrollo

La solución que se propone se encuentra enmarcada dentro del proyecto Identidad Cuba, como ya se explicó se tiene como cliente principal al MININT y en sus inicios fue definido el ambiente de desarrollo que regiría, quedando especificados el Sistema Gestor de Base de Datos a utilizar, así como las herramientas de consulta y desarrollo de manera general. La metodología que se definió fue para el desarrollo de las aplicaciones, pero no fue definida una específica para la creación de un almacén de datos.

1.6.1 Justificación de la metodología

En los últimos años un gran número de metodologías han llegado a estar disponibles para guiar la creación de proyectos de Almacenamiento de Datos. Todas se basan en un conjunto común de tareas: análisis de requerimientos del negocio, diseño de datos, diseño de la arquitectura, implementación y despliegue (12) (15). Sin embargo, con tantas metodologías para elegir, la preocupación radica en cuál emplear para el desarrollo de un proyecto dado.

Para este trabajo se revisaron y compararon las principales metodologías de Almacenamiento de Datos basándose en un grupo de atributos comunes, las cuales se muestran a continuación:

Kimball: su arquitectura *bottom-up*² (26) restringe el alcance a las necesidades de un área o departamento de la organización. Se recomienda para pequeños proyectos, donde además se quiera asegurar un alto acceso a los datos para los usuarios finales, y el rápido desarrollo de la solución. Su enfoque de modelado dimensional facilita su comprensión a los usuarios.

Inmon: su arquitectura *top-down*³ (26) permite el alcance global de las necesidades de la organización. Es más apropiada para sistemas complejos donde no se requiera asegurar un alto acceso a los datos para los usuarios finales, pero sí la perdurabilidad y consistencia del sistema aunque cambien los procesos de negocio en la organización.

HEFESTO: Permite la construcción de un Almacén de Datos de forma sencilla, ordenada e intuitiva. Se destaca el éxito de su uso en proyectos de diversa naturaleza en distintos países (10). Su enfoque de modelado dimensional le hace más fácil de comprender a los usuarios finales. Se basa en los requerimientos del cliente lo que permite construir un almacén consistente ante los cambios en el negocio.

² *bottom-up*: Arquitectura ascendente que comienza por los *Datamarts*, los cuales después se integran en el *Datawarehouse*.

³ *top-down*: Arquitectura que comienza por el Almacén de Datos y luego se descompone en *Datamarts* en dependencia de las necesidades.



Teniendo en cuenta que se cubrirá la organización de manera global, la amplitud de los modelos de datos relacionales disponibles, la perspectiva de crecimiento en el orden de los Gigas que experimentan mensualmente las fuentes de datos en su conjunto y que se debe garantizar que el contenido sea entendible y navegable, para que los usuarios finales puedan definir salidas no previstas sin depender de los diseñadores se escoge la metodología HEFESTO., la cual propone cuatro fases para la construcción del Almacén de Datos (ver [anexo 3](#)).

Esta metodología está fundamentada en una amplia investigación, comparación de metodologías existentes y experiencias propias en procesos de confección de Almacenes de Datos (10). Los objetivos, artefactos y resultados en cada fase son sencillos de comprender para quienes nunca han implementado un Almacén de Datos, pues utiliza modelos conceptuales y lógicos que son sencillos de interpretar. No cuenta con un número excesivo de artefactos. Con la ayuda de esta metodología es posible obtener en poco tiempo una visión integral del problema suficientemente sólida para desarrollar la implementación.

1.6.2 Herramienta de diseño y modelado de Almacenes de Datos: Embarcadero ER/Studio 8.0

Esta herramienta es una solución intuitiva y visual de modelado de datos para el diseño y mantenimiento de bases de datos transaccionales, de soporte a la toma de decisiones y para la Web (27). ER/Studio es la única herramienta de modelado que soporta todos los sistemas de gestión de las más importantes bases de datos, entregando a los usuarios la flexibilidad de trabajar con cualquiera de ellas (28).

Con el propósito de apoyar el diseño y el soporte de la integración de un Almacén de Datos, se destacan las siguientes funcionalidades de la herramienta: realiza el modelado dimensional de los modelos lógico y físico, permite la integración de metadatos y el modelo, brinda documentación para trazabilidad de los datos, permite exportar el modelo de datos y la salida de esquemas XML desde modelos físicos y lógicos (27).

Esta herramienta es usada por noventa de las empresas de la lista “*Fortune 100*”⁴ y por una comunidad activa de tres millones de usuarios en todo el mundo que contribuye con mejoras y amplia información para su uso (29).

1.6.3 Herramienta de Almacenamiento de Datos: Oracle Database

Básicamente, Oracle es un Sistema de Gestión de Base de Datos Relacional (o RDBMS por sus siglas en inglés), desarrollado por Oracle Corporation. Se considera a Oracle como uno de los sistemas de bases de

⁴ Fortune 100: Lista de las 100 compañías líderes de la economía norteamericana



datos más completos, destacando su: soporte de transacciones, estabilidad, escalabilidad y soporte multiplataforma. (30)

Oracle Database es el líder del mercado de Almacenes de Datos. Su versión Oracle Database 11gR2 es una plataforma completa de Base de Datos para Almacén de datos e inteligencia empresarial que combina: rendimiento, escalabilidad, calidad de datos e integración. Oracle Database 11g proporciona una plataforma exclusivamente integrada para el análisis. Al incorporar OLAP y funciones estadísticas directamente en la Base de Datos, Oracle ofrece toda la funcionalidad de un motor analítico independiente, con la confiabilidad, seguridad y escalabilidad empresarial de Oracle Database. Puesto que la integración de datos es un requerimiento central de cualquier almacén de datos, Oracle Database 11g incluye Oracle Warehouse Builder una herramienta líder para el desarrollo de ETL (en el epígrafe 1.6.5 se detalla más sobre las características de la herramienta.), que utiliza las capacidades de acceso de datos heterogéneos y transformación de datos escalable de Oracle (31). La versión 11gR2 de Oracle Database, que es la que se utiliza actualmente en el Sistema Único de Identificación Nacional, crea las bases para “brindar información a los usuarios incluso con una mayor calidad de servicio y de manera oportuna.” (32)

1.6.4 Herramienta de reportes: Oracle Business Intelligence Enterprise Edition

Oracle Business Intelligence Enterprise Edition (OBIEE) es una plataforma de negocios completa con una amplia oferta en capacidades de análisis y reporte, proveyendo a las empresas de análisis profundo y métricas de desempeño, capacitándolas así para la rápida toma de decisiones.(33)

El OBIEE integra componentes que conforman un juego de herramientas para incluir una arquitectura orientada a servicios, servicio de acceso a datos, una infraestructura de cálculo y análisis, servicios de administración de metadatos, un modelo de seguridad, preferencias de usuario y herramientas administrativas (34).

La herramienta brinda la capacidad de solicitar reportes, gráficos, tablas comparativas e interactuar con el cuadro de mando digital de manera intuitiva; tampoco es necesario preocuparse por la complejidad de las estructuradas de datos porque se presenta una interfaz lógica donde dicha información es comprensible. Esto les permite generar salidas informativas que no hayan sido previstas sin la necesidad de consultar a un diseñador.



1.6.5 Herramienta para la realización de los ETL: Oracle Warehouse Builder

Oracle Warehouse Builder (OWB) es una herramienta ETL desarrollada por Oracle, que ofrece un entorno gráfico para construir, administrar y mantener los procesos de integración de datos en los sistemas de Inteligencia de Negocios.

OWB permite diseñar modelos relacionales y dimensionales, brinda soporte para OLAP, cubos organizados MVs (Materialized Views) así como para el modelado y carga de SCD (Slowly Changes Dimensions) de tipo 2 y 3, permite integración con herramientas como OBIEE; admite integración de datos y técnicas ETL.

Esta herramienta también forma parte de la suite Oracle 11gR2 (35) en la cual están soportadas las bases de datos del SUIN y que es de donde se extraerán los datos, además, al ser la Base de Datos origen Oracle incrementa el desempeño como herramienta de ETL.

1.6.6 Herramienta para la gestión de Base de Datos: PL/SQL Developer

PL/SQL Developer 7.1.4 es un ambiente de desarrollo integrado que está específicamente orientado para el desarrollo, prueba, depuración de errores y optimización de PL/SQL de Oracle. La herramienta contiene ayuda sensitiva al contexto, descripciones de bases de datos de objetos, sintaxis resaltada, edición y búsqueda de datos y un buscador. (36)

El PL/SQL Developer 7.1.4 se enfoca en la facilidad de uso, calidad del código y productividad, ventajas claves durante el desarrollo de aplicaciones en Oracle.

1.7 Conclusiones parciales

En el estudio realizado dentro del campo de la Inteligencia de Negocios se decide utilizar los principios de la Inteligencia Institucional para el desarrollo del Almacén de Datos, debido a la naturaleza Institucional del cliente. Como esquema de diseño se usará el esquema estrella pues optimiza los tiempos de respuesta ante las consultas de los usuarios. Se escoge como modo de almacenamiento de la Base de Datos ROLAP, pues constituye una arquitectura flexible que crece para dar soporte a amplios requerimientos y resulta la más idónea de acuerdo a los grandes volúmenes de información histórica a tratar.

En esta primera fase se ha hecho un análisis profundo de las metodologías, que ayudó a determinar y fundamentar según las necesidades de la solución la utilización de HEFESTO como metodología para la construcción del Almacén de Datos. Según el ambiente de desarrollo definido en el marco del proyecto en



Capítulo 1: Fundamentos teóricos

que se encuentra la solución propuesta, se decide utilizar las herramientas: Embarcadero ER/Studio 8.0, como herramienta de diseño y modelado de Almacenes de Datos; PL/SQL Developer, como herramienta para la gestión de Base de Datos y la Suite de Oracle 11gR2 como entorno de desarrollo que incluye: Oracle Database como herramienta de Almacenamiento de Datos, Oracle Warehouse Builder como herramienta para realizar los ETL y Oracle Business Intelligence Enterprise Edition como herramienta para diseñar e implementar los reportes candidatos.



Capítulo 2 Análisis y diseño del Almacén de Datos

2.1 Introducción

En el presente capítulo se muestran los resultados y pasos a seguir para el análisis y el diseño del Almacén de Datos según propone la metodología HEFESTO. Se elabora un modelo conceptual a partir de los requisitos de información obtenidos de las entrevistas con los clientes, se amplía el modelo conceptual luego de establecer las correspondencias con los OLTP, se diseña el modelo lógico del almacén, para finalmente construir el modelo de arquitectura.

2.2 Análisis de los requerimientos

Primeramente, se identificaron los requisitos de información del MININT a través de preguntas de negocio⁵ que explicitan los objetivos de la organización. Se analizaron estas preguntas con el objetivo de identificar cuáles son los indicadores y perspectivas a tomar en cuenta para la construcción del almacén. Por último, se confecciona un modelo conceptual en donde se puede visualizar el resultado obtenido de esta primera fase, de manera que se refleje claramente lo que se espera del Almacén de Datos.

2.2.1 Identificar preguntas

Para identificar las preguntas, resulta necesario identificar las necesidades de información, que serán esenciales para trazar las metas y estrategias de la organización en aras de ganar una mayor eficiencia en la toma de decisiones. Esto se logra mediante el uso de disímiles técnicas: como entrevistas, cuestionarios y observaciones.

Se realizó una entrevista a los funcionarios del MININT en busca de sus necesidades de información, donde se les solicitó que escogieran los temas más relevantes a tratar que estuvieran soportados por algún OLTP; como resultado expresaron la necesidad de obtener la información referente a las personas, los trámites y los cambios de direcciones, y de informatizar el CMI integral del MININT, con el objetivo de lograr un mayor control.

Luego se procedió a identificar qué era lo que realmente interesaba acerca de estos temas: cuáles eran los indicadores o medidas que representan lo que se desea analizar concretamente, y las perspectivas desde las cuales deseaban examinarlos.

⁵ Las preguntas del negocio se corresponden con las salidas informativas que debe poseer la aplicación.



Capítulo 2: Análisis y diseño del Almacén de Datos

Como consecuencia del análisis realizado se determinó que las preguntas del negocio eran las siguientes:

- Se desea conocer la “Cantidad de personas que realizaron cambios de dirección permanentes entre provincias en un período determinado”.
- Se desea conocer la “Cantidad de personas que realizaron cambios de dirección entre municipios de una misma provincia en un período determinado”.
- Se desea conocer la “Cantidad de personas que realizaron cambios de dirección en un mismo municipio en un período determinado”.
- Se desea conocer el “Total de personas que realizaron cambios de dirección según el territorio de procedencia⁶ en período un determinado”.
- Se desea conocer la “Cantidad de personas de cada género⁷ que pertenecen a un territorio⁸ en un período determinado”.
- Se desea conocer la “Cantidad de personas que presentan un tipo de color de piel⁹ de acuerdo al territorio que pertenecen”.
- Se desea conocer la “Cantidad de personas de cada género por cada color de piel que pertenecen a un territorio determinado”.
- Se desea conocer la “Cantidad de personas según el rango de edad ¹⁰que pertenecen a un territorio determinado”.
- Se desea conocer la “Cantidad de defunciones por territorio en un período determinado”.
- Se desea conocer la “Cantidad de personas según el género por cada rango de edad que pertenecen a un territorio determinado”.
- Se desea conocer la “Cantidad de personas mayor o menor de edad de acuerdo al color de piel que pertenecen a un territorio determinado”.
- Se desea conocer la “Cantidad de personas atendidas por trámite en horario extendido por territorio en un período determinado”.
- Se desea conocer la “Cantidad de personas atendidas por información en horario extendido por territorio en un período determinado”.

⁶ Por ejemplo, de La Habana hacia otra provincia, hacia otro municipio dentro de la misma provincia o dentro de un mismo municipio en la provincia.

⁷ Género: Masculino o Femenino

⁸ Territorio: En este caso se refiere a provincia o municipio

⁹ Color de Piel: Blanca, Mestiza o Negra

¹⁰ Rangos de Edades: menores de 16,16-25, 26-35, 36-45, 46-55, 56-65, mayores de 65.



Capítulo 2: Análisis y diseño del Almacén de Datos

- Se desea conocer el “Total de personas atendidas en horario extendido por territorio en un período determinado”.
- Se desea conocer el “Total de personas atendidas en horario extendido por día de la semana por territorio en un período determinado”.
- Se desea conocer la “Cantidad de trámites según el tipo de trámite¹¹ por territorio¹² en un período determinado”.
- Se desea conocer el “Total de trámites según el tipo de trámite en un período determinado”.
- Se desea conocer el “Total de trámites por territorio en un período determinado”.
- Se desea listar las personas de acuerdo a su: “provincia de residencia, municipio de residencia, número de identidad, primer nombre, segundo nombre, primer apellido, segundo apellido, edad, género, color de piel, nombre padre, nombre madre, provincia de procedencia, municipio de procedencia, dirección de residencia y fecha de defunción”.

2.2.2 Identificar indicadores y perspectivas

Una vez identificadas las preguntas de negocio, se procede a su análisis y descomposición para definir los indicadores que serán utilizados y las perspectivas o dimensiones desde las cuales se analizarán.

Ejemplo de análisis:

Total de personas que realizaron cambios de dirección según la provincia de procedencia¹³ en un período determinado.

Cantidad de trámites según el tipo de trámite por territorio en un período determinado.

● Perspectivas

● Indicadores

A continuación se muestran los indicadores y las perspectivas identificados:

Indicadores	Perspectivas
Cantidad de personas	Persona
Total de personas	Cambio de dirección
Porcentaje de personas	Fecha

¹¹ Nacimientos, Arribos, Pérdida, Actualización (Subsanación de error, Deterioro, Cambio legal), Solicitud de Documentos (Salida FAR, Salida de Prisiones)

¹² Para los trámites el territorio es: provincia, municipio y la oficina donde se realiza el trámite.

¹³ Por ejemplo, de La Habana hacia otra provincia, hacia otro municipio dentro de la misma provincia o dentro de un mismo municipio en la provincia.



Capítulo 2: Análisis y diseño del Almacén de Datos

Cantidad de trámites	Trámite
Total de trámites	Oficina

Tabla 2.1 Indicadores y perspectivas

2.2.3 Modelo Conceptual

En esta etapa se construye el modelo conceptual a partir de los indicadores y perspectivas obtenidas, lo que permite observar con claridad el alcance del almacén. Además, permite interpretar el negocio sin poseer demasiados conocimientos previos, comprender cuáles son los resultados que se obtendrán, cuáles son las variables que se utilizarán para analizarlos y cuál es la relación que existe entre ellos.

En la siguiente figura se muestra el modelo conceptual resultante del análisis de los indicadores y perspectivas, donde a la izquierda se colocan las perspectivas seleccionadas, que serán unidas a un óvalo central que representa y lleva el nombre de la relación que existe entre ellas. La relación, constituye el proceso o área de estudio elegida. De dicha relación se desprenden los indicadores que se ubican en la derecha:

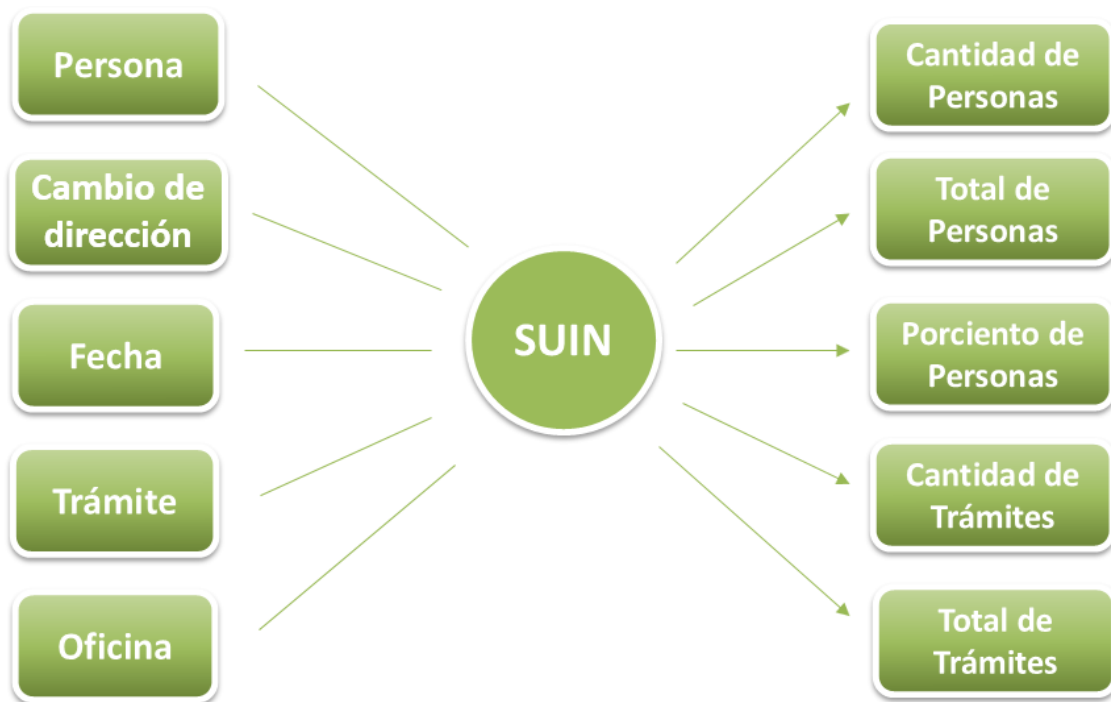


Figura 2.1 Modelo Conceptual



2.3 Análisis de los OLTP

En este paso se analizan las fuentes OLTP (Esquemas Identidad¹⁴ y ADM¹⁵ de la Base de Datos del SUIN) (ver [anexos del 4 al 7](#)), para determinar cómo calcular los indicadores y para establecer las respectivas correspondencias entre el modelo conceptual creado en el paso anterior y las fuentes de datos. Luego, se definen qué campos se incluyen en cada perspectiva, para finalmente ampliar el modelo conceptual.

2.3.1 Conformar indicadores

A continuación se explicita cómo se calcularán los indicadores, detallando cada uno de ellos:

- “Cantidad de Personas”: representa la cantidad de personas a calcular según un parámetro¹⁶ en particular.
- Función de agregación: SUM.
- “Total de Personas”: representa el total de personas a calcular según un parámetro en particular.
- Función de agregación: SUM.
- “Porcentaje de Personas”: representa el porcentaje de las personas a calcular según un parámetro en particular.
- Función de Porcentaje: $(\text{cantidad de personas}) / (\text{total de personas}) * 100$
- “Cantidad de trámites”: representa la cantidad de trámites a calcular según un trámite en particular.
- Función de agregación: SUM.
- “Total de trámites”: representa el total de trámites a calcular según un trámite en particular.
- Función de agregación: SUM.

Además de los indicadores identificados a partir de las preguntas del negocio, se definieron los indicadores automatizables del CMI del MININT, los cuales se muestran a continuación:

- Nivel de fiabilidad del servicio:

$$\left(1 - \frac{\text{cantidad de irregularidades detectadas}}{\text{total de trámites realizados}}\right) * 100$$

- Nivel de estabilidad del tiempo de servicio:

¹⁴ Esquema Identidad: esquema que contiene toda la información referente a la información de la población y los trámites que se realizan en las oficinas de identificación y registro.

¹⁵ Esquema ADM: esquema que contiene toda la información referente a la administración de identidades del SUIN

¹⁶ Parámetros: Color de Piel, Color Cabello, Color Ojos, Condición Migratoria, Cambio de dirección, Género, Territorio,



Capítulo 2: Análisis y diseño del Almacén de Datos

$$\left(\frac{\text{cantidad de servicios que son atendidos en la unidad en el tiempo establecido}}{\text{total de solicitudes de servicio atendidas que no interrumpen el trámite}} \right) * 100$$

- Nivel de estabilidad en el tiempo de respuesta a las irregularidades:

$$\left(\frac{\text{cantidad de irregularidades a las que se da respuesta en el tiempo establecido}}{\text{total de irregularidades atendidas}} \right) * 100$$

- Coeficiente de fiabilidad del servicio:

$$\left(\frac{\text{cantidad de trámites con irregularidades detectadas}}{\text{total de solicitudes de servicio}} \right) * 100$$

- Tiempo promedio de servicio en la unidad CIRP=(tiempo promedio de los trámites realizados)

2.3.2 Establecer correspondencias

Posteriormente, se examinan los OLTP disponibles, que contienen la información necesaria para poder identificar las correspondencias entre las perspectivas del modelo conceptual y las fuentes de datos.

En las siguientes figuras se exponen las correspondencias identificadas entre las perspectivas del modelo conceptual y las tablas del ambiente operacional.

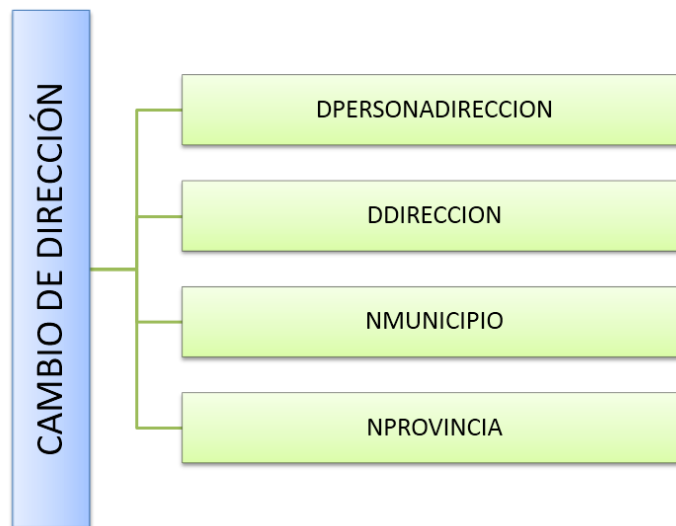


Figura 2.2 Correspondencia entre la perspectiva Cambio de Dirección con las tablas del Esquema Identidad

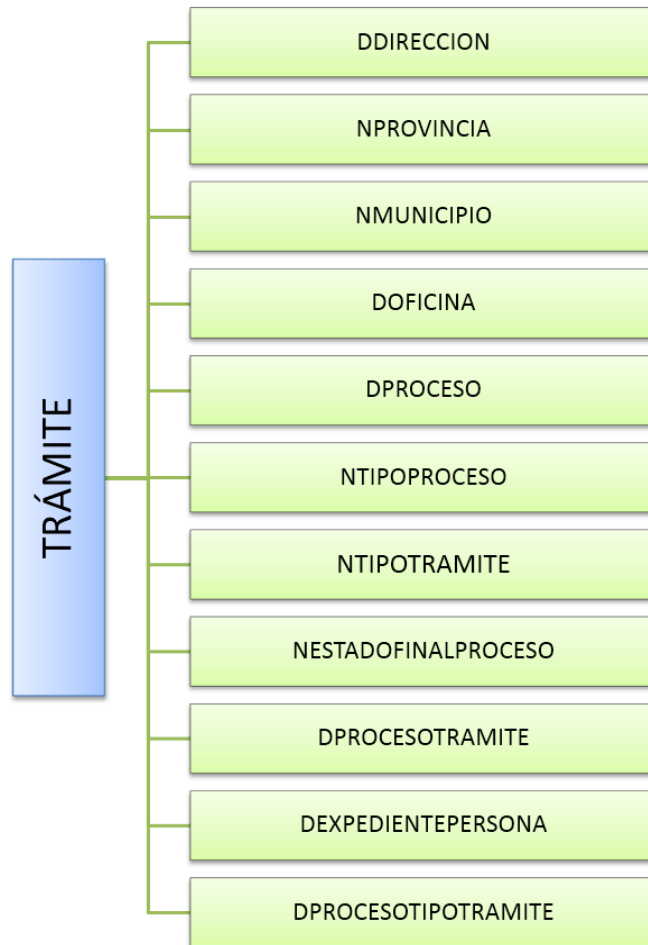


Figura 2.3 Correspondencia entre la perspectiva Trámite con las tablas del Esquema Identidad

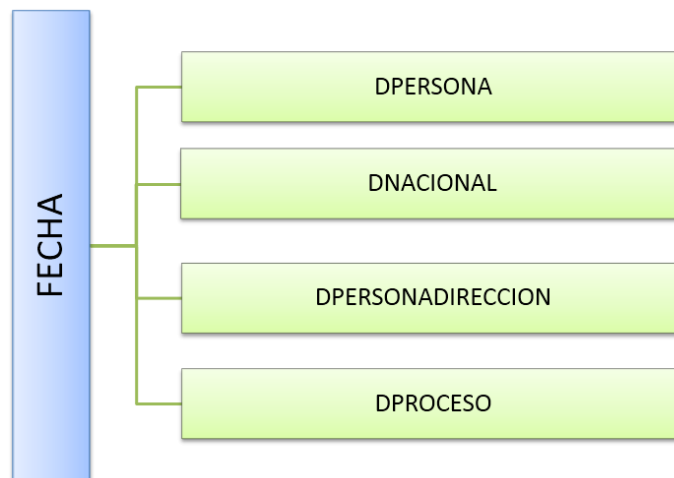


Figura 2.4 Correspondencia entre la perspectiva Fecha con las tablas del Esquema Identidad



Capítulo 2: Análisis y diseño del Almacén de Datos

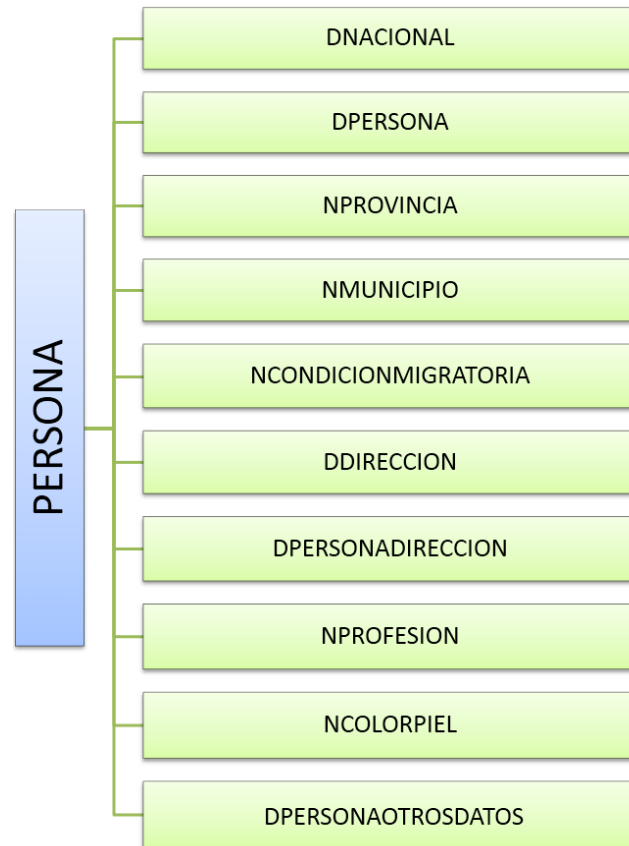


Figura 2.5 Correspondencia entre la perspectiva Persona con las tablas del Esquema Identidad

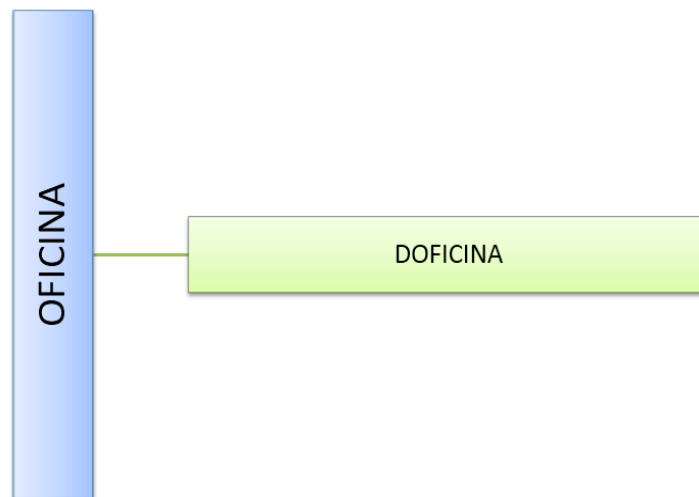


Figura 2.6 Correspondencia entre la perspectiva Oficina con las tablas del Esquema ADM



Capítulo 2: Análisis y diseño del Almacén de Datos

2.3.3 Nivel de granularidad

Una vez establecidas las correspondencias con los OLTP, se analizaron los campos de cada tabla a la que se hacía referencia, mediante dos métodos: primero se examinó el modelo de la Base de Datos del SUIN para intuir los significados de cada campo, y luego se consultó con especialistas para aclarar algunas dudas del negocio y aspectos pendientes que no se pudieron inferir, además de corroborar los campos identificados. Finalmente, se seleccionaron los campos que integran cada perspectiva, pues a través de estos se examinarán y filtrarán los indicadores.

En las siguientes tablas se describen los campos que componen cada perspectiva, especificándose además su origen OLTP.

Perspectiva Persona		
Campo	Origen OLTP	Descripción
IdPersona	dPersona	Es la llave primaria de la tabla dPersona e identifica a una persona en particular
CondicionMigratoria	nCondicionMigratoria	Indica la condición migratoria que tiene la persona
Provincia	nProvincia	Representa la provincia de residencia de la persona
Municipio	nMunicipio	Representa el municipio de residencia de la persona
RangodeEdad	dNacional	Indica en que rango de edad se encuentra la persona
Genero	dNacional	Indica el género de la persona
ColorPiel	nColorPiel	Indica el color de piel de la persona
Numeroidentidad	dNacional	Indica el número de identidad de la persona
PrimerNombre	dNacional	Indica el primer nombre de la persona
SegundoNombre	dNacional	Indica el segundo nombre de la persona, puede ser un valor nulo
PrimerApellido	dNacional	Indica el primer apellido de la persona



Capítulo 2: Análisis y diseño del Almacén de Datos

SegundoApellido	dNacional	Indica el segundo apellido de la persona
NombrePadre	dPersonaOtrosDatos	Indica el nombre del padre de la persona
NombreMadre	dPersonaOtrosDatos	Indica el nombre de la madre de la persona
Direccion	dDireccion	Indica la tira de dirección de la persona

Tabla 2.1 Descripción de los campos de la perspectiva Persona

Perspectiva Cambio de Dirección		
Campo	Origen OLTP	Descripción
ProvinciaProcedencia	nProvincia	Indica la provincia donde residía la persona antes de realizar el cambio de dirección
MunicipioProcedencia	nMunicipio	Indica el municipio donde residía la persona antes de realizar el cambio de dirección
ProvinciaDestino	nProvincia	Indica la provincia donde reside actualmente la persona
MunicipioDestino	nMunicipio	Indica el municipio donde reside actualmente la persona
DireccionProcedencia	dDireccion	Indica la tira de dirección de la persona antes de realizar el cambio de dirección
DireccionDestino	dDireccion	Indica la tira de dirección actual de la persona

Tabla 2.2 Descripción de los campos de la perspectiva Cambio de Dirección



Capítulo 2: Análisis y diseño del Almacén de Datos

Perspectiva Trámite		
Campo	Origen OLTP	Descripción
NombreProceso	nTipoProceso	Indica el proceso que da inicio al trámite
TipoHorario	dProceso	Indica el tipo de horario en que se inicia el trámite
IdSerialProceso	dProceso	Es la clave primaria de la tabla dProceso e identifica a un proceso en particular
EstadoFinal	nEstadoFinalProceso	Indica el estado en que el trámite finalizó
NombreTramite	nTipoTramite	Indica el trámite que se realiza

Tabla 2.3 Descripción de los campos de la perspectiva Trámite

Perspectiva Fecha	
Campo	Descripción
Año	Indica número del año
Trimestre	Indica número del trimestre
Mes	Indica número del mes
Día	Indica número del día

Tabla 2.4 Descripción de los campos de la perspectiva Fecha

Perspectiva Oficina		
Campo	Origen OLTP	Descripción
idOficina	dOficina	Es la clave primaria de la tabla dOficina e identifica a una oficina en particular
Oficina	dOficina	Indica el nombre de la oficina
Provincia	nProvincia	Indica la provincia donde radica la oficina
Municipio	nMunicipio	Indica el municipio donde radica la oficina

Tabla 2.5 Descripción de los campos de la perspectiva Oficina



2.3.4 Modelo Conceptual ampliado

De acuerdo a los resultados obtenidos en los pasos anteriores, se amplía el modelo conceptual, colocando bajo cada perspectiva los campos seleccionados y bajo cada indicador su respectiva fórmula de cálculo, con este modelo que se muestra a continuación culmina el análisis del Almacén de Datos.

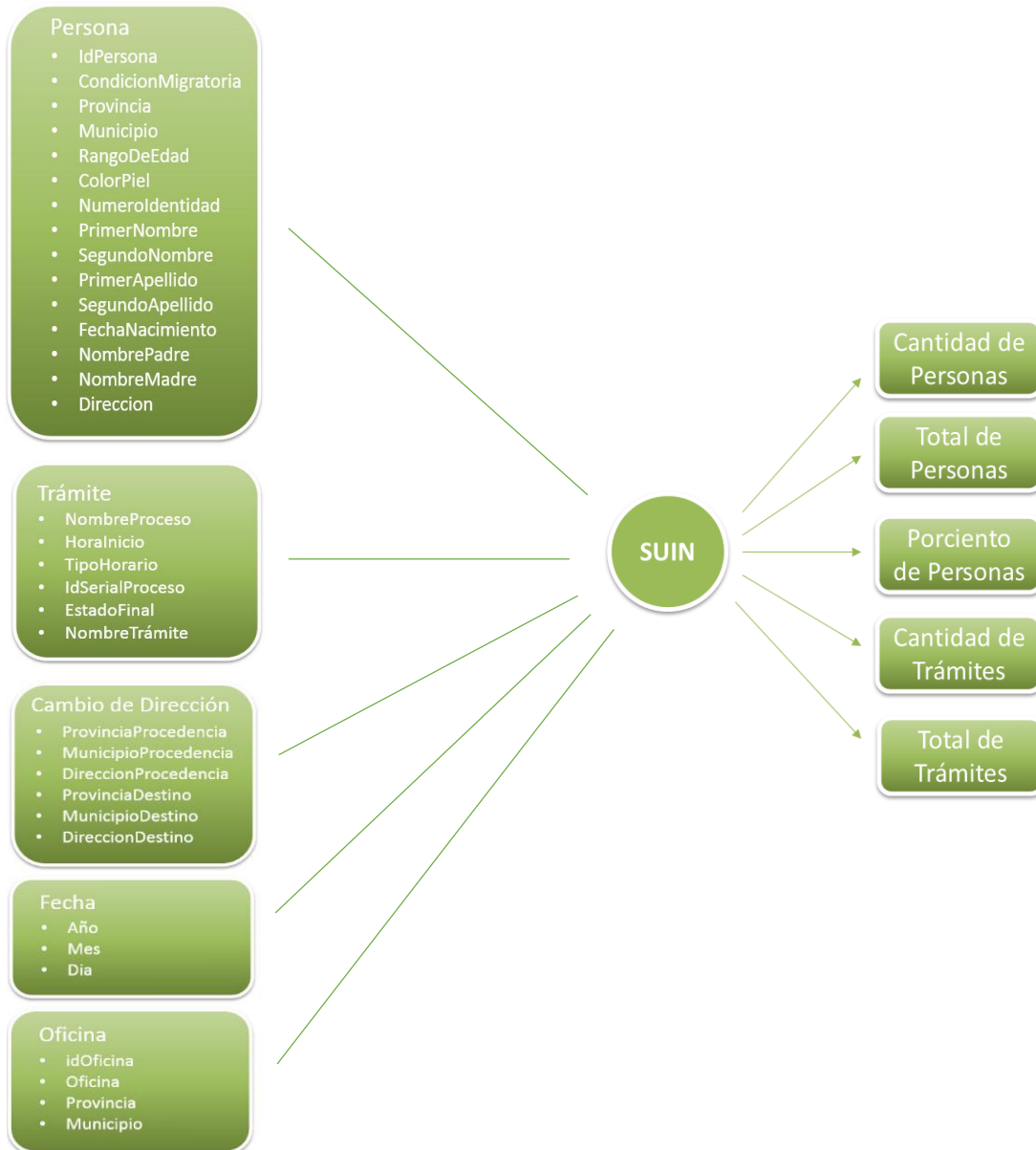


Figura 2.7 Modelo Conceptual Ampliado



2.4 Modelo Lógico de Almacén de Datos

En este paso se confecciona el modelo lógico del almacén, teniendo como base el modelo conceptual que ya ha sido creado. Para ello, primero se define el tipo de modelo a utilizar, luego se diseñan las tablas de dimensiones y de hechos, para finalmente realizar las uniones pertinentes entre estas tablas.

2.4.1 Tipo de Modelo Lógico del Almacén de Datos

Luego del análisis realizado en el capítulo 1 de los tipos de esquemas a utilizar para contener la estructura del depósito de datos, se escoge el esquema estrella para diseñar el modelo lógico, pues resulta ser el que mejor se adapta a los requerimientos y necesidades del cliente.

2.4.2 Tablas de dimensiones

Una vez escogido el esquema, se diseñan las tablas de dimensiones, donde cada perspectiva definida en el modelo conceptual constituye una tabla de dimensión.

A continuación se muestran las tablas de dimensiones diseñadas para el Almacén de Datos:

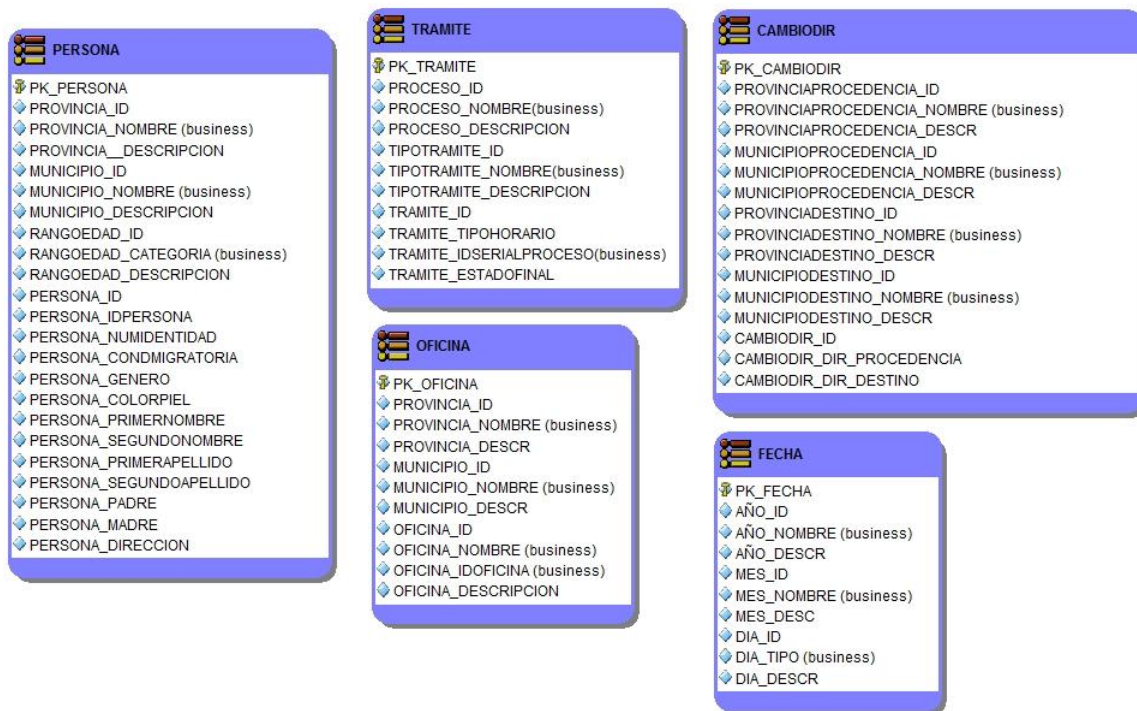


Figura 2.8 Dimensiones



Capítulo 2: Análisis y diseño del Almacén de Datos

2.4.3 Tablas de hechos

Para diseñar las tablas de hecho se siguieron algunas pautas:

- Se asignó un nombre a cada tabla de hechos que represente la información a analizar o el área de negocio enfocada.
- Se definió la clave primaria, que se compone de la combinación de las claves primarias de cada tabla de dimensión relacionada.
- Se crearon las medidas para el cálculo de los indicadores.
- Se crearon variables binarias para la agregación de las medidas.

A continuación se muestran las tablas de hechos diseñadas para el Almacén de Datos:

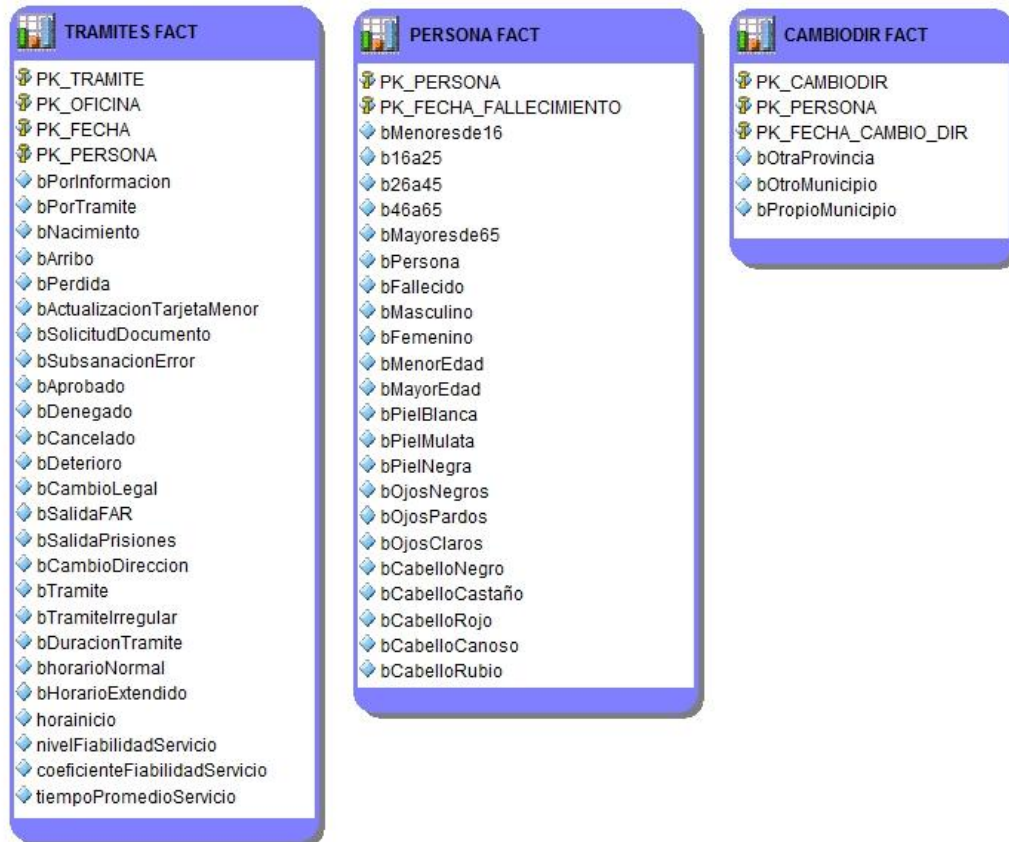


Figura 2.9 Hechos



Capítulo 2: Análisis y diseño del Almacén de Datos

2.4.4 Uniones

Luego de diseñar las tablas de hechos y de dimensiones se realizan las uniones¹⁷ correspondientes, quedando construido el modelo lógico como se muestra en la siguiente figura:

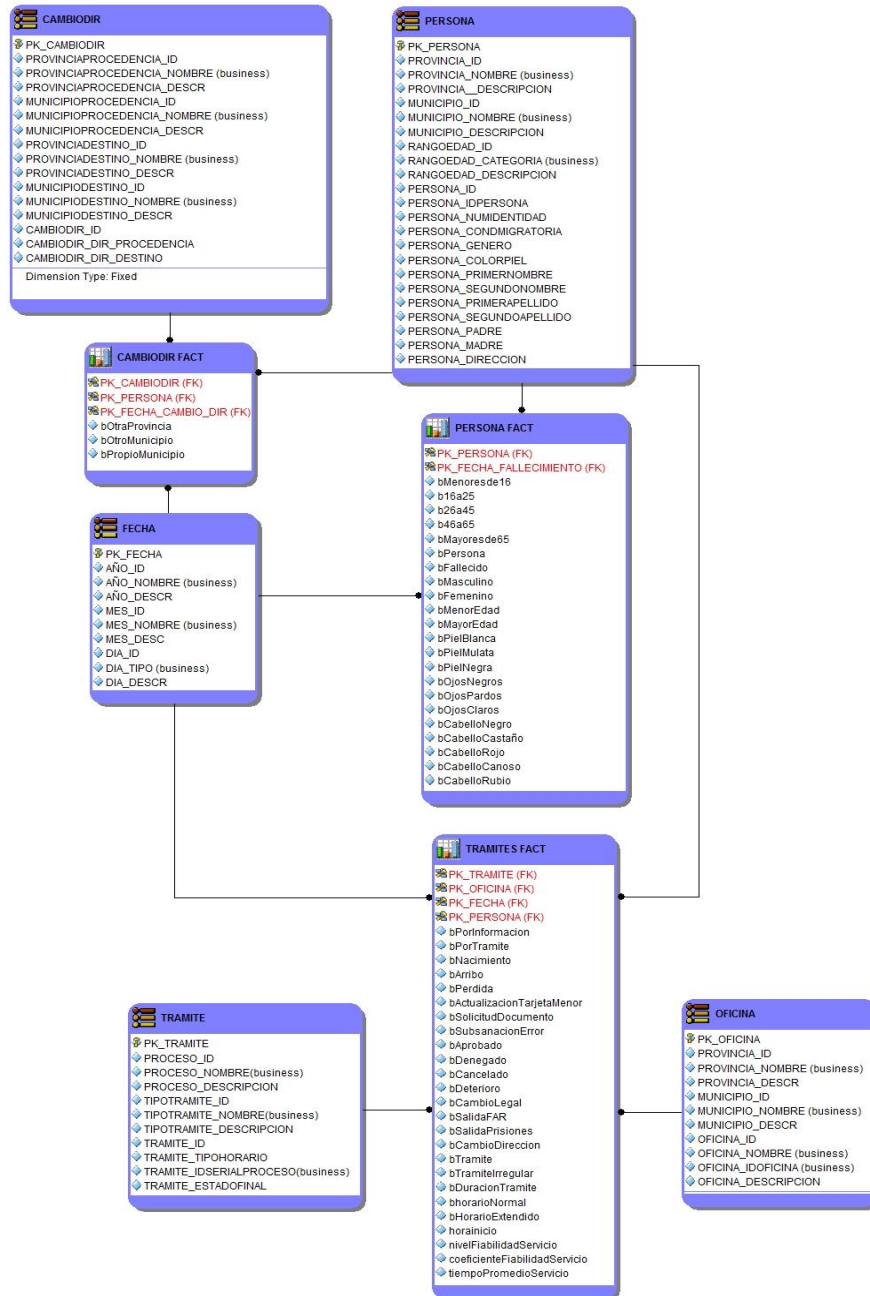


Figura 2.10 Modelo Lógico

¹⁷ Uniones: Según la metodología las relaciones que se establecen entre las tablas de dimensiones y las tablas de hechos.



2.5 Modelo de Arquitectura del Almacén de Datos

Teniendo en cuenta todo lo descrito anteriormente se define la arquitectura de la Solución de Inteligencia Institucional para el SUIN, en la cual se ilustran las etapas que definen el ambiente del Almacén de Datos.



Figura 2.11 Arquitectura de Solución de Inteligencia Institucional para el SUIN

Para el modelo de arquitectura a seguir, primeramente se definen las fuentes de datos, en este caso se escogieron los Esquemas Identidad (ver [Anexos del 4 al 6](#)) y ADM (ver [Anexo 7](#)) de la Base de Datos del SUIN, luego se extraen los datos hacia el Área de Almacenamiento Temporal donde se le aplican un conjunto de transformaciones y finalmente se cargan los datos en el Almacén donde se consultan para finalmente confeccionar los reportes que serán mostrados a los usuarios.

2.6 Conclusiones parciales

En este capítulo, teniendo en cuenta los procesos de las fases y pasos a seguir según define la metodología seleccionada y atendiendo a los requerimientos del cliente, se identificaron los indicadores resultantes de las interrogantes y sus respectivas perspectivas de análisis, mediante las cuales se construyó el modelo conceptual del almacén. Se analizaron los OLTP para establecer las correspondencias y se seleccionaron los campos de cada perspectiva con el objetivo de ampliar el modelo conceptual, lo que ayudó a visualizar el alcance del Almacén de Datos. Se realizó el diseño del modelo lógico, lo cual sienta las bases para la implementación de la solución. Se confeccionó el modelo de arquitectura, que detalla el proceso de flujo de información con el cual se define el flujo de trabajo a realizar durante el desarrollo de la solución que se propone.



Capítulo 3 Implementación y pruebas

3.1 Introducción

En el presente capítulo se realiza la implementación de la solución de Inteligencia Institucional para el SUIN. Se realiza el proceso de integración de datos mediante la técnica ETL. Se establecen los flujos de procesos para la carga y actualización del almacén. Se diseñan e implementan los reportes candidatos que responden a los requisitos de información del cliente. Por último, se realizan las pruebas para verificar el correcto funcionamiento del software y que cuente con la calidad requerida.

3.2 Implementación del almacén

Una vez construido el modelo lógico, se procede a poblar el Almacén de Datos mediante el proceso de integración de datos, que agrupa una serie de técnicas y subprocesos que se encargan de llevar a cabo todas las tareas relacionadas con la extracción, manipulación, integración, depuración de datos, carga y actualización del almacén, donde la técnica utilizada será efectuar los procesos ETL y se definirán las reglas y políticas para su respectiva actualización.

Para los procesos ETL en correspondencia con el modelo de arquitectura definida, se crearon las Áreas de Almacenamiento Temporal, que constituyen una línea divisoria entre el sistema operacional y el depósito de datos y van a almacenar toda la información que posteriormente será cargada en el Almacén de Datos.

Este almacenamiento intermedio permite:

- Integrar las distintas fuentes de datos.
- Manipular los datos sin interrumpir los OLTP ni el Almacén de Datos.
- No depender de la disponibilidad de los OLTP, pues una vez cargados a este los datos ya no es necesario que los OLTP estén en línea para efectuar las transformaciones y la carga posterior.
- Almacenar y gestionar los metadatos que se generarán en los procesos ETL.

Una vez creado el almacenamiento intermedio se extrajeron los datos relevantes desde los OLTP, para luego efectuar las transformaciones correspondientes con el objetivo de elevar la calidad de los datos y eliminar inconsistencias; posteriormente fue realizada la carga de los datos que residen en el almacenamiento intermedio hacia el almacén. A continuación se describen los procesos ETL diseñados para el Almacén de Datos:



Proceso ETL de Persona

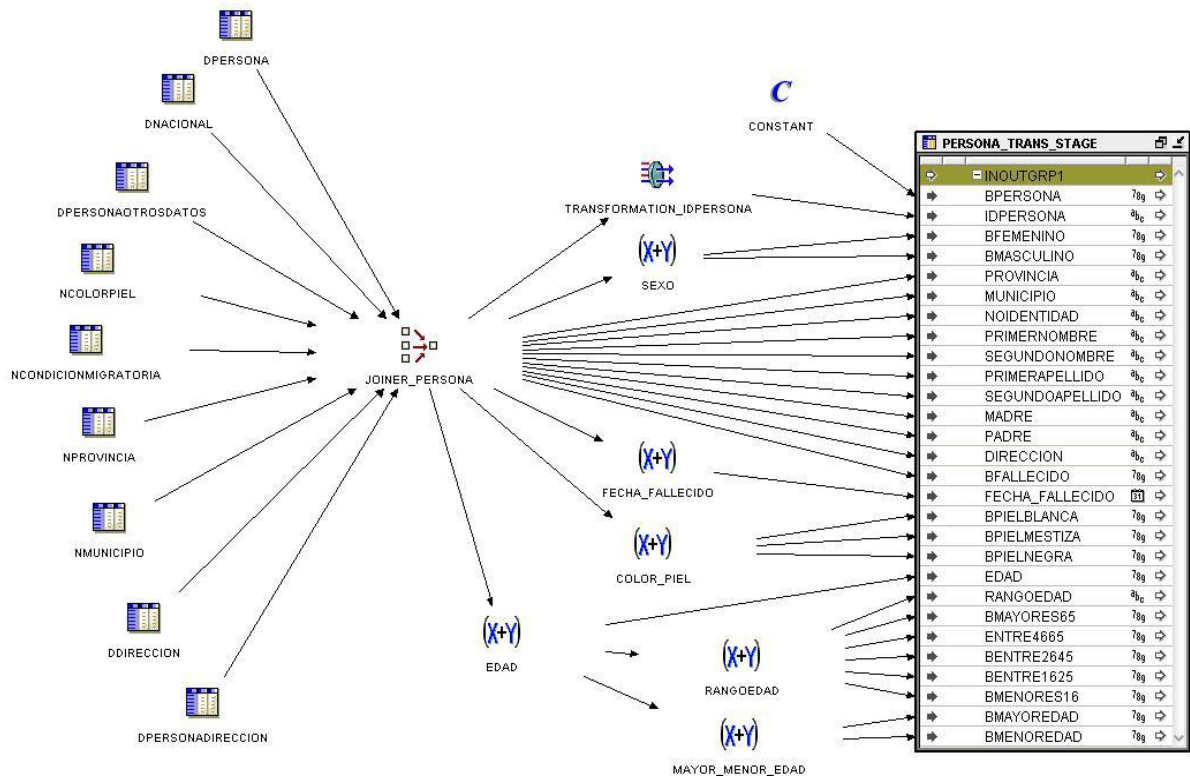


Figura 3.1 Correspondencia de la tabla temporal de persona

El ETL persona inicia seleccionando los atributos claves de las fuentes de datos origen correspondientes (tablas definidas cuando se establecieron las correspondencias entre los atributos de las perspectivas y los OLTP). Posteriormente se utiliza el componente *JOINER* para establecer las relaciones entre las tablas y determinar las condiciones de unión. Luego, se utilizan expresiones para generar nuevos datos como la edad, el rango de edad y si es mayor o menor de edad; para calcular los valores de las variables que integrarán el hecho, como el género y el color de piel. Además, se realiza limpieza de datos pues se reemplazan valores para datos faltantes como la fecha de fallecido, evitando que se carguen valores nulos en el almacén y no se pueda cargar el cubo *PERSONA_FACT*, y para datos inconsistentes como personas con fecha de fallecido mayor que la fecha actual. Se estandarizan formatos de representación acorde al nivel de granularidad definido, por ejemplo el truncamiento a la fecha de fallecido hasta el nivel de día, se usan convenciones de nombramiento para esclarecer la semántica de algunos atributos y se realizan conversiones de tipos de datos como al campo *id_persona*. Finalmente, se trasladan todos los datos hacia la tabla de almacenamiento intermedio *PERSONA_TRANS_STAGE*. Para posteriormente cargarlos en la



Capítulo 3: Implementación y pruebas

tabla de dimensión (ver [anexo 8](#)) y la tabla de hecho (ver [anexo 9](#)) correspondiente. La Figura 3.1 muestra el proceso de carga de los datos hacia la tabla de almacenamiento intermedio.

Proceso ETL de Cambio de Dirección

El ETL cambio de dirección inicia seleccionando los atributos claves de los OLTP correspondientes. Por la complejidad para obtener la tabla de almacenamiento temporal de cambio de dirección resulta necesario agrupar en el *JOINER_DESTINO* el grupo de tablas que permiten obtener los datos asociados a la dirección de residencia actual, para posteriormente enlazarlo al componente *JOINER* junto a las tablas correspondientes al grupo lógico de la dirección de procedencia, y mediante las condiciones de unión, poder seleccionar los cambios de dirección. Se utiliza el componente *AGREGATOR* para agrupar los datos según el orden de los campos especificados, en este caso *provincia_procedencia*, *provincia_destino*, *municipio_procedencia* y *municipio_destino*. Luego se utilizan expresiones para calcular los valores de las medidas que integrarán el cubo *CAMBIODIRECCION_FACT* como *otra_provincia*, *otro_municipio* y *mismo_municipio*. Se estandarizan formatos de representación acorde al nivel de granularidad definido, por ejemplo el truncamiento a la fecha del cambio de dirección hasta el nivel de día y se realizan conversiones de tipos de datos a campos como *id_persona*, *id_direccionproc*, *id_direcciondes*. Finalmente, se trasladan todos los datos hacia la tabla de almacenamiento intermedio *CAMBIODIRECCION_TRANS_STAGE* para posteriormente cargarlos en la tabla de dimensión (ver [anexo 10](#)) y la tabla de hecho (ver [anexo 11](#)) correspondiente. La figura 3.2 muestra el proceso de carga de los datos hacia la tabla de almacenamiento intermedio.

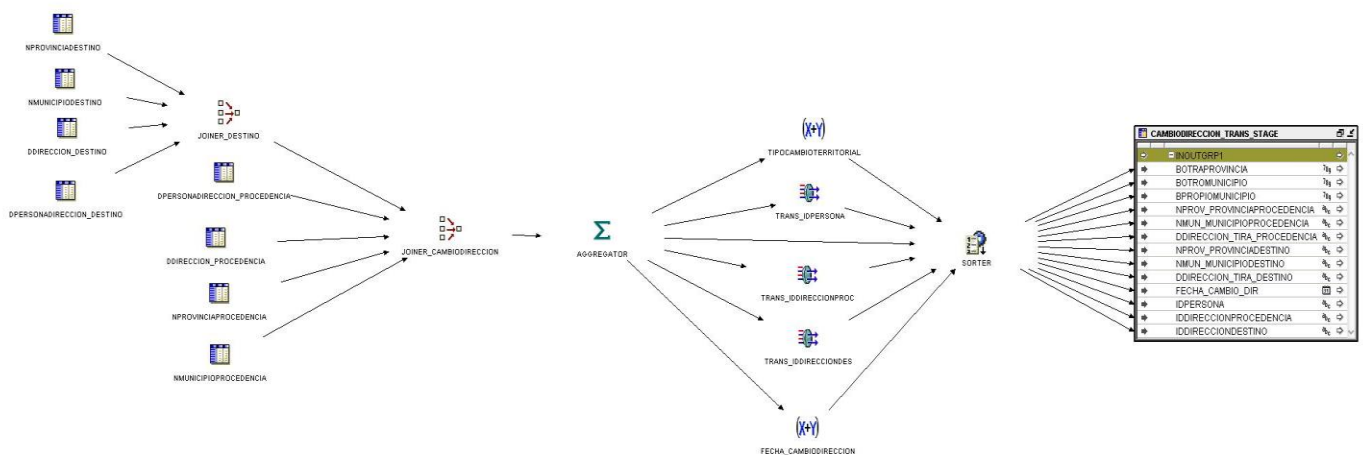


Figura 3.2 Correspondencia de la tabla temporal de cambio de dirección



Proceso ETL de Trámite

El ETL trámite inicia seleccionando los atributos claves de las fuentes de datos origen correspondientes (tablas definidas cuando se establecieron las correspondencias entre los atributos de las perspectivas y los OLTP). Posteriormente se utiliza el componente *JOINER* para establecer las relaciones entre las tablas y determinar las condiciones de unión. Luego, se utilizan expresiones para generar nuevos datos: como la fecha de inicio del trámite, el día de la semana en que se realiza y la duración del mismo, y para calcular los valores de las medidas que integrarán el cubo TRAMITE_FACT: como el tipo de horario en que se realiza el trámite, el estado final del trámite y el tipo de trámite. Finalmente, se trasladan todos los datos hacia la tabla de almacenamiento intermedio TRAMITE_TRANS_STAGE para posteriormente cargarlos en la tabla de dimensión (ver [anexo 12](#)) y la tabla de hecho (ver [anexo 13](#)) correspondiente. La figura 3.3 muestra todo el proceso ETL realizado.

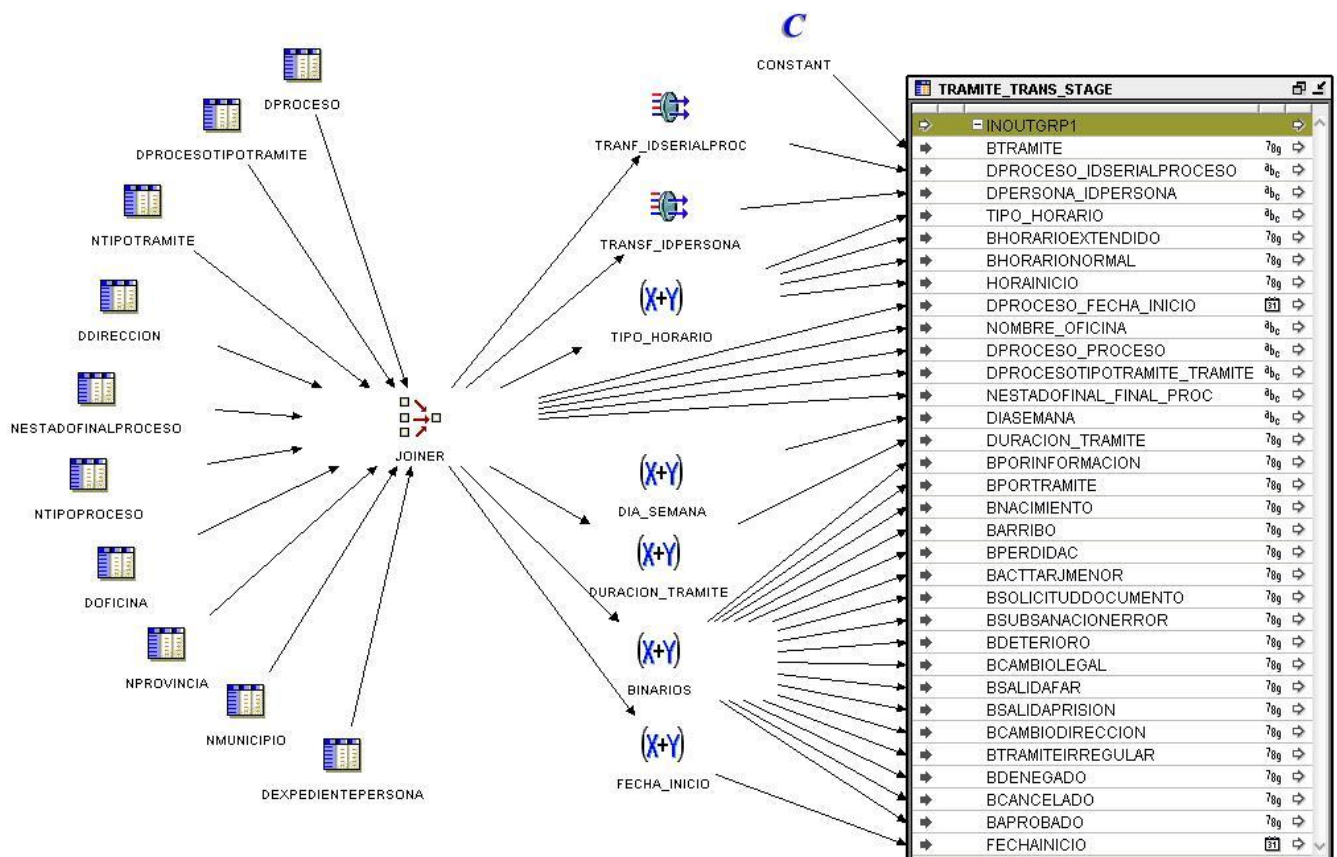


Figura 3.3 Correspondencia de la tabla temporal de trámite

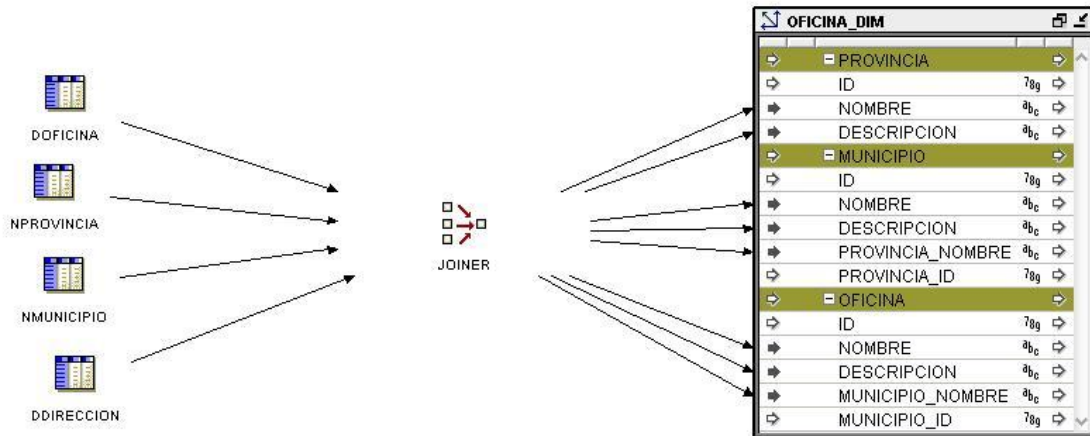


Figura 3.4 Correspondencia de la dimensión oficina

Las oficinas de trámites se encuentran distribuidas por todo el país en sus respectivas provincias y municipios. Debido a la simplicidad de la correspondencia para cargar la dimensión y buscando ganar en tiempo no resulta necesaria una tabla de almacenamiento intermedio. Por tanto, la correspondencia de la dimensión oficina como se aprecia en la figura 3.4 se carga directamente de las fuentes de datos.

Es preciso señalar que en el caso de la dimensión fecha se utilizó la implementación que ofrece OWB (ver [anexo 14](#)).

Tablas

El almacén cuenta con 11 tablas, 3 de almacenamiento temporal, 5 de dimensiones y 3 de hechos, las cuales se muestran a continuación:

Tablas		
Almacenamiento Temporal	Dimensiones	Hechos
PERSONA_TRANSTAGE	PERSONA_DIM	PERSONA_FACT
CAMBIODIRECCION_TRANSTAGE	CAMBIODIRECCION_DIM	CAMBIODIRECCION_FACT
TRAMITE_TRANSTAGE	TRAMITE_DIM	TRAMITE_FACT
	OFICINA_DIM	
	FECHA_DIM	

Tabla 3.1 Tablas del Almacén de Datos



Capítulo 3: Implementación y pruebas

DAY_IDX	FECHA_DIM	DAY_START_DATE
CALENDAR_YEAR_IDX	FECHA_DIM	CAL_YEAR_START_DAY
CALENDAR_QUARTER_IDX	FECHA_DIM	CAL_QUARTER_START_DAY
CALENDAR_MONTH_IDX	FECHA_DIM	CAL_MONTH_START_DAY
PERSONA_DIM_IDX	FACT_PERSONA	PERSONA_DIM
FECHA_DIM_IDX	FACT_PERSONA	PERSONA_DIM
CAMBIODIRECCION_DIM_IDX	FACT_CAMBIODIRECCION	CAMBIODIRECCION_DIM
PERSONA_DIM_IDX_1	FACT_CAMBIODIRECCION	PERSONA_DIM
FECHA_DIM_IDX_1	FACT_CAMBIODIRECCION	FECHA_DIM
TRAMITE_DIM_IDX	FACT_TRAMITE	TRAMITE_DIM
PERSONA_DIM_IDX_2	FACT_TRAMITE	PERSONA_DIM
OFICINA_DIM_IDX	FACT_TRAMITE	OFICINA_DIM
FECHA_DIM_IDX_2	FACT_TRAMITE	FECHA_DIM
PERSONA_DIM_DIMENSION_KEY_PK	PERSONA_DIM	DIMENSIÓN_KEY
CAMBIODIR_DIMENSION_KEY_PK	CAMBIO_DIRECCION_DIM	DIMENSIÓN_KEY
OFICINA_D_DIMENSION_KEY_PK	OFICINA_DIM	DIMENSIÓN_KEY
TRAMITE_D_DIMENSION_KEY_PK	TRAMITE_DIM	DIMENSIÓN_KEY
FECHA_DIM_DAY_START_DAY_PK	FECHA_DIM	DIMENSIÓN_KEY

Tabla 3.3 Índices del Almacén de Datos

Flujos de Procesos

Oracle Warehouse Builder brinda la posibilidad de crear los Flujos de Procesos los cuales permiten modelar, automatizar y mejorar continuamente los procesos de negocios de acuerdo a las reglas definidas por los usuarios. Además permiten ejecutar de forma paralela o secuencial distintos eventos de negocio (37) (38). Una vez diseñados todos los procesos ETL, estos son cargados a través de los flujos que han sido diseñados y no de manera independiente.



Capítulo 3: Implementación y pruebas

La carga inicial del almacén se ejecuta a través de un Flujo de Procesos como se muestra en la siguiente figura:

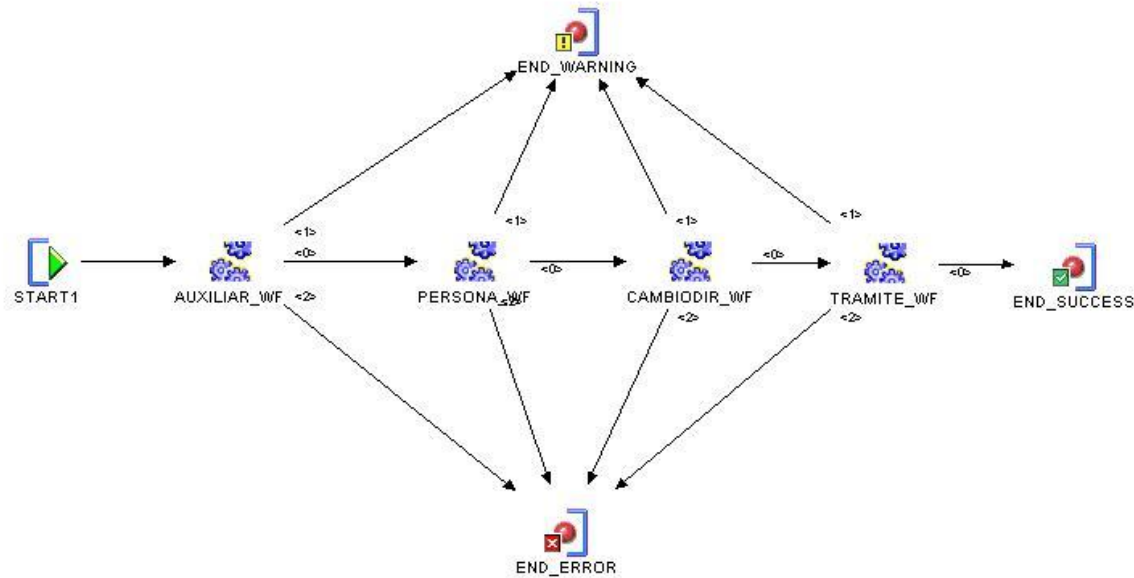


Figura 3.5 Flujo de proceso general de carga

Todos los flujos inician con el componente STAR1, luego ejecutan la secuencia de las actividades que se realizan y finalmente culminan con el componente END_SUCCESS. Como se ilustra en la figura 3.5, para cargar el almacén primeramente se ejecuta el subproceso que carga las dimensiones que no dependen del almacenamiento temporal (dimensión fecha y dimensión oficina) y luego se define el orden en que se cargan los restantes subprocesos. Para todos los subprocesos se han diseñado flujos independientes.

A continuación se muestra el flujo de procesos del ETL de persona donde primeramente se carga la tabla de almacenamiento temporal de persona, luego se carga la tabla de la dimensión persona y finalmente se carga la tabla del hecho¹⁸ persona (Ver el resto de los flujos de procesos en los [anexos 15 y 16](#)).

¹⁸ En la herramienta Oracle las tablas de hechos son las tablas físicas de los cubos.

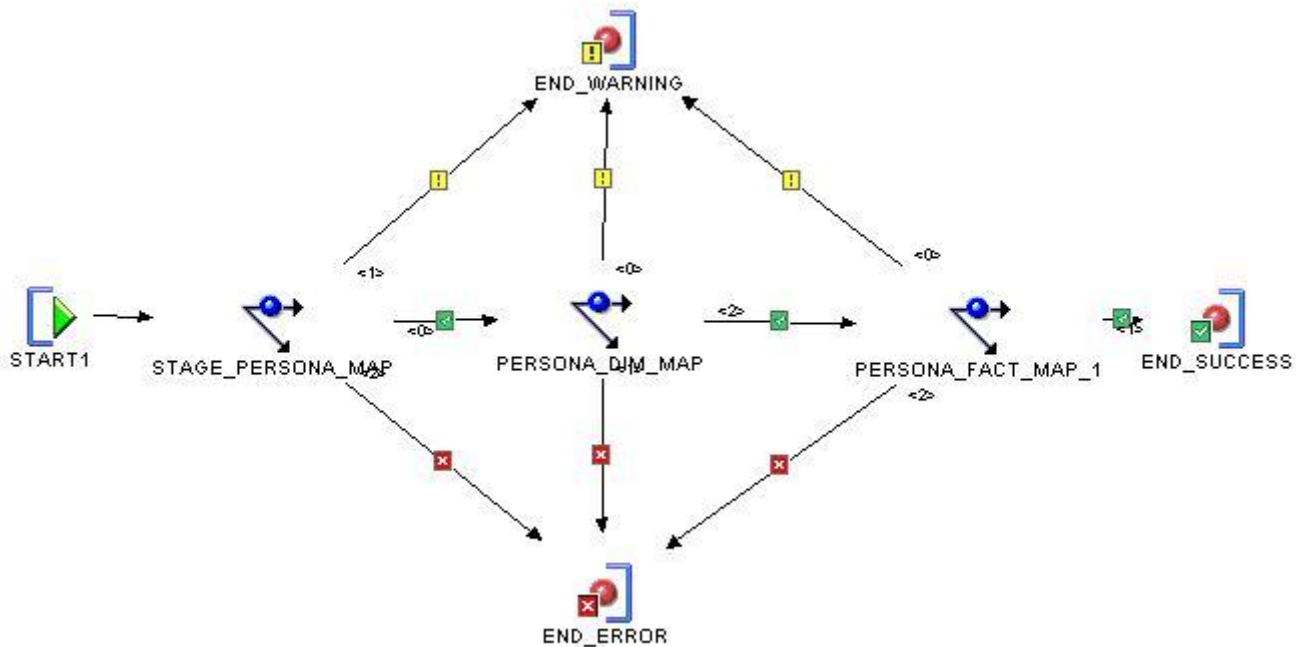


Figura 3.6 Flujo del proceso ETL de Persona

Una vez cargado el Almacén de Datos, se establece la política de actualización de los datos.

- La información se actualizará todos los días laborales a las 12:00 de la noche para no afectar el rendimiento del SUIN en las oficinas de trámites.
- Serán modificados todos los datos que presenten cambios, excepto aquellos campos que fueron definidos para almacenar información histórica (ver [anexo 17](#)).
- Serán añadidos todos los nuevos datos resultados de las transacciones diarias del SUIN.

3.3 Implementación de la aplicación web

Luego de concluido el diseño y la implementación del almacén están creadas las condiciones para comenzar el desarrollo de la aplicación web, que permitirá mostrar al usuario los principales reportes que responden a sus necesidades de información en función de los datos almacenados.

El procesamiento analítico en línea (OLAP) es un factor esencial para la construcción de la solución pues constituye el motor de consultas especializado del almacén. Por tanto, en este punto del desarrollo se



Capítulo 3: Implementación y pruebas

conforman los cubos OLAP, los cuáles son la estructura fundamental que permite extraer de forma rápida y eficiente la información que se requiere.

Los cubos OLAP se implementaron uno por cada tabla de hecho, asociando sus dimensiones correspondientes. Para cada dimensión se establecieron las jerarquías de los atributos de las mismas, lo cual es fundamental para organizar los niveles por los que se efectuarán las búsquedas. La principal ventaja de manejar jerarquías reside en poder navegar y profundizar en los datos por las diferentes dimensiones y niveles de detalles, para tener una visión dinámica de los mismos.

A continuación se muestran las jerarquías de las dimensiones cambio de dirección y fecha, donde cada número se corresponde con el nivel de la jerarquía (ver las restantes jerarquías en los [anexos del 18 al 20](#)):

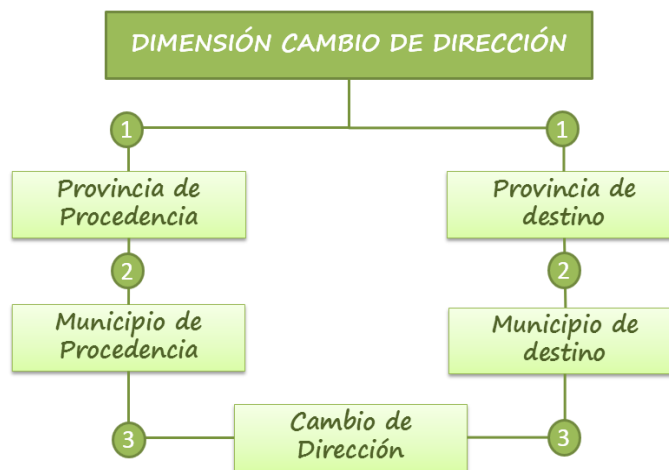


Figura 3.7 Jerarquía de la dimensión Cambio de Dirección

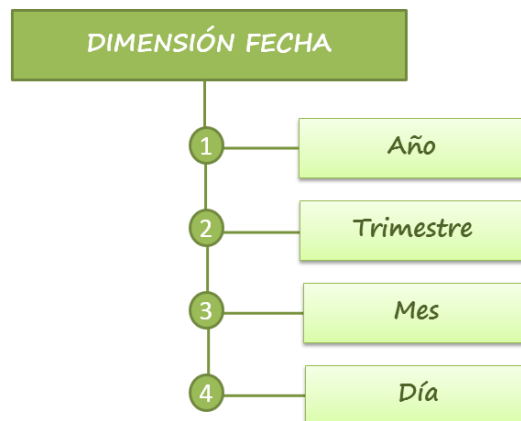


Figura 3.8 Jerarquía de la dimensión Fecha



Capítulo 3: Implementación y pruebas

Las jerarquías definidas para el almacén poseen las siguientes características:

- En una dimensión pueden existir varias.
- Están compuestas por dos o más niveles. Este requisito es obligatorio según ORACLE.
- Entre un nivel superior y uno inferior existe una relación de uno a muchos.

Reportes candidatos

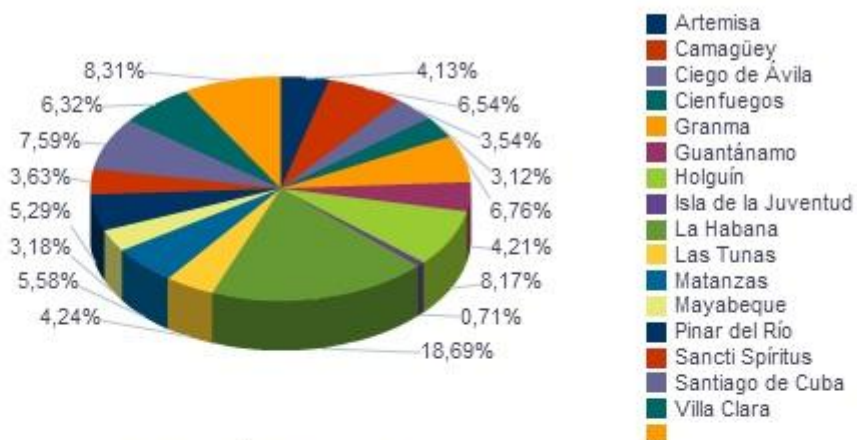
Fueron implementados los reportes candidatos según los requisitos de información del cliente. En dependencia de los tipos de reportes varía la complejidad de los mismos. Para el diseño de los reportes estadísticos se hizo uso de tablas para mostrar los datos y gráficas para ilustrar los indicadores. A continuación se muestran algunos de los reportes:



Figura 3.9 Reporte de grupos etarios en Cuba



Porcentaje de Personas x Provincias



Distribución de personas

Nombre de provincia	No. Personas
Artemisa	563.501
Camagüey	893.303
Ciego de Ávila	484.207
Cienfuegos	426.792
Granma	923.153
Guantánamo	574.542
Holguín	1.116.226
Isla de la Juventud	96.787
La Habana	2.553.106
Las Tunas	579.004
Matanzas	761.563
Mayabeque	434.327
Pinar del Río	722.550
Sancti Spiritus	495.533
Santiago de Cuba	1.036.295
Villa Clara	863.622
	1.135.275
Suma Total	13.659.786

Figura 3.10 Reporte de densidad poblacional



Nivel de fiabilidad del servicio

Provincia	Municipio	Trámite irregular	Trámite	Nivel de Fiabilidad del Servicio
Artemisa	Alquízar	13	675	98,07%
	Artemisa	41	1.431	97,13%
	Bahía Honda	8	545	98,53%
	Bauta	11	236	95,34%
	Caimito	11	392	97,19%
	Candelaria	11	403	97,27%
	Guanajay	21	577	96,36%
	Güira de Melena	14	447	96,87%
	Mariel	3	508	99,41%
	San Antonio de los Baños	17	723	97,65%
	San Cristóbal	0	68	100,00%

Nivel de fiabilidad

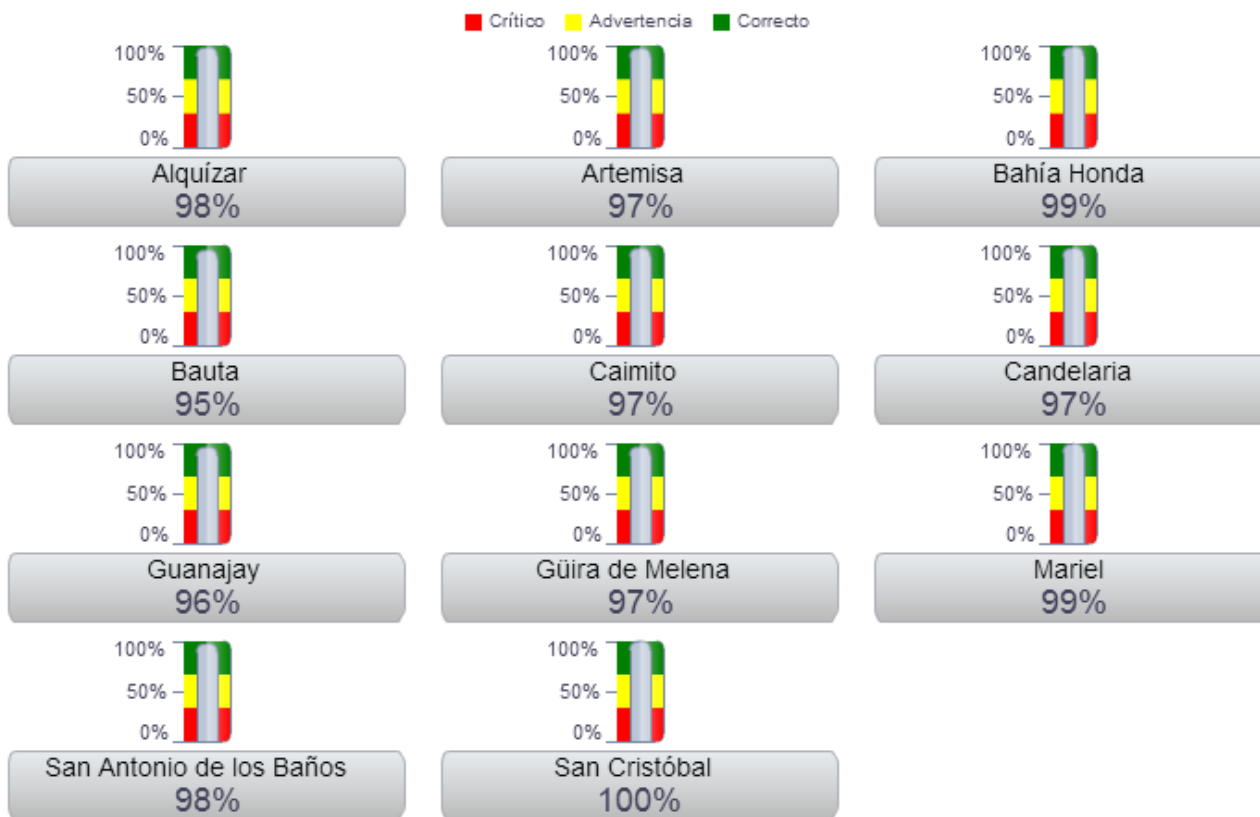


Figura 3.11 Reporte del indicador nivel de fiabilidad del servicio del CMI



3.4 Despliegue de la solución

El diagrama de despliegue permite mostrar la arquitectura en tiempo de ejecución del sistema respecto a hardware y a software (39). A continuación se muestra el diagrama de despliegue propuesto para la solución:

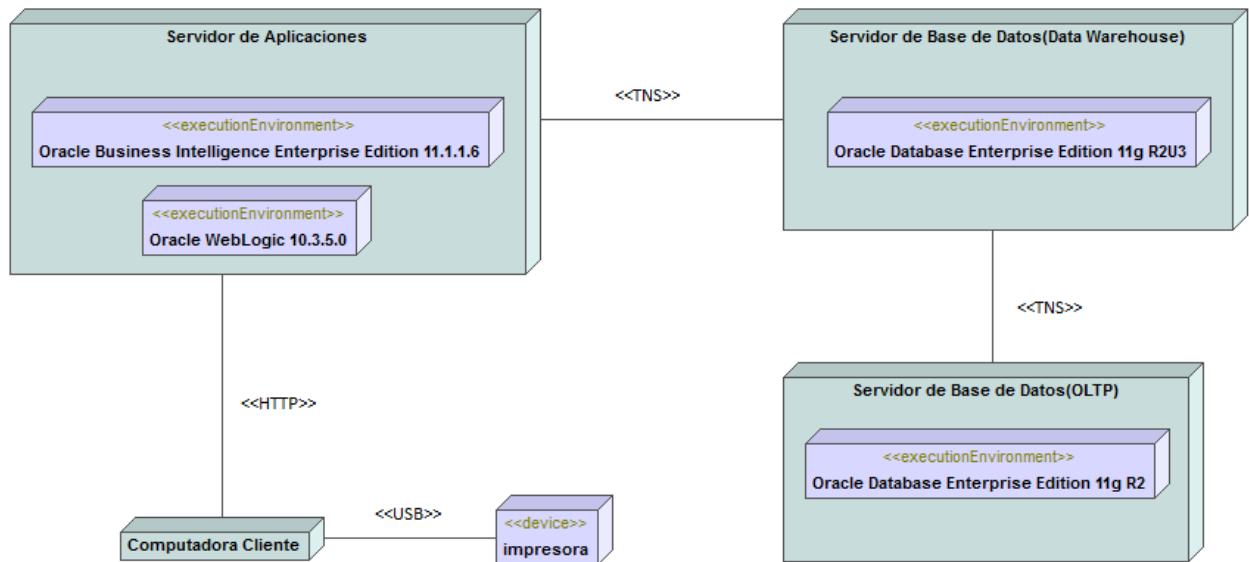


Figura 3.12 Diagrama de despliegue

Para el despliegue de la solución se propone utilizar un servidor de Base de Datos que posea Oracle Database Enterprise Edition 11gR2 (sería el servidor del SUIN que ya se encuentra operativo), de donde se extraen los datos a través del protocolo TNS hacia otro servidor de Base de Datos que requiere Oracle Database Enterprise Edition 11gR2U3 (sería el servidor del Data Warehouse), el cual se comunica mediante el protocolo TNS con el servidor de aplicaciones que debe contar con Oracle Business Intelligence Enterprise Edition 11.1.1.6 y Oracle WebLogic 10.3.5.0. Las computadoras clientes (serían las que ya se encuentran operativas en las oficinas de trámites) consultarán el servidor de aplicaciones mediante el protocolo HTTP y contarán con una impresora conectada a través del puerto USB.

El servidor de Bases de Datos del almacén debe contar con un mínimo de 250 Giga Bytes de capacidad de almacenamiento en disco, 8 Gigas de RAM y un procesador de 4 núcleos. El servidor de aplicaciones debe contar con un mínimo de 100 Giga Bytes de capacidad de almacenamiento en disco, 8 Gigas de RAM y un procesador de 8 núcleos.



3.5 Pruebas

Las pruebas de software se aplican en la fase de prueba y validación del desarrollo de un sistema de información y son un conjunto de herramientas técnicas y métodos que tienen como objeto comprobar los requerimientos establecidos con los usuarios o beneficiarios del producto. Estas técnicas tienen como objeto detectar problemas y van desde el uso del ingenio por parte del personal de pruebas hasta herramientas automáticas que ayudan a aliviar el peso y el costo de esta actividad. Las pruebas no pueden asegurar la ausencia de errores; sólo puede demostrar que existen defectos en el software (40).

Para validar el software y garantizar que presenta la calidad requerida se realizaron varios tipos de pruebas:

3.5.1 Lista de chequeo

Para evaluar la calidad en la construcción de la solución se utilizaron listas de chequeo. Una lista de chequeo contiene un listado de preguntas en forma de cuestionario que sirven para verificar el grado de cumplimiento de determinadas reglas establecidas. Cada pregunta tiene asociada una evaluación que da una medida del grado de cumplimiento. Además, sirven de guía para realizar tanto control interno como externo, pues el personal externo al proceso de construcción del software puede verificar de manera rápida y eficaz la calidad del mismo.

Los elementos que forman parte de la estructura de la lista de chequeo son:

- **Peso:** define si el indicador a evaluar es crítico o no.
- **Indicador:** representa la pregunta o indicador a evaluar.
- **Evaluación:** es la forma de evaluar el indicador en cuestión. El mismo se evalúa de 1 en caso de que exista alguna dificultad y 0 en caso contrario.
- **No. Errores:** especifica la cantidad de errores encontrados sobre el indicador.
- **Comentario:** especifica los señalamientos o sugerencias que quiera incluir la persona que aplica la lista de chequeo. Pueden o no existir señalamientos o sugerencias.

A continuación se muestra la lista de chequeo para evaluar el diseño e implementación del almacén:

Nro.	Peso	Indicador	Evaluación	No. Errores	Comentarios
1	Crítico	¿Se realizó un análisis profundo del área de la organización, responsable de la utilización del DWH?			



Capítulo 3: Implementación y pruebas

2	Crítico	¿Se realizó un análisis profundo de las fuentes de datos que poblarán el DWH?			
3	Crítico	¿Se realizaron entrevistas, cuestionarios u observaciones para identificar las preguntas del negocio y los requisitos de información?			
4		¿Se identificaron correctamente los indicadores por los cuales se medirán las preguntas del negocio?			
5	Crítico	¿Se identificaron correctamente las perspectivas desde las cuales se analizarán las preguntas del negocio?			
6		¿Los indicadores son valores numéricos y representan lo que se desea analizar concretamente?			
7		¿Quedó explícito cómo se calculará cada indicador?			
8		¿El modelo conceptual permite observar con claridad el alcance del Almacén de Datos?			
9		¿Se definieron las correspondencias que existen entre de las tablas de las fuentes de datos con las perspectivas representadas en el modelo conceptual?			
10		¿Se seleccionaron correctamente los campos que integrarán cada perspectiva y su correspondiente origen OLTP?			
11		¿La granularidad definida representa el nivel de detalle con el que se desea almacenar la información sobre el negocio analizado?			
12		¿La granularidad definida permite tener mayores posibilidades analíticas?			
13	Crítico	¿La tipología de esquema seleccionada es la que mejor se adapta a los requerimientos y necesidades del usuario?			
14		¿Se diseñaron correctamente las tablas de dimensiones?			
15		¿Se diseñaron correctamente las tablas de hechos?			



Capítulo 3: Implementación y pruebas

16		¿Se realizaron las uniones correspondientes entre las tablas de dimensiones y las tablas de hechos?			
17		¿La jerarquía planteada tiene una relación lógica “padre-hijo”?			
18	Crítico	¿Permite el modelo lógico tener una primera representación de la estructura de los datos?			
19	Crítico	¿Se implementaron las correspondencias de las tablas de almacenamiento temporal correctamente?			
20		¿Se implementaron correctamente las correspondencias de las dimensiones y los cubos?			
21		¿Se implementaron los ETL en correspondencia con el diseño?			
22		¿Los reportes se corresponden con las preguntas identificadas en el negocio?			
23		¿Los reportes pueden configurarse a través de la interfaz del sistema?			
24		¿La interfaz está orientada a facilitar el uso de las funciones del sistema por parte de los usuarios?			
25		¿Los usuarios son capaces de manipular los resultados de manera que se ajusten a sus necesidades, conformando nuevos reportes?			
26		¿El sistema responde de una forma rápida a la información solicitada por el usuario?			
27	Crítico	¿Los datos mostrados apoyan la toma de decisiones en el MININT?			
28	Crítico	¿Los cambios en los datos se reflejan en los reportes?			

Tabla 3.4 Lista de chequeo de la Solución de Inteligencia Institucional

La siguiente figura refleja en un gráfico de barras el comportamiento de los indicadores definidos en la lista de chequeo. En total se analizaron 28 indicadores, de los cuales 9 son críticos. Luego de aplicada la lista de chequeo se generaron cero no conformidades.

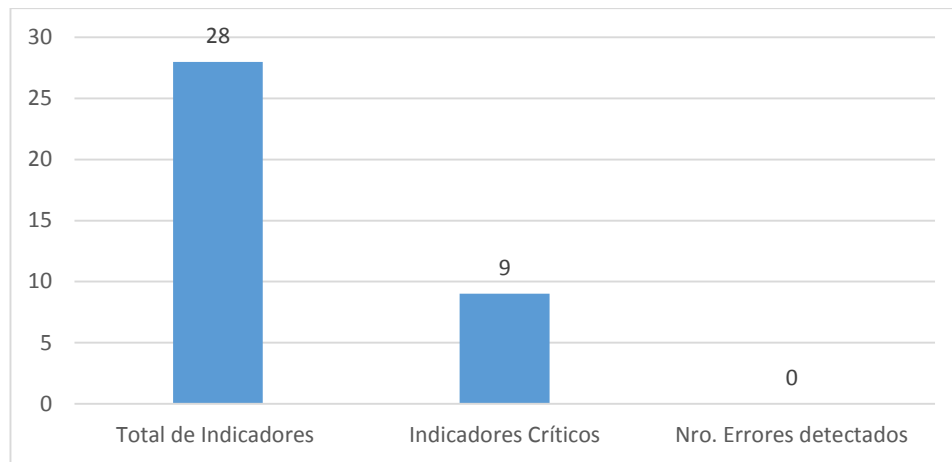


Figura 3.13 Comportamiento de los Indicadores

3.5.2 Análisis de los tiempos de respuesta de los reportes

Para comprobar la agilidad en el acceso a la información, fue realizada una comparación entre los tiempos de respuesta de la Base de Datos del SUIN y la aplicación web, para esta prueba fue tomada una muestra de 6 reportes del proceso de persona, que es el de más altos tiempos de respuesta. A continuación se muestra una tabla comparativa que evidencia la eficacia de la solución de Inteligencia Institucional.

Reportes	Tiempo SUIN (min)	Tiempo Solución de Inteligencia Institucional (min)
Cantidad de personas por sexo	4:24	1:16
Cantidad de personas por color de la piel tanto mayores como menores de edad	10:16	3:42
Cantidad de personas por sexo y rango de edad	22:15	4:29
Porcentajes por grupos etarios	15:12	1:47
Cantidad de personas por sexo y color de la piel	10:09	3:59
Cantidad de personas por color de la piel	5:16	1:19
Promedio	11:15	2.52

Tabla 3.5 Comparación de los tiempos de respuesta en el acceso a la información



3.5.3 Prueba de integridad de datos

La prueba de integridad de datos es el proceso de encontrar datos imprecisos en el Almacén de Datos. Esta prueba, tiene como objetivo encontrar errores en la manera en que se almacenan los datos, de manera que los mismos no sean comprometidos en el proceso de actualización, restauración o recuperación. Además, este tipo de pruebas se realiza para descubrir deficiencias de diseño que pueden provocar corrupción en los datos, acceso no autorizado a los datos, pérdida de integridad de los datos y pérdida de un rendimiento adecuado de las transacciones (41) (42).

Luego de aplicar la prueba de integridad de datos sobre el almacén de datos para verificar el correcto diseño e implementación de los proceso ETL y que los datos fueron cargados correctamente en el Almacén de Datos, y sobre la aplicación web para verificar que los datos mostrados se correspondían con los datos guardados en el almacén: se detectaron cero no conformidades.

A continuación se muestra el resultado de una consulta realizada sobre la Base de Datos del SUIN (ver figura 3.9) y del reporte diseñado en el centro de reportes (ver figura 3.10) para verificar la integridad de los datos.

	PROVINCIA_NOMBRE	PERSONA	FALLECIDO	PIEL_BLANCA	PIEL_MESTIZA	PIEL_NEGRA	MASCULINO	FEMENINO
1	Artemisa	563501	21217	323106	55158	53628	290949	272552
2	Camagüey	893303	33267	523872	75026	68696	459276	434027
3	Ciego de Ávila	484207	18537	368556	48103	30996	249599	234608
4	Cienfuegos	426792	14614	270928	49577	32566	218250	208542
5	Granma	923153	31374	258405	405535	23841	473151	450002
6	Guantánamo	574542	16948	83011	250871	84962	291065	283477
7	Holguín	1116226	40178	706529	108267	39743	572712	543514
8	Isla de la Juventud	96787	3090	44583	21189	8163	49937	46850
9	La Habana	2553106	136884	1205410	430409	383144	1249358	1303748
10	Las Tunas	579004	18669	364149	66860	30976	297911	281093
11	Matanzas	761563	30689	454978	80747	78822	388396	373167
12	Mayabeque	434327	19171	259542	40888	32133	224348	209979
13	Pinar del Río	722550	25446	416536	70077	78575	372464	350086
14	Sancti Spiritus	495533	20222	344867	32145	26667	254749	240784
15	Santiago de Cuba	1036295	35808	218680	384355	190871	523057	513238
16	Villa Clara	863622	39652	601989	68165	45466	442388	421234
17		1135275	810009	306268	60800	50149	596614	538661

Figura 3.14 Resultado de la consulta efectuada sobre la Base de Datos del SUIN



Capítulo 3: Implementación y pruebas

Provincia Nombre	Persona	Fallecido	Pielblanca	Pielmestiza	Pielnegra	Masculino	Femenino
Artemisa	563.501	21.217	323.106	55.158	53.628	290.949	272.552
Camagüey	893.303	33.267	523.872	75.026	68.696	459.276	434.027
Ciego de Ávila	484.207	18.537	368.556	48.103	30.996	249.599	234.608
Cienfuegos	426.792	14.614	270.928	49.577	32.566	218.250	208.542
Granma	923.153	31.374	258.405	405.535	23.841	473.151	450.002
Guantánamo	574.542	16.948	83.011	250.871	84.962	291.065	283.477
Holguín	1.116.226	40.178	706.529	108.267	39.743	572.712	543.514
Isla de la Juventud	96.787	3.090	44.583	21.189	8.163	49.937	46.850
La Habana	2.553.106	136.884	1.205.410	430.409	383.144	1.249.358	1.303.748
Las Tunas	579.004	18.669	364.149	66.860	30.976	297.911	281.093
Matanzas	761.563	30.689	454.978	80.747	78.822	388.396	373.167
Mayabeque	434.327	19.171	259.542	40.888	32.133	224.348	209.979
Pinar del Río	722.550	25.446	416.536	70.077	78.575	372.464	350.086
Sancti Spíritus	495.533	20.222	344.867	32.145	26.667	254.749	240.784
Santiago de Cuba	1.036.295	35.808	218.680	384.355	190.871	523.057	513.238
Villa Clara	863.622	39.652	601.989	68.165	45.466	442.388	421.234
	1.135.275	810.009	306.268	60.800	50.149	596.614	538.661

Figura 3.15 Reporte mostrado en la aplicación web.

3.5.4 Prueba de Aceptación

Luego de concluir el desarrollo de la Solución de Inteligencia Institucional para el Sistema Único de Identificación Nacional, así como las validaciones expuestas anteriormente, se realizó un encuentro con los especialistas del SUIN, en el cual se hizo entrega de una carta de aceptación del producto, que se muestra a continuación.



"Aniversario 55 de la Revolución"

Mayo 20 de 2013.

ACTA DE ACEPTACIÓN DEL PRODUCTO:

Solución de Inteligencia Institucional para el SUIN

Por este medio se certifica que el producto realizado por los estudiantes con motivo de su trabajo de diploma cumple con los requisitos establecidos y aprobados por el proyecto Identidad Cuba.

Han sido entregados para trabajos futuros los principales artefactos resultantes de la investigación: software, especificación de requerimientos y documento de tesis.

Otras observaciones:

Para que así conste firman:

 _____ Jefe de Proyecto Msc. Ing. Erick de la Vega García	 _____ Jefe de Desarrollo Ing. Reynier Blanco Zambrano	 _____ Jefe de Producto Msc. Ing. Liudnet Caballero Duboy
--	--	--

Figura 3.16 Acta de aceptación del producto



3.6 Conclusiones parciales

En este capítulo, la estructuración del Almacén de Datos mediante tablas, índices y secuencias, permitió soportar la correcta carga del almacén y lograr mayor rapidez al realizar las consultas. La implementación del proceso de integración de datos permitió cargar el Almacén de Datos en correspondencia con el modelo de arquitectura definido. La implementación de los flujos de procesos para cargar y actualizar el Almacén de Datos contribuyó a mejorar el rendimiento del almacén y permitió agilizar el proceso de carga de los datos. La implementación de los reportes candidatos en la aplicación web posibilitó mostrar la información en correspondencia con los requerimientos del cliente. La definición de listas de chequeo permitió evaluar la calidad en la construcción del software, identificándose cero no conformidades luego de aplicada. La prueba de integridad de datos permitió asegurarse no solo que la información fuera correcta, sino que también que estuviera actualizada. La prueba de aceptación permitió avalar el producto y obtener una carta de aceptación por parte de los funcionarios del SUIN.



Conclusiones

- Luego del análisis de los principales conceptos asociados a la Inteligencia de Negocio así como de las necesidades de información presentes en el MININT se logró realizar un diseño flexible, acorde a los grandes volúmenes de información a tratar.
- Con el diseño lógico del almacén y la arquitectura de la solución propuesta se garantiza la integridad y el carácter histórico de la información disponible en el Almacén de Datos, así como la optimización en los tiempos de respuesta para la generación de cada uno de los reportes que se realicen.
- Al digitalizar el Cuadro de Mando Integral y contar con un medio de almacenamiento de la información actual e histórica, de las personas y trámites realizados en el SUIN, se permite mejorar el proceso de toma de decisiones en cada uno de los niveles del MININT.
- Al realizar las pruebas de integridad al Almacén de Datos, las Listas de Chequeo y de validar con el cliente el almacén y la aplicación web, se avaló la calidad de la propuesta de solución.



Recomendaciones

Integrar al Almacén de Datos información referente a otras Áreas del MININT o procesos del SUIN, para elevar el alcance de la información y se permita aumentar los conocimientos como soporte al proceso de toma de decisiones.

Confeccionar el manual de usuario de la solución, para facilitar el trabajo con la herramienta de Inteligencia Institucional.

Implementar técnicas de minerías de datos sobre el Almacén de Datos, para extraer información oculta, patrones y tendencias mediante el análisis de los datos.



Referencias bibliográficas

- (01) Dresner, H. (1989). *Business Intelligence*. Gartner Group.
- (02) Antunez, I. M. (2008). *La inteligencia de negocios desde la perspectiva cubana. Retos y tendencias*. Instituto de Información Científica y Tecnológica (IDIC), Consultoría Biomundi.
- (03) González Santos, A. P. (2012). *Solución de Inteligencia de Negocios para el SUIN*. Universidad de las Ciencias Informáticas, Facultad 1.
- (04) Martínez García, J. H. (2010). *La inteligencia de negocios como herramienta para la toma de decisiones estratégicas en las empresas: análisis de su aplicabilidad*. Universidad Nacional de Colombia.
- (05) Alejandro, P. A. (2006). *Inteligencia de Negocios: una propuesta para su desarrollo en las organizaciones*. Instituto Politécnico Nacional, Dirección de publicaciones, Mexico.
- (06) Curto Díaz, J. C. (2010). *Introducción al Business Intelligence*. Barcelona: Editorial UOC.
- (07) Oramas, J. (2009). La inteligencia de negocios, un concepto informático. *Revista Sistemas N° 111*.
- (08) Hernández López, A. C. (s.f.). *Almacenes de datos aplicados a la seguridad ciudadana*. Universidad de Ciencias Informáticas. Recuperado el 18 de Febrero de 2013, de <https://publicaciones.uci.cu/index.php/SC/article/download/38/39>
- (09) Del Giudice, G. A. (2010). *Proyecto de implementación de un Data Warehouse para universidades nacionales*. INSTITUTO UNIVERSITARIO AERONAUTICO, FACULTAD DE INGENIERIA, Córdoba.
- (10) Bernabeu, R. D. (2007). *DATA WAREHOUSING: Investigación y Sistematización de Conceptos - HEFESTO: Metodología propia para la construcción de un Datawarehouse*. Córdoba, Argentina.
- (11) Mendez, A. M.-M. (s.f.). *Fundamentos de Data Warehouse*. Instituto Tecnológico de Buenos Aires, Centro de Actualización Permanente en Ingeniería del Software, Buenos Aires.
- (12) Inmon, W. H. (1996). *Building the Data Warehouse*. New York: Computer, Wiley.
- (13) Inmon, W. H. (2000). *What is a data warehouse?*
- (14) Casares, C. (s.f.). Data Warehousing. Obtenido de <http://www.programacion.com/bbdd/tutorial/warehouse/>
- (15) Kimbal, R. y. (1998). *The Data Warehouse Lifecycle Toolkit*. Wiley Edition.
- (16) Gutiérrez Echeverría, D. (s.f.). *Data Warehouse*. Universidad Iberoamericana, Unidad de Recursos Informáticos, México, D.F.
- (17) Hurtado Torres, M. y. (s.f.). *BASES DE DATOS Y DATA WAREHOUSE: HERRAMIENTAS ESTRATÉGICAS PARA LA EFICACIA COMERCIAL*. Facultad de Ciencias Económicas y Empresariales. Universidad de Granada, Departamento de Lenguajes y Sistemas Informáticos.



- (18) DATAMART. (26 de Septiembre de 2009). Recuperado el 20 de Enero de 2013, de TodoTecnologia.com:
<http://todotecnology.blogspot.com/2009/09/datamart.html>
- (19) Domínguez, A. B. (s.f.). Recuperado el 2 de 2 de 2013, de
http://www.programacion.com/articulo/data_warehousing_201/6.
- (20) Mateo Morales, Y. S. (2012). *Integración de Datos (ETL) y Almacenes de Datos*. Maestría en Ingeniería de Software.
- (21) Cerezal Tamargo, L. (s.f.). *Diseño Físico del Data Warehouse*. Banco Central de Cuba . Revista BETSIME. Recuperado el 13 de Febrero de 2013, de http://www.betsime.disaic.cu/secciones/tec_ma_05.htm
- (22) Gloria Wolff, C. (2002). *Modelamiento multidimensional*.
- (23) Rob, P. C. (s.f.). *Sistemas de bases de datos: diseño, implementación y administración*. Cengage Learning.
- (24) Moliner López, F. J. (2005). *Grupos A y B de Informática. Bloque específico de la Generalitat Valenciana*. Centro de Estudios SIPEUROCENTER.
- (25) Microsoft Corporation. (s.f.). *Procesamiento y modos de almacenamiento de particiones*. Recuperado el 14 de Febrero de 2013, de <http://msdn.microsoft.com/es-es/library/ms174915.aspx>
- (26) Inc Exforsys. (2005). *Design of the data warehouse: Kimball Vs Inmon*. Recuperado el 15 de Febrero de 2013, de <http://www.exforsys.com/tutorials/msas/data-warehouse-design-kimball-vs-inmon.html>
- (27) Guia de Soluciones TIC. (s.f.). *Embarcadero ER/Studio. Diseño y Mantenimiento de Bases de Datos Transaccionales*. Recuperado el 10 de Febrero de 2013, de <http://www.guiadesolucionestic.com/software-del-sistema/herramientas-de-desarrollo/herramientas-de-desarrollo-de-software/972-embarcadero-erstudio>
- (28) Diario TI. (2008). *Embarcadero presenta ER/Studio 8.0*. Obtenido de <http://diarioti.com/embarcadero-presenta-erstudio-8-0/20644?lang=es>
- (29) Baum, M. (2010). *Embarcadero ER/Studio XE Introduce More Scalable Environment for Database and Data Warehouse Modeling*. Chase Communications. Obtenido de <http://datamodel.wordpress.com/tag/data-warehousing/>
- (30) Oracle Corporation. (s.f.). Recuperado el 10 de Febrero de 2013, de www.oracle.com
- (31) Oracle Corporation. (s.f.). Recuperado el 12 de Febrero de 2013, de <http://www.oracle.com/technetwork/es/database/enterprise->
- (32) Oracle Corporation. (s.f.). Recuperado el 12 de Febrero de 2013, de <http://www.oracle.com/technetwork/es/database/enterprise-edition/overview/index.html>
- (33) Oracle Corporation. (s.f.). Obtenido de <http://www.oracle.com/lad/corporate/press/pr-lad-110427-367397-esa.html>



Referencias bibliográficas

- (34) BI-Spain.com. (2008). *Análisis de Oracle Business Intelligence Suite Enterprise Edition Plus, que ya incorpora soluciones de Hyperion*. Obtenido de <http://www.bi-spain.com/articulo/50851/oracle/analisis-de-oracle-business-intelligence-suite-enterprise-edition-plus-que-ya-incorpora-soluciones-de-hyperion>
- (35) Oracle Corporation. (s.f.). Recuperado el 12 de Febrero de 2013, de <http://www.oracle.com/technetwork/developer-tools/warehouse/index.html>
- (36) Allround Automations. (s.f.). Recuperado el 17 de Febrero de 2013, de <http://www.allroundautomations.com/plsqldev.html>
- (37) Oracle Corporation. (2009). *Oracle Warehouse Builder User's Guide 11g Release (11.1)*.
- (38) Fischer, L. (2005). *WORKFLOW HANDBOOK*.
- (39) Falgueras Campderrich, B. (2003). *Ingeniería de software*. Editorial UOC.
- (40) GALEA BRACHO, A. R. (s.f.). *GUIA DE LA UNIDAD I TECNICAS DE PRUEBA*. UNIVERSIDAD POLITECNICA MARACAIBO, DEPARTAMENTO DE INFORMATICA.
- (41) ELFRIEDE, D. y. (1999). *Automated Software Testing: Introduction, Management and Performance*. Addison-Wesley.
- (42) Glendford J, M. y. (2012). *The Art of Software Testing*. John Wiley & Sons.

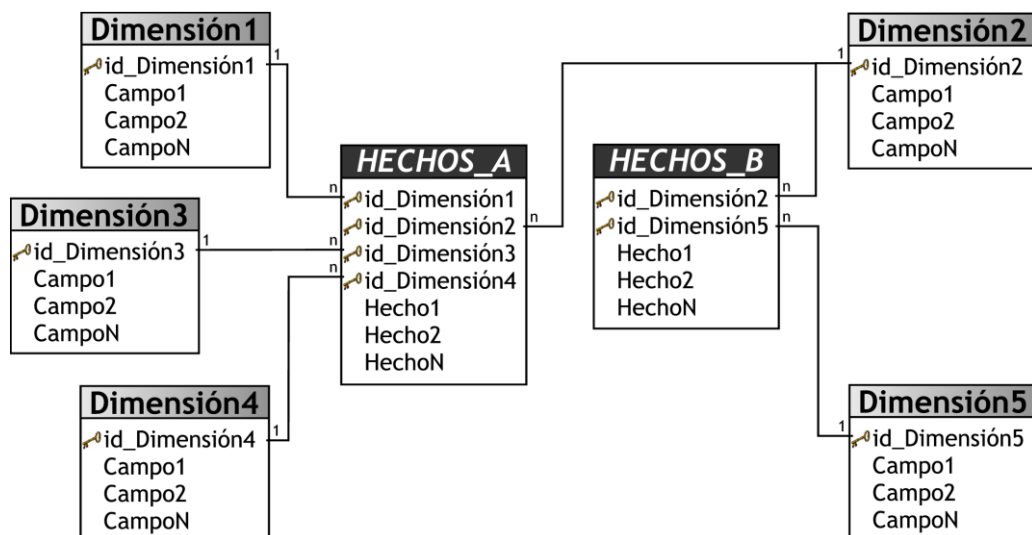


Anexos

Anexo 1: OLTP VS Data Warehouse

	OLTP	Data Warehouse
Objetivo	Soportar actividades transaccionales diarias.	Consultar y analizar información estratégica y táctica.
Tipo de datos	Operacionales.	Para la toma de decisiones.
Modelo de datos	Normalizado.	Desnormalizado.
Consulta	SQL.	SQL más extensiones.
Datos consultados	Actuales.	Actuales e históricos.
Horizonte de tiempo	60 - 90 días.	5 - 10 años.
Tipos de consultas	Repetitivas, predefinidas	No previsibles, dinámicas
Nivel de almacenamiento	Nivel de detalle.	Nivel de detalle y diferentes niveles de sumarización.
Acciones disponibles	Alta, baja, modificación y consulta.	Carga y consulta.
Número de transacciones	Elevado	Medio o bajo
Tamaño	Pequeño - Mediano.	Grande.
Tiempo de respuesta	Pequeño (segundos - minutos).	Variable (minutos - horas).
Orientación	Orientado a las aplicaciones.	Orientado al negocio.
Sello de tiempo	La clave puede o no tener un elemento de tiempo.	La clave tiene un elemento de tiempo.
Estructura	Generalmente estable.	Generalmente varía de acuerdo a su propia evolución y utilización.

Anexo 2: Esquema Constelación



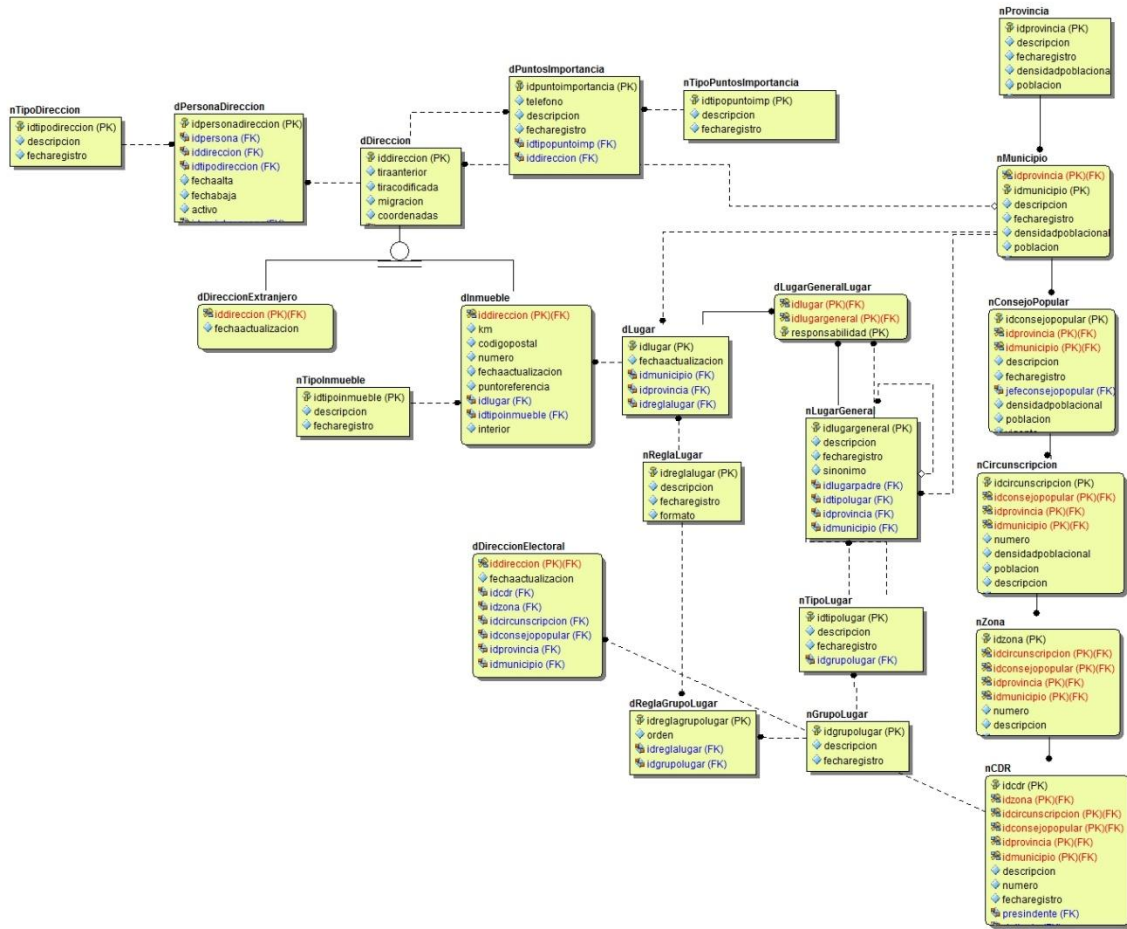


Anexo 3: Fases de la Metodología HEFESTO

- 1) Análisis de requerimientos
 - a) Identificar preguntas
 - b) Identificar indicadores y perspectivas
 - c) Modelo Conceptual
- 2) Análisis de los OLTP
 - a) Conformar indicadores
 - b) Establecer correspondencias
 - c) Nivel de granularidad
 - d) Modelo Conceptual ampliado
- 3) Modelo Lógico de Almacén de Datos
 - a) Tipo de Modelo Lógico del Almacén de Datos
 - b) Tablas de dimensiones
 - c) Tablas de hechos
 - d) Uniones
- 4) Integración de datos
 - a) Carga inicial
 - b) Actualización

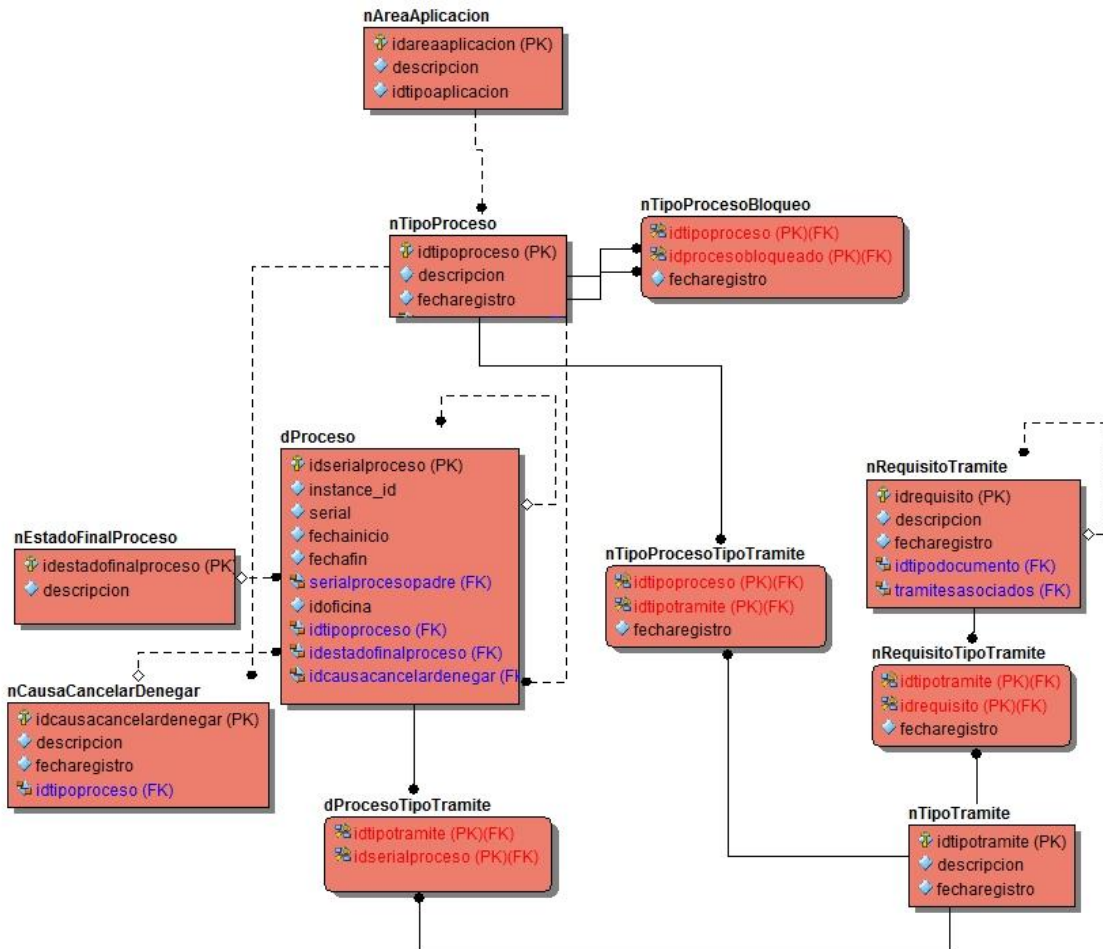


Anexo 4: Modelo de Datos del SUIN.- Submodelo Direcciones



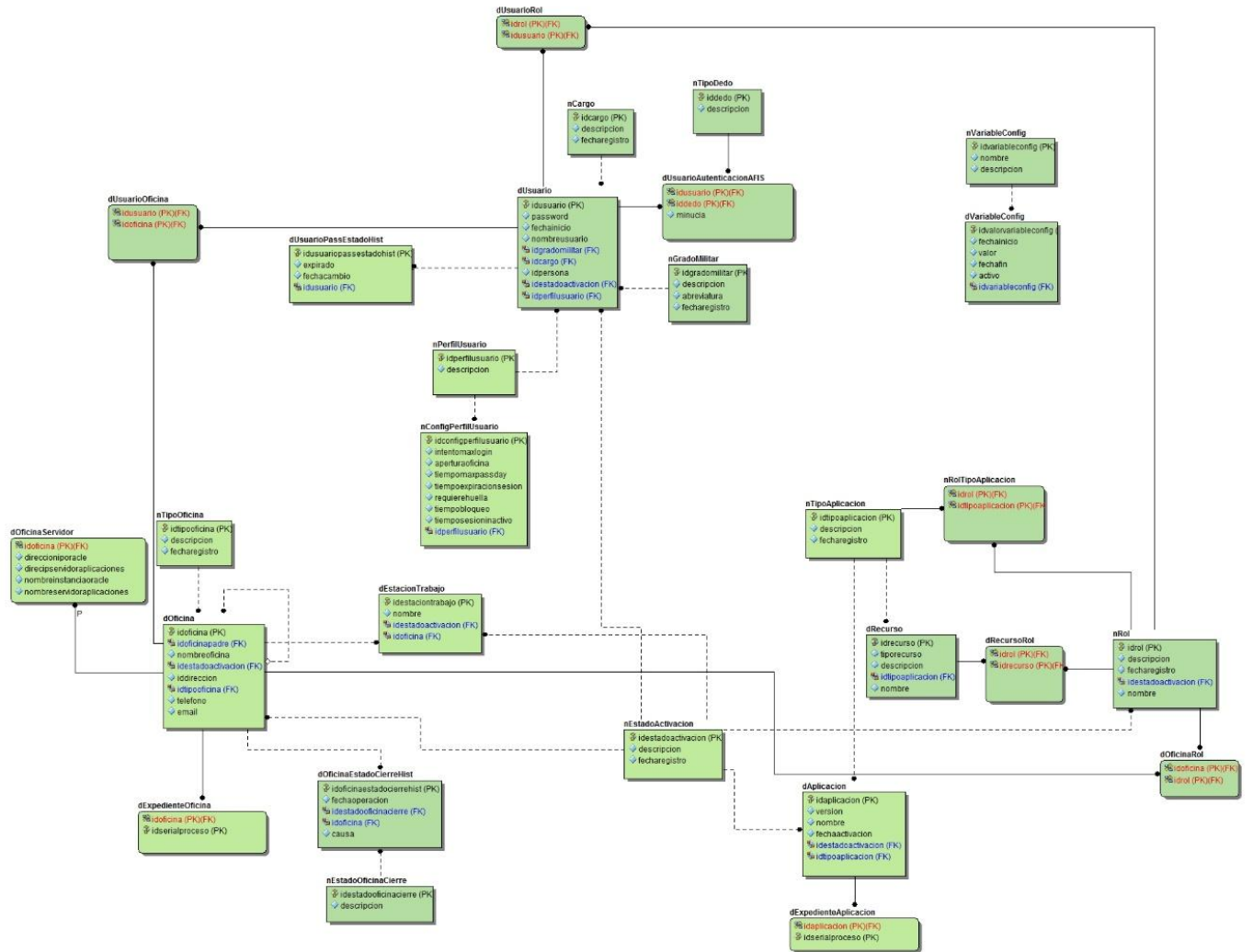


Anexo 6: Modelo de Datos del SUIN.- Submodelo Procesos



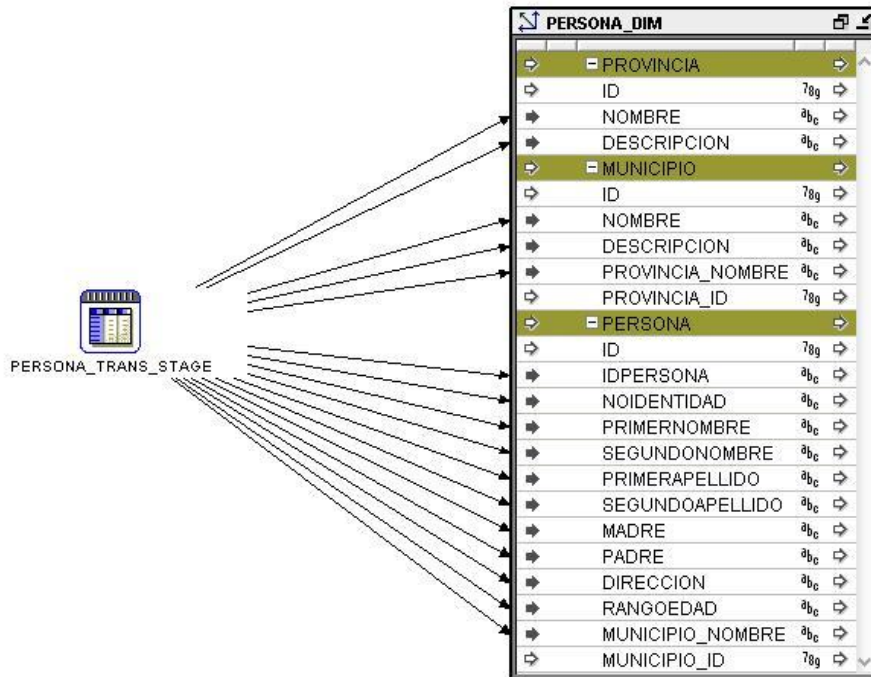


Anexo 7: Modelo de Datos del SUIN. Submodelo ADM

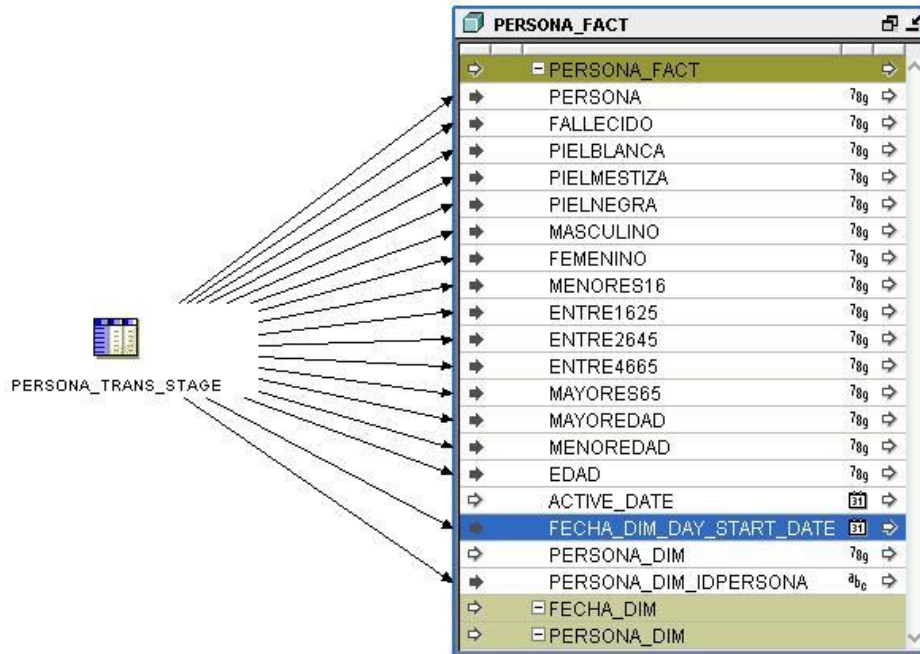




Anexo 8: Correspondencia de la dimensión Persona

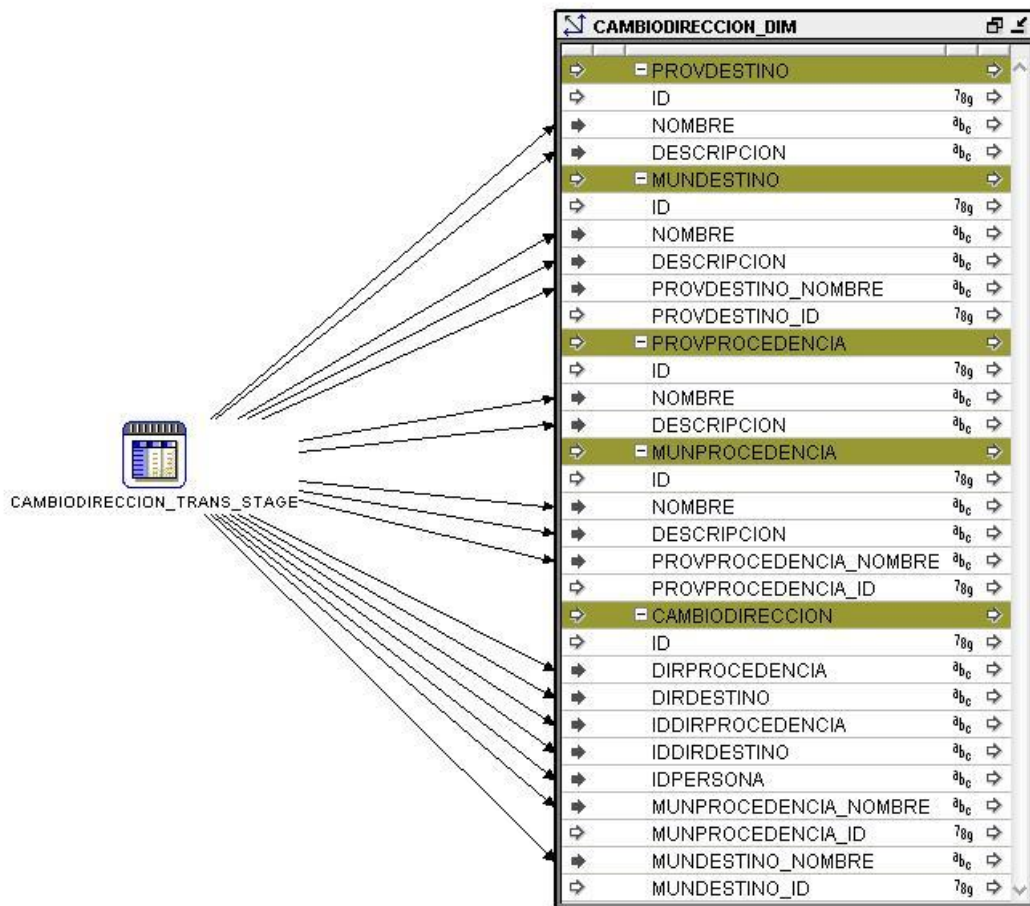


Anexo 9: Correspondencia del cubo Persona

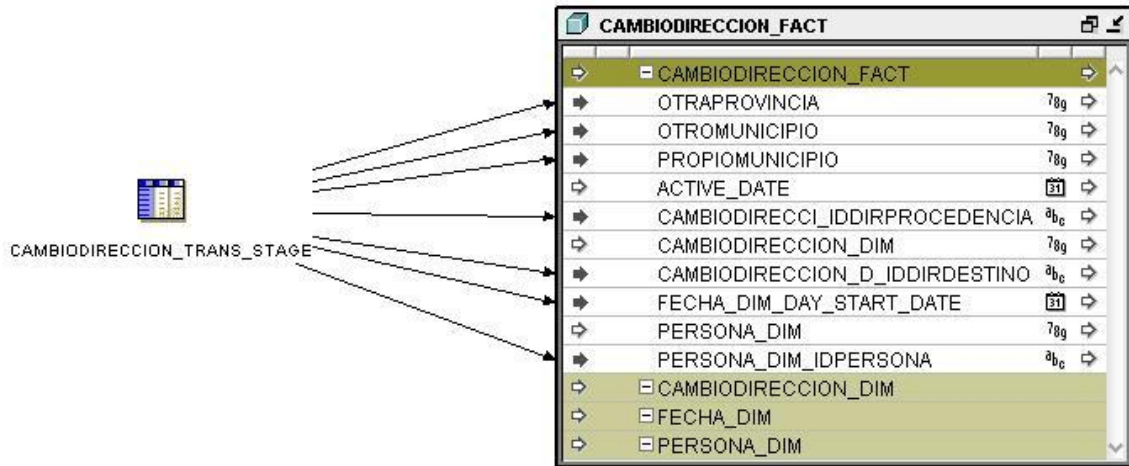




Anexo 10: Correspondencia de la dimensión Cambio de Dirección

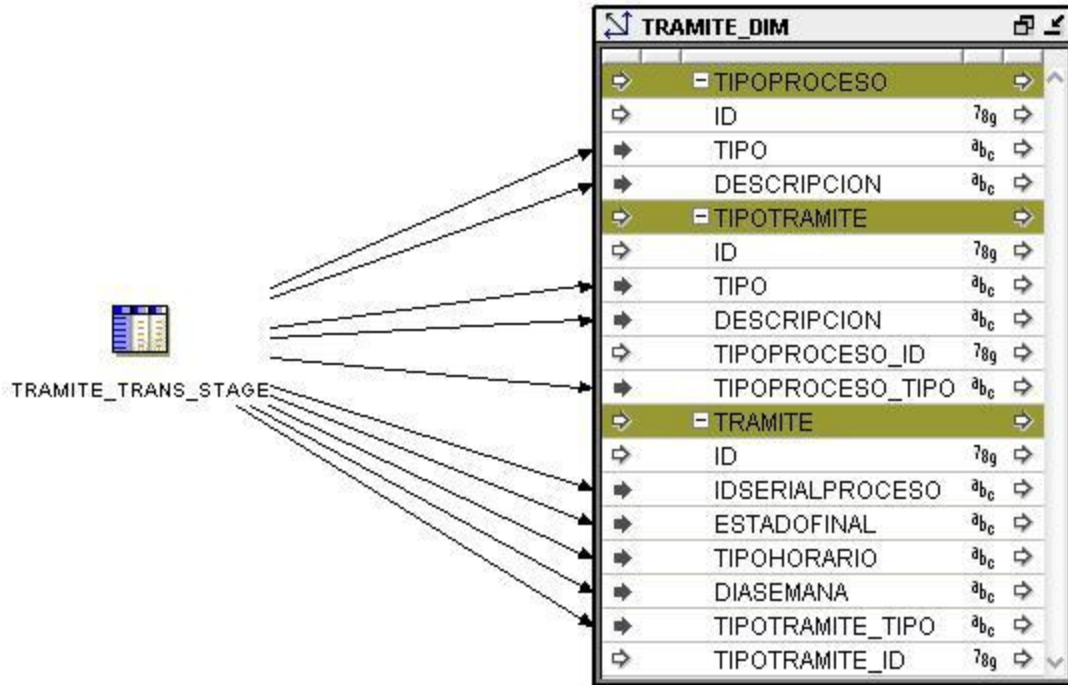


Anexo 11: Correspondencia del cubo Cambio de Dirección



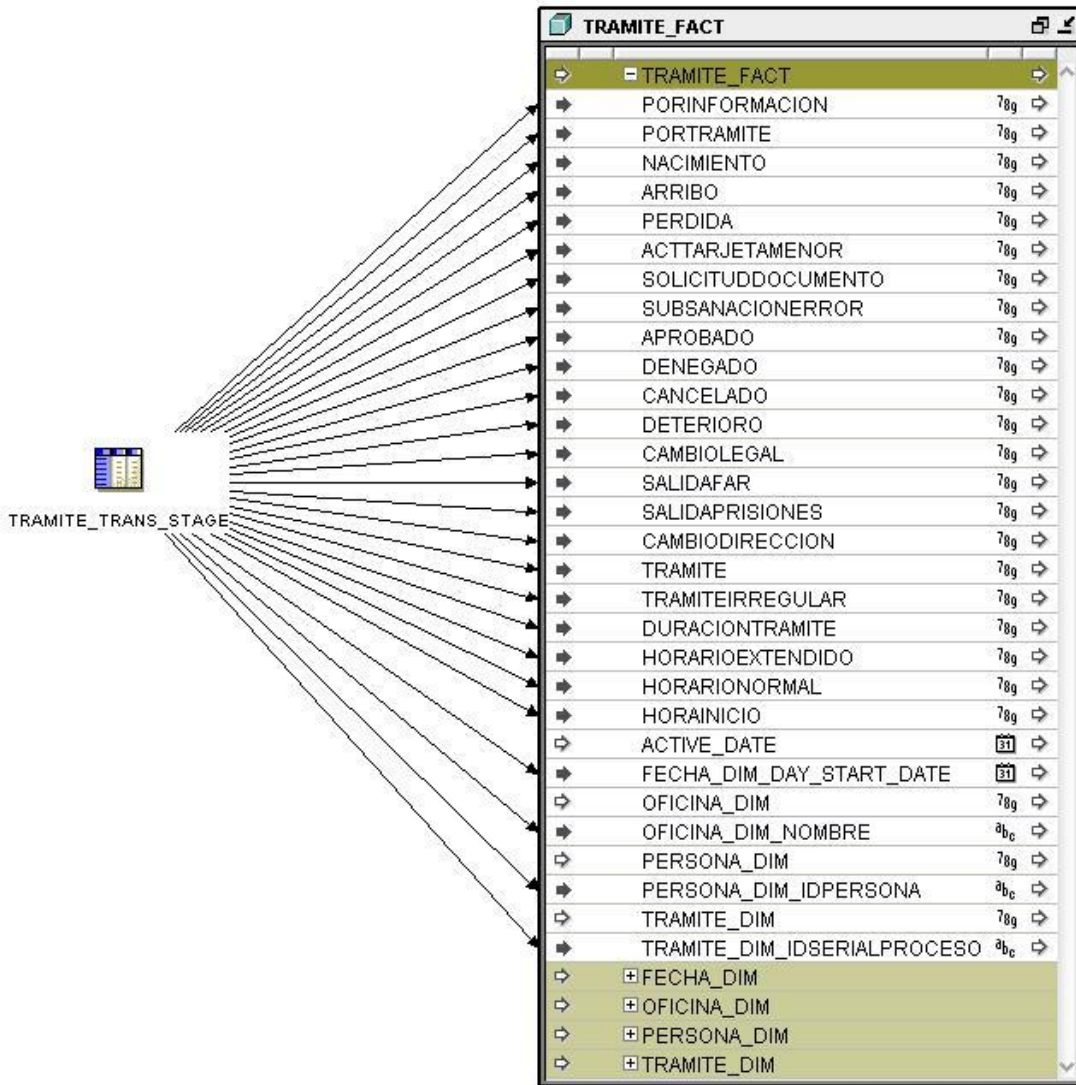


Anexo 12: Correspondencia de la dimensión Trámite



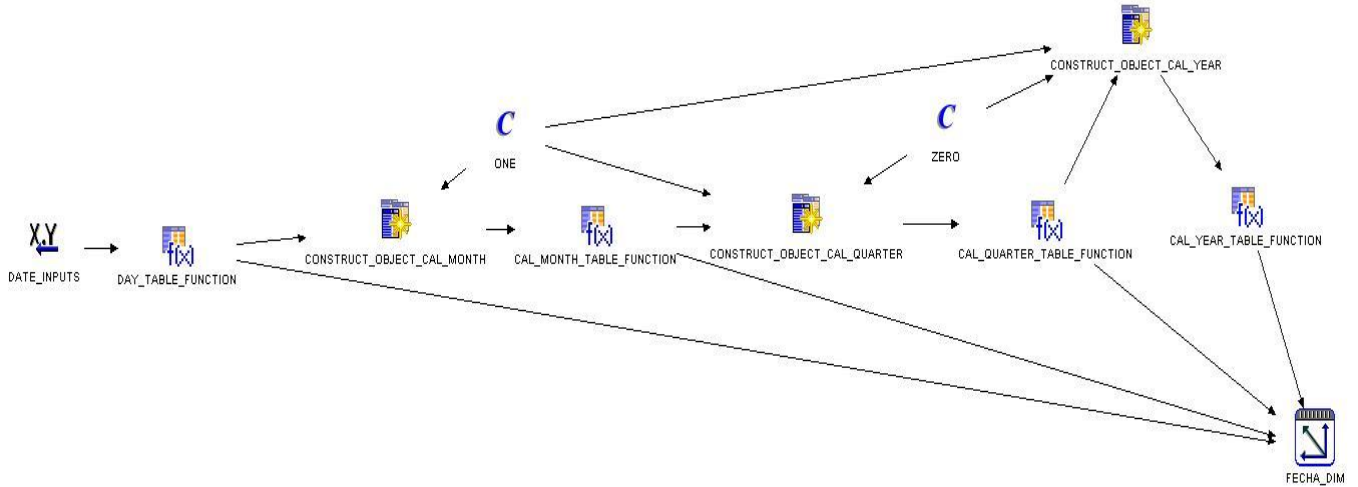


Anexo 13: Correspondencia del cubo Trámite

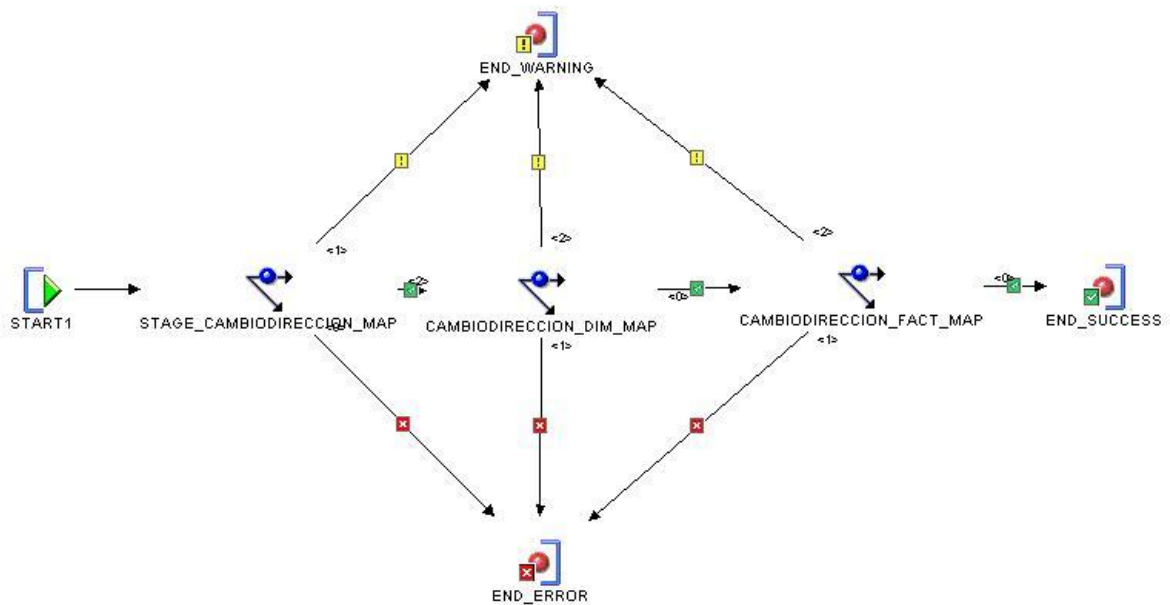




Anexo 14: Correspondencia de la dimensión Fecha

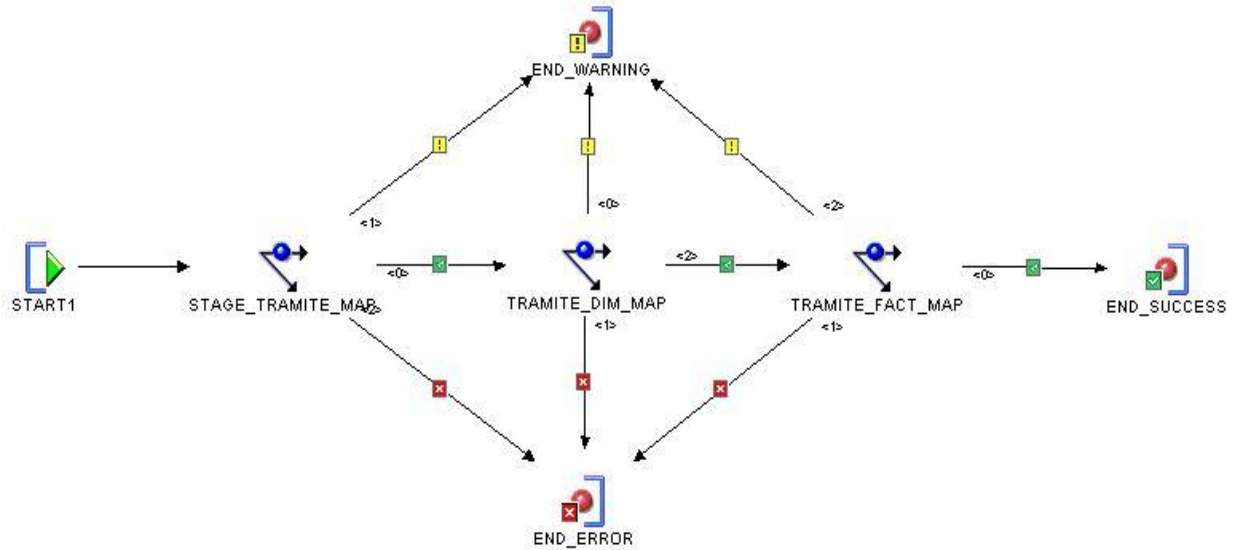


Anexo 15: Flujo del proceso ETL de Cambio de Dirección





Anexo 16: Flujo del proceso ETL de Trámite

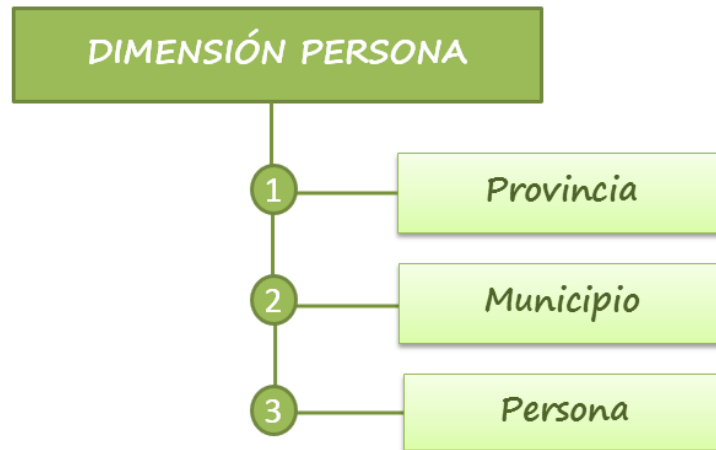


Anexo 17: Valores históricos que guarda el Almacén de Datos

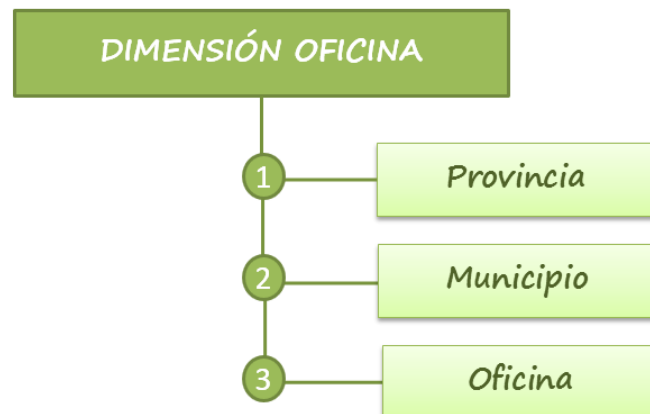
CAMPO	TABLA
DIRECCION	PERSONADIM
NUMERO_IDENTIDAD	PERSONADIM
NOMBRE_MADRE	PERSONADIM
NOMBRE_PADRE	PERSONADIM
PRIMER_NOMBRE	PERSONADIM
SEGUNDO_NOMBRE	PERSONADIM
PRIMER_APELLIDO	PERSONADIM
SEGUNDO_APELLIDO	PERSONADIM



Anexo 18: Jerarquías de la dimensión persona



Anexo 19: Jerarquías de la dimensión oficina



Anexo 20: Jerarquías de la dimensión trámite

