Universidad de las Ciencias Informáticas

Facultad 2

Técnicas de Extracción, Transformación y Carga de Datos del Sistema de Información Nacional de Seguridad Ciudadana en la República Bolivariana de Venezuela

Trabajo de Diploma

Presentado para optar por el título de

Ingeniero en Ciencias Informáticas

Autor: Doris Medina Mustelier

Tutores: Lic. Iván M. Cárdenas Tandrón

Co-Tutores Ing. Yanet Peña Vázquez

Lic. Lianne Guillén Pérez

"Año del 50 Aniversario del Triunfo de la Revolución"

Ciudad de la Habana, Cuba. Marzo de 2009.

Dedicatoria

…….A mis padres y mis familiares

Agradecimientos

A mi mamá por ser la luz que guía mi camino.

A mi papá por ser mi apoyo incondicional.

A mi hermanito por ser él más lindo del mundo y él que yo más quiero.

A mi papito por su amor y comprensión.

A mi abuela por haber contribuido a mi educación, y ser algo imprescindible para mí.

A mis tíos que son como mis hermanos.

A mis profesores, Lianne, Mialín, Yurelkis, Yadira por apoyarme.

A mi tutor por ser compañero y amigo.

A mis amistades tanto de la escuela como fuera de ella.

A todos aquellos que contribuyeron a la realización de este trabajo.

Resumen

La seguridad es una premisa necesaria para el funcionamiento de la sociedad y uno de los principales criterios para asegurar la calidad de vida de la población. Constituye, además, un derecho de todo ser humano y es la facultad de toda persona a desenvolverse dentro de una sociedad libre de amenazas que atenten contra su vida, integridad física, psíquica o cultural.

Para sustentar la capacidad de análisis y tratamiento de la información altamente eficiente y validada, orientada a estudiar la naturaleza y causas de la inseguridad, con vistas a disminuir el índice delictivo y de esta forma contribuir a la mejora del sentir ciudadano, surge en la República Bolivariana de Venezuela, la necesidad de relacionar las informaciones contenidas en las distintas fuentes de información empleando herramientas informáticas que propicien su integración, transformación y almacenamiento de datos y de esta forma contribuir a la toma de decisiones en aras de aumentar la seguridad de cada ciudadano. Por tal motivo, es sumamente importante establecer los procesos de Integración de la Información brindada por los diferentes Órganos de Seguridad. La definición y establecimiento de un patrón de integración, es un punto vital para garantizar una solución exitosa ante esta problemática.

Índice

Introducción

La seguridad ciudadana constituye la base principal para el desarrollo de los pueblos. Se considera como una condición necesaria para el funcionamiento de la sociedad y uno de los principales criterios para determinar la calidad de vida. Su presencia se manifiesta tanto en la vida diaria de las personas como en la vida pública en la sociedad, y su ausencia atenta contra la tranquilidad y el bienestar social. Es por ello que el estado está en la obligación de implementar diversas leyes, políticas preventivas y estrategias en aras de lograr seguridad de sus pueblos.

La falta de seguridad ciudadana ha pasado a ser uno de los temas centrales de preocupación de los ciudadanos, y por lo tanto, una de las cuestiones a resolver por los responsables políticos de principios de este siglo. La crisis de la seguridad y su correlativa necesidad de reforma ha propiciado que en ocasiones, desde el entorno de gobierno se formulen leyes y políticas simplistas que actúan sobre los síntomas y no sobre las causas que lo generan. Para solucionar este problema es necesario integrar, procesar y finalmente analizar información de todos los ámbitos sociales que propicien la formulación de estrategias, políticas y acciones encaminadas a factores claves de la sociedad previniendo así acciones que generen inseguridad en la sociedad.

La República Bolivariana de Venezuela no deja de tener este problema social, sin embargo el gobierno tiene como sus principales premisas la seguridad de los ciudadanos, esto se manifiesta claramente en la Constitución que se estableció en 1999, la cual plantea que: "*Toda persona tiene derecho a la protección del estado, a través de los Órganos de Seguridad Ciudadana regulados por Ley, frente a situaciones que constituyan amenazas, vulnerabilidad o riesgos para la integridad física de las personas, sus propiedades, el disfrute de sus derechos y el cumplimiento de sus deberes*". (Oficial, 2002)

El Ministerio del Poder Popular para Relaciones Interiores y Justicia (MPPRIJ) de la República Bolivariana de Venezuela, en aras de enfrentar los problemas de Inseguridad Ciudadana y disminuir los altos índices de delincuencia que presenta el país, creó el Centro de Tratamiento y Análisis de Información de Seguridad Ciudadana (CTAISC), formado por un conjunto amplio de Analistas y Especialistas en temas de Seguridad Ciudadana, que se encargan de analizar todos los datos que brindan los diferente Órganos de Seguridad, así como proponer medidas al gobierno, tras el estudio de los posibles factores que dan lugar

a la inseguridad de la población. Los principales problemas que permitieron el surgimiento del CTAISC fueron los siguientes:

1. Ausencia, a nivel nacional, de un Sistema de Información Integrado sobre los órganos de seguridad ciudadana.
2. Dificultad para lograr efectividad en las estrategias y políticas en materia de seguridad; en muchos casos, se debe a que se diseñan en base a hipótesis y suposiciones por carencia de información fidedigna.
3. Limitaciones para el acceso a la información de las bases de datos de los organismos de seguridad.
4. Ausencias de mecanismos centralizados que permitan dar seguimiento y control a políticas y estrategias en relación con hechos y situaciones extraordinarias o relevantes.
5. Expectativas que tiene la ciudadanía en todas las transformaciones anunciadas por el Estado en materia de seguridad ciudadana.

Para sustentar la capacidad de análisis y tratamiento de la información altamente eficiente y validada, orientada a estudiar la naturaleza y causas de la inseguridad, con vistas a disminuir el índice delictivo y de esta forma contribuir a la mejora del sentir ciudadano en este sentido, el CTAISC creó un Sistema de Información de Seguridad Ciudadana (SINSEC). A partir de este momento los datos manejados por los órganos de seguridad seleccionados, pasan a ser información disponible para el Análisis y la Toma de Decisiones en manos de analistas, expertos y altos ejecutivos del MPPRIJ mediante el uso de tecnologías de punta.

Para que este sistema tenga sentido se necesita integrar los datos disponibles en los diferentes órganos de seguridad en un almacén de datos único y centralizado; sin embargo, este proceso requiere el uso de técnicas avanzadas de integración de datos; la complejidad del problema aumenta debido a que cada fuente posee diferentes arquitecturas y conceptualizaciones de los parámetros, variables y valores con los que interactúa. Por lo que surge el siguiente **problema científico**.

¿Cómo integrar los datos de los Órganos de Seguridad Ciudadana de la República Bolivariana de Venezuela en un Almacén de Datos centralizado para el Tratamiento y Análisis de la Información?

El **objeto de investigación** son las técnicas de integración de datos, centrando su **campo de acción** en las técnicas de extracción, transformación y carga de datos.

El **objetivo general** de este trabajo se centra en aplicar técnicas de integración de datos, utilizando las herramientas existentes para integrar los principales Órganos de Seguridad con que cuenta el Gobierno de la República Bolivariano de Venezuela. Se tienen como **objetivos específicos**:

1. Determinar las necesidades de integración del Almacén de Datos del SINSEC.
2. Determinar las condiciones de integración de los órganos de seguridad en la República Bolivariana de Venezuela.
3. Migrar los datos de los órganos de seguridad seleccionados aplicando las técnicas de integración de datos

Para cumplir con los objetivos y resolver la situación problémica planteada, se proponen las siguientes **tareas de la investigación**:

1. Identificar y realizar una propuesta de los principales Órganos de Seguridad según la necesidad de información.
2. Realizar un análisis del dominio informativo de los Órganos de Seguridad, definiendo un lenguaje común para el proceso de transformación del almacén de datos.
3. Realizar un análisis de las herramientas y las técnicas de integración de datos y seleccionar la apropiada dadas las características y los requisitos que se tienen.

Este trabajo consta de distintos capítulos, que reflejan las decisiones tomadas para la solución del problema, quedando estructurado de la siguiente manera:

Capítulo 1 Técnicas de Integración de Datos de Seguridad Ciudadana: contiene todo lo referente a los aspectos teóricos que soportan este trabajo, además hace referencia a algunas de las herramientas más utilizadas en el mundo que permiten el desarrollo del sistema y las metodologías más utilizadas.

Capítulo 2 Técnicas de Extracción, Transformación y Carga de Datos: describe las características fundamentales que detallan el diseño de un proceso de integración de datos (Extracción, Transformación y Carga de Datos).

Capítulo 3 Descripción de la Solución: se define la descripción de la solución con el uso de las herramientas propuestas para la aplicación de las técnicas de Integración de Datos como base para el funcionamiento del SINSEC.

Capítulo 4 Resultados de la Investigación: define los resultados obtenidos luego de la puesta en práctica de la solución propuesta para el proceso de integración de datos del SINSEC.

Capítulo 1: Técnicas de Integración de Datos de Seguridad Ciudadana

1.1 Introducción

En el presente capítulo se hará una valoración de la realidad de los Órganos de Seguridad del Estado Venezolano, la creación del Centro de Tratamiento y Análisis de Información de Seguridad Ciudadana como principal Órgano para proponer el camino para luchar contra la inseguridad, se describe y justifica la necesidad para el CTAISC de contar con mecanismos de integración de datos para lograr su reglamento orgánico y objetivos funcionales.

Se abordan además algunos de los conceptos teóricos que fueron necesarios investigar para la concepción de este trabajo, así como la descripción de los principales conceptos asociados al dominio del problema que son necesarios para desarrollar los procesos de integración de datos al almacén de datos; por último se enumeran y analizan las opciones tecnológicas que permiten dar solución a esta problemática.

1.2 Seguridad Ciudadana

Se entiende por Seguridad Ciudadana "*el estado de sosiego, certidumbre y confianza que debe proporcionarse a la población, residente o de tránsito, mediante acciones dirigidas a proteger su integridad física y propiedades*" (ILPES, 1998).

La seguridad ciudadana está muy ligada a lo que comúnmente denominamos orden público, el cual se establece como garantía y límite de libertad. Consiste en que ningún ciudadano puede realizar una acción que sea perjudicial a los demás. Es, también, la facultad que tiene toda persona, natural o jurídica, de desenvolverse cotidianamente libre de amenazas a su vida, libertad, integridad física, psíquica y cultural. Es un derecho humano.

La ciudad de Caracas, junto a Río de Janeiro y Bogotá, es considerada una de las ciudades más violentas del mundo (INE, 2006).

En la República Bolivariana de Venezuela los valores de incidencia delictiva alcanzan altos niveles que han despertado el sentido de alarma por parte de los ciudadanos y el Estado. Por esas razones, la lucha contra la criminalidad es una de las principales prioridades del gobierno venezolano.

1.3 Tratamiento y Análisis de la Información

El análisis y el tratamiento de la información se han convertido en una herramienta indispensable para la toma de decisiones.

Hoy en día, el aumento del volumen y la cantidad de información que se encuentra digitalizada, ha aumentado constituyendo la memoria de toda organización. Esta importante función se puede utilizar además para explicar el comportamiento de la institución o empresa en el pasado, entender el presente y poder predecir el futuro.

El volumen de información que se produce ha crecido tanto, que la eficiencia en el manejo de los factores de la producción depende cada vez más de la manera en que se administra y planifica el ciclo de creación y utilización de la información. La información es una activo, pero su utilización no esta ligada a los procesos que la generan: se produce en un tiempo y lugar distinto al de su utilización y aplicación, por lo que su valor esta determinado por quien la usa, y no por quien la produce. Los gobiernos no están exentos de esta situación.

En el Reporte Mundial del Sector Público de 2003, la Organización de las Naciones Unidas (ONU) indica que uno de los factores clave para el éxito de la *e-gobierno*[1] es que los gobiernos aprendan a administrar información para crear conocimiento. Dicho de otro modo: para que los gobiernos generen valor público deben utilizar eficientemente la información, esto es, información que genere conocimiento.

En este proceso es importante tener en cuenta que la calidad de los datos muchas veces está en diferentes medios de almacenamientos con diferentes estructuras, en otros casos la calidad de los datos no es la óptima, siendo necesario someter a un proceso de extracción de los datos, limpieza y luego almacenarlos en un lugar común donde estarían preparados para la toma de decisiones.

[1] e-gobierno o gobierno electrónico consiste en el uso de las tecnologías de la información y el conocimiento en los procesos internos de gobierno.

Después que lo datos han sido tratados estarán listo para ser analizados por los analistas o ejecutivos de la empresa o institución. El análisis y tratamiento de la información ha permitido la implementación de estrategias, proyectos y acciones dirigidas a factores claves, constituyendo una de las formas de contribuir progresivamente a mejorar el comportamiento de las instituciones, la vida, en fin en toda las áreas de la sociedad.

1.4 Inteligencia Institucional

Cuando trabajamos los términos de análisis y tratamiento de la información en la actualidad se nombra con frecuencia el término de inteligencia de negocio (Business Intelligence) el cual se describe como:

"…Son los procesos, tecnologías, y herramientas que se necesitan para convertir los datos en información, la información en conocimiento, y el conocimiento en planes que impulsan acciones rentables para el negocio. La Inteligencia de Negocios abarca el almacenamiento de datos, herramientas analíticas, y contenido y gestión del conocimiento…" (Armstrong-Smith, 2006).

Actúa como un factor estratégico para una empresa u organización, que no es otra que proporcionar información privilegiada para responder a los problemas de negocio: entrada a nuevos mercados, promociones u ofertas de productos, eliminación de islas de información, control financiero, optimización de costes, planificación de la producción, análisis de perfiles de clientes, rentabilidad de un producto concreto, etc.

Sin embargo, cuando no se pueden medir los resultados de la inversión en mejoras en los procesos de negocios sino en la sociedad, en el aumento de los empleos, en la calle, en las ciudades y pueblos, o dicho de otro modo, en la satisfacción ciudadana y en una mejor calidad de vida para todos, se puede decir que se está hablando de **Inteligencia Institucional.** (Rivera Victoria, 2007)

La información es un arma poderosa para la toma de decisiones, que permite corregir oportunamente posibles error y atacarlos permitiendo que se gane en precisión y pertinencia al advertir en su desarrollo las complejidades y rasgos característicos de las tareas institucionales.

Es importante destacar que las técnicas de integración forman parte tanto de la Inteligencia de negocio como la institucional. Su diferencia radica en la complejidad de los procesos de integración. Cuando

hablamos de una empresa, establecer relaciones entre sus conceptos puede resultar complejo, pero cuando lo hacemos en un organismo gubernamental las fuentes de información que se manejan se enmarcan en diferentes ámbitos sociales.

Algunas de estas herramientas, tecnologías y actividades se realizan dentro de la inteligencia institucional.

- ✓ Data Warehousing y Data Marts (almacenamiento de datos).
- ✓ Data Mining (herramientas para minería de datos).
- ✓ OLAP (herramientas de procesamiento analítico de datos).
- ✓ Herramientas de consulta y reporte de datos.
- ✓ Herramientas de producción de reportes personalizados.

1.5 Tratamiento y Análisis de la Información en República Bolivariana de Venezuela

En la República Bolivariana de Venezuela, actualmente se proyectan e implementan un conjunto de medidas encaminadas a disminuir el delito, a elevar la percepción de seguridad en la población y a fortalecer la capacidad para mejorar la eficiencia de los órganos con competencia en materia de seguridad ciudadana.

Uno de los principales problemas con que cuentan las instituciones estadales es que se encuentra limitada su capacidad informativa y analítica para proyectar estrategias integrales de prevención e intervención en materia de seguridad ciudadana.

En estos momentos el Gobierno cuenta con información necesaria para toma de decisiones, sin embargo no tiene un medio que le permita tratar y analizar la información. Todos los órganos de seguridad tienen su propio dominio informativo y no está centralizado. Además que todos manejan diferentes conceptos de la información lo que no permite la toma de medidas encaminadas a mejorar la seguridad ciudadana.

1.6 Centro de Tratamiento y Análisis de la Información de Seguridad Ciudadana

El CTAISC es una Organización de alto nivel del Sector Público venezolano, especializada en el análisis de información para las políticas públicas estadales. Este centro tiene varios especialistas y analistas de temas de seguridad en la población. Su creación surge por la necesidad de un centro que unificara la información de todos los órganos de seguridad en aras de lograr la efectividad en las estrategias y políticas de seguridad, hasta el momento las tácticas de seguridad estaban basadas en hipótesis y no en la realidad.

Para analizar la información de interés para el país, los analistas del CTAISC realizan la grabación en tiempo real de 12 canales de televisión, 10 estaciones de radio, el análisis de los medios de prensa en Internet y prensa escrita del país.

Asimismo realiza el enlace, seguimiento, evaluación y cierre del cumplimiento de los órganos de seguridad sobre las decisiones y políticas emanadas de los entes decidores superiores en materia de seguridad ciudadana con la finalidad de fortalecer la efectividad de la toma de decisiones para la disminución de la criminalidad y violencia a nivel nacional.

Este centro tiene como misión:

1. Garantizar la integración y el análisis multidisciplinario de la información generada por las diferentes bases de datos de los órganos de seguridad ciudadana del país, para la generación de políticas y estrategias, mejorando los niveles de seguridad ciudadana.

2. Brindar seguimiento a hechos y situaciones extraordinarias, para ello cuenta con tecnología de punta y un talento humano caracterizado por su honestidad, profesionalidad, responsabilidad, y sentido de cooperación.

1.7 Sistema de Información Nacional de Seguridad Ciudadana

El Sistema de Información Nacional de Seguridad Ciudadana de la República Bolivariana de Venezuela, es un sistema que contiene el Almacén de Datos de Seguridad Ciudadana donde se encuentran

disponibles a los analistas (usuarios), los datos operacionales de los diferentes Órganos de Seguridad u otros Organismos para el análisis y toma de decisiones. Este sistema se definió de la siguiente manera:

"El SINSEC constituye una herramienta que garantiza la integración y el análisis multidisciplinario de la información generada por los diferentes órganos de seguridad ciudadana del país para elevar la efectividad de las estrategias y políticas diseñadas y así mejorar los niveles de seguridad ciudadana." (ALBET.S.A., 2008)

La creación del Sistema de Información Nacional de Seguridad Ciudadana o SINSEC, sustenta la capacidad de análisis y tratamiento de la información altamente eficiente y validada, orientada a estudiar la naturaleza y causas de la inseguridad, con vistas a disminuir el índice delictivo y de esta forma contribuir de forma objetiva a la mejora del sentir ciudadano en este sentido.

Este sistema es desarrollado con un *enfoque iterativo incremental*, orientado a la asimilación paulatina de la información por parte de los analistas y del equipo de soporte del MPPRIJ, así como en el aumento del Dominio Informativo del propio almacén. Dicho modelo anuncia el adelanto continuo de la organización en cuanto a la capacidad de análisis y a la calidad de la gestión.

A partir de este momento la Información de Gestión de las Orígenes de Información seleccionadas, pasa a ser Información disponible para el Análisis y la Toma de Decisiones en manos de analistas, expertos y altos ejecutivos del MPPRIJ mediante el uso de Tecnología de Punta.

Contiene el Almacén de Datos de Seguridad Ciudadana donde se encuentran disponibles a los analistas (usuarios finales), los datos operacionales de los diferentes Órganos de Seguridad u otros Organismos para el análisis y toma de decisiones. En el primer Dominio Informativo, se cuenta con los datos que brinda el Cuerpo de Investigaciones Científicas, Penales y Criminalísticas (CICPC), el Instituto Nacional de Estadística (INE), y la Coordinación Nacional de Ciencias Forenses (CNCF).

1.7.1 Almacén de datos

Cuando se habla de almacenes de datos existen 2 figuras que son muy conocidas por sus aportes en esta área, ellos son Bill Inmon y Ralph Kimball.

Según Bill Inmon define un almacén de datos en términos de las características del repositorio de datos:"Un almacén de datos es una colección de datos orientados por temas, integrados, variables en el tiempo y no volátil para el apoyo de la toma de decisiones" (William H., 2005)

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

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

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

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

Inmon defiende una metodología descendente (top-down) a la hora de diseñar un almacén de datos, ya que de esta forma se considerarán mejor todos los datos corporativos. En esta metodología los *Data Marts*[2] se crearán después de haber terminado el almacén de datos completo de la organización.

Sin embargo Ralph Kimball conocido autor en el tema de los almacenes de datos, define un almacén de datos como: "una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis". También fue Kimball quien determinó que un almacén de datos no era más que: "la unión de todos los Data Marts de una entidad". Defiende por tanto una metodología ascendente (bottom-up) a la hora de diseñar un almacén de datos (Kimball & Ross, 2002).

Las definiciones anteriores se centran en los datos en sí. Sin embargo, los medios para obtener y analizar esos datos, para extraerlos, transformarlos y cargarlos, así como las diferentes formas para realizar la gestión de datos son componentes esenciales de un almacén de datos. Muchas referencias a un almacén de datos utilizan esta definición más amplia. Por lo tanto, en esta definición se incluyen herramientas para

[2] Son subconjuntos de datos con el propósito de ayudar a que un área específica dentro del negocio, se pueda tomar mejores decisiones.

la inteligencia empresarial, herramientas para extraer, transformar y cargar datos en el almacén de datos, y herramientas para gestionar y recuperar los metadatos.

1.7.2 Metodologías de desarrollo de almacenes de datos

Cuando se hace referencia a metodologías de desarrollo de Almacenes de Datos no podemos dejar de hablar de Inmon y Kimball que se consideran padres en estos temas .Ambos desarrollaron metodologías que funcionan bien en diferentes tipos de almacenes de datos. Sin embargo, hay situaciones en las que Kimball aporta un enfoque rápido y eficaz de almacenamiento de datos y hay otros en los que el enfoque Inmon conduce a una solución más limpia. Ambas metodologías tienen ventajas y desventajas.

La metodología Kimball de dividir el mundo de *Inteligencia de Negocio* entre el hecho y las dimensiones es muy eficaz y conduce a una solución completa en una cantidad muy pequeña de tiempo. La técnica de Kimball tiene una gran cantidad de documentación y se puede encontrar respuestas a muchas preguntas. Se puede empezar de cero y dar al usuario una primera información sobre sus datos en cuestión de días. Después de tales prototipos, se inicia la solución normal del ciclo de vida de inteligencia de negocio.

Los Data Marts resultantes son muy fáciles de consultar para ambos desarrolladores y herramientas de usuario final. Las relaciones directas entre los hechos y las dimensiones de conceder a nadie la capacidad de construir preguntas muy simples, la mayor parte del tiempo sin echar un vistazo a la documentación de metadatos. Esta es, sin duda, para corregir los esquemas estrellas simples. Al empezar a usar copo de nieve, puente de tablas y / o otras técnicas avanzadas la simplicidad es algo perdido y la complejidad de las preguntas comienza a subir. Sin embargo, siempre es mucho más fácil entonces consultar el original de bases de datos OLTP[3]. La metodología de Kimball es ideal para los primeros pasos de la introducción de inteligencia de negocios a un cliente.

La visión Inmon de un almacén de datos es muy diferente a la de Kimball. La estructura Inmon se basa en un complejo empresarial de bases de datos relacionales, mientras que Kimball se basa en un hecho de dimensiones.

[3] OLAP es el acrónimo en inglés de Procesamiento Analítico en Línea (**O**n-**L**ine **A**nalytical **P**rocessing). Es una solución utilizada en el campo de Inteligencia de Negocios, cuyo objetivo es agilizar la consulta de grandes cantidades de datos

Ambos métodos pueden coexistir felizmente en un solo almacén de datos. La solución Kimball es muy rápida de lograr. Podemos empezar con un puro almacén de datos Kimball, transferir los datos directamente desde el OLTP a los hechos y las dimensiones de los Data Marts (Russo, 2008).

1.8 Técnicas de Integración de Datos

El proceso de integración de datos esta sustentado por la necesidad de reunir los datos de diferentes fuentes. Siendo un problema la heterogeneidad de los datos de cada uno de las fuentes, convirtiéndose este proceso de unificación de datos un trabajo muy costoso y complejo.

Pero más allá de integrar datos disímiles en cuanto a la mezcla dato-a-dato de diferentes "entidades", de forma tal que la integración tenga sentido, y puedan obtenerse resultados comparables y compatibles esta la calidad de esos datos. Es por eso que las Técnicas y Tecnologías de Integración tienen tanto fortalezas como debilidades. Existen límites claros entre los diferentes Proyectos de Integración más apropiados para cada tecnología.

La integración se puede enfocar de varias formas diferentes dependiendo de la idea de "Integración" que se tenga. Fundamentalmente existen cuatro:

I. Replicación de Datos

II. EAI [4] (Morgenthal, 2000)

III. EII [5] (Morgenthal, 2005)

IV. ETL[6] (Kimball & Caserta, 2004)

[4] Del ingles Enterprise Application Integration

5 Del ingles Enterprise Information Integration

6 Del ingles Extract, Transform and Load

1.8.1 Replicación de Datos

La replicación de datos no es más que el transporte de datos entre dos o más servidores, permitiendo que ciertos datos de la base de datos estén almacenados en más de un sitio, y así aumentar la disponibilidad de los datos y mejorar el rendimiento de las consultas globales.

Entre las principales características de esta tecnología se puede decir que utiliza un mecanismo de Integración de Bases de datos a Bases de Datos, basado en Tablas lo cual es una desventaja para la resolución del problema a resolver dada la heterogeneidad de las fuentes de datos. Además es un mecanismo de Baja complejidad, bajo costo, se puede decir que es Unidireccional, es una forma simplificada de ETL (Ver más adelante) entre otras características.

1.8.2 Integración de Aplicaciones Empresariales (EAI)

Otra de las tecnologías de integración es Enterprise Application Integration (EAI) o Integración de Aplicaciones de Empresa que se define como el uso de software y principios de arquitectura de sistemas para integrar un conjunto de aplicaciones.

Las soluciones **EAI** habilitan la automatización de procesos de negocio end-to-end por medio de la coordinación de secuencias de tareas y recursos (ambos sistemas y personas) que las ejecutan. Estas soluciones deben estar preparadas para la modificación dinámica de procesos, incluso estando en marcha. Desarrolla una “Vista Unificada” de la Organización y sus aplicaciones, buscando siempre la forma en que las aplicaciones existentes acoplan dentro de la nueva vista desarrollada. Las clave para elevar la eficiencia operacional y los valores individuales de cada uno de estos sistemas es asegurando que ellos puedan comunicarse e interactuar.

Se necesita dotar al negocio de un acceso a la información completo y transparente, y habilitar movimientos sin fisuras de la Información de una aplicación a otra. Con vistas a esto EAI, se encarga de aliviar estos problemas, así como crear un nuevo paradigma y obtener realmente organizaciones proactivas.

Las características más importantes de esta tecnología es que se utiliza para la integración de Aplicaciones – a– Aplicaciones. Además esta orientado a mensajes, no a la integración de datos en lotes y abundantes transformaciones lo cual es un desventaja significativa a la hora de realizar un problema

donde se involucra el tratamiento y el análisis de datos. Además su principal utilización es para soporte a BPM[7].

1.8.3 Integración de Información Empresarial (EII)

Entre las tecnologías de integración también se encuentra EII o Integración de Información Empresariales que es un mecanismo de transformación y acceso a datos transparente y optimizado para suministrar una única interfaz a lo largo de los datos de las organizaciones.

No se basa en integrar bases de datos en si mismas (Morgenthal, 2005; C.Imhoff, 2005) ya que la información se mantiene en las fuentes de información. En lugar de eso, se desarrolla una Interfaz Programática para el Acceso a los Datos que permita recuperar los datos. Usualmente el resultado de este método es un Sistema de Información Heterogéneo Distribuido, Virtualmente Integrado. Por lo general este tipo de solución consiste en crear un Broker[8] de tal forma que contengan directorios de Bases de Datos y que a su vez pueda dar, sirva de canal de consulta y representación de la información recuperada. EII es el mejor método para la Integración de Datos a Tiempo Real. La edad relativa de los datos depende de las fuentes de datos ya que la información es capturada en tiempo real de las fuentes de datos, esto implica que las fuentes tienen que tener una estructura tecnológica sólida y bien establecida.

1.8.4 Extracción, Transformación y Carga de Datos (ETL)

Por otra parte está las ETL (Extract, Transform and Load) es la tecnología enfocada a la Integración de Datos, tanto por lote, como a tiempo real hacia almacenes de datos.

[7] Administración a el proceso de negocio (Del ingles **B**usiness **P**rocess **M**anagement)

[8] Mediador o Intermediario

Las ETL proporcionan consolidación de datos para la construcción de bases de datos permanentes utilizadas para el análisis o la generación de informes, la federación de datos para la creación de dashboards[9], y la propagación de datos para la transferencia de datos entre aplicaciones.

Estos tres procesos se combinan en una herramienta para extraer datos de bases de datos fuentes, archivos u otro sistema, y colocarlas en bases de datos destino. ETL se utiliza para migrar datos de una o más bases de datos a terceros, para formar repositorios de datos, Data Marts, almacenes de datos y también para convertir bases de datos de un tipo o formato a otro. ETL se utiliza además para sincronizar datos desde diversas aplicaciones e involucra procesos de manipulación de datos que van más allá de un simple movimiento desde el punto A hasta el punto B.

Extracción - el proceso de lectura de datos desde los sistemas fuentes. Los datos pueden ser extraídos de las siguientes maneras

- **Schedule-Driven Pull Mode** (ETL por lote).
- **Event-driven Push Mode** (Proceso Online que propaga las modificaciones de los datos hacia las bases de datos destinos).

Transformación - el proceso de conversión de los datos extraídos de su forma actual en el formato que debe ser, en la que se pueden colocar en otros sistemas o bases de datos. La transformación se produce mediante el uso de normas o tablas o a través de la combinación de los datos con otros datos.

Carga - el proceso de creación y ejecución de flujos de trabajo para escribir los datos en los sistemas destinos. La carga de datos puede provocar el refrescado completo de un almacén de datos o puede hacerse mediante la actualización de la base de datos destino. Las Interfaces para ello incluyen las normas como ODBC[10], JBDC[11] o interfaces de aplicación.

[9] Referente a Tablero de Control Digital. Es la aplicación digital del Cuadro de Mando Integral. Brinda una visión general del comportamiento de la institución.

[10] **O**pen **D**ata**b**ase **C**onnectivity (ODBC) es un estándar de acceso a Bases de Datos utilizado para acceder a cualquier dato desde cualquier aplicación, sin importar qué Sistema Gestor de Bases de Datos se este utilizando.

Esta tecnología es usada cuando las fuentes de datos que se quieren integrar trabajan con diferentes nomenclaturas de los valores con los que interactúan, por lo que el proceso de integración implica aplicar reglas de transformación para depurar los datos y almacenarlos de forma que la información que se genere tenga coherencia, sea verídica y no pierda su valor informacional.

Para lograr los objetivos del CTAISC, no es necesaria la obtención de la información a tiempo real, como si lo necesitaría un centro militar, un centro de emergencia, un hospital, etc., sino que la integración de los nuevos registros de datos provenientes de los Órganos de Seguridad pueden ser adicionados al Almacén de Datos en forma de bloque de datos, incluso con una latencia. Para realizar esta integración se necesitan extraer de los diferentes órganos de seguridad, los datos significativos que puedan aportar información valiosa para los analistas del CTAISC, luego transformar estos datos en un lenguaje común para finalmente ser cargados al almacén de datos.

1.9 Protocolos de comunicación

Los protocolos de comunicaciones definen las normas que posibilitan que se establezca una comunicación entre varios equipos o dispositivos, ya que estos equipos pueden ser diferentes entre sí, es decir es un conjunto de reglas y procedimientos que deben respetarse para el envío y la recepción de datos a través de una red. Existen diversos protocolos de acuerdo a cómo se espera que sea la comunicación. Algunos protocolos, por ejemplo, se especializarán en el intercambio de archivos (FTP).

Pueden estar implementados bien en hardware (tarjetas de red), software (drivers), o una combinación de ambos. Los protocolos implantados en sistemas de comunicación con un amplio impacto, suelen convertirse en estándares, debido a que la comunicación e intercambio de información (datos) es un factor fundamental en numerosos sistemas.

[11] **J**ava **D**ata**b**ase **C**onnectivity (JDBC), es una API que permite la ejecución de operaciones sobre bases de datos, independientemente del sistema operativo donde se ejecute o de la base de datos a la cual se accede.

En el campo de las redes informáticas, los protocolos se pueden dividir en varias categorías, una de las clasificaciones más estudiadas es la OSI[12].

Protocolos orientados a conexión: estos protocolos controlan la transmisión de datos durante una comunicación establecida entre dos máquinas. En tal esquema, el equipo receptor envía acuses de recepción durante la comunicación, por lo cual el equipo remitente es responsable de la validez de los datos que está enviando. Los datos se envían entonces como flujo de datos. TCP[13] es un protocolo orientado a conexión;

Protocolos no orientados a conexión: éste es un método de comunicación en el cual el equipo remitente envía datos sin avisarle al equipo receptor, y éste recibe los datos sin enviar una notificación de recepción al remitente. Los datos se envían entonces como bloques (datagramas). UDP[14] es un protocolo no orientado a conexión.

Según la clasificación OSI, la comunicación de varios dispositivos ETD[15] se puede estudiar dividiéndola en 7 niveles, que son expuestos desde su nivel más alto hasta el más bajo.

Las 4 capas superiores trabajan con problemas particulares a las aplicaciones, y las 3 capas inferiores se encargan de los problemas pertinentes al transporte de los datos.

[12] El modelo de referencia de Interconexión de Sistemas Abiertos (OSI, Open System Interconnection) modelo de red descriptivo creado por ISO (Organización Internacional para la Estandarización).

[13] **T**ransmission-**C**ontrol-**P**rotocol (TCP) en español Protocolo de Control de Transmisión.

[14] **U**ser **D**atagram **P**rotocol (UDP) es un protocolo del nivel de transporte basado en el intercambio de datagramas.

[15] ETD es un Equipo Terminal de Datos. Se considera ETD a cualquier equipo informático, sea receptor o emisor final de datos.

Fig. 1 Capas del Modelo OSI

Ofrece a las aplicaciones (de usuario o no) la posibilidad de acceder a los servicios de las demás capas y define los protocolos que utilizan las aplicaciones para intercambiar datos, como correo electrónico (POP y SMTP), gestores de bases de datos y servidor de ficheros (FTP). Hay tantos protocolos como aplicaciones distintas y puesto que continuamente se desarrollan nuevas aplicaciones el número de protocolos crece sin parar.

Es importante aclarar que el usuario normalmente no interactúa directamente con el nivel de aplicación. Suele interactuar con programas que a su vez interactúan con el nivel de aplicación pero ocultando la complejidad subyacente.

De los protocolos de comunicación más usados podemos destacar HTTP el cual es nombrado bajo las siglas WWW, el FTP para transferencia de ficheros, el SMTP para el trabajo con correo electrónico, el TCP/IP entre otros.

1.9.1 HTTP

El protocolo de transferencia de hipertexto (HTTP, HyperText Transfer Protocol) es el protocolo usado en cada transacción de la Web (WWW). HTTP define la sintaxis y la semántica que utilizan los elementos software de la arquitectura web (clientes, servidores, proxies) para comunicarse. Entre las características que brinda el protocolo de http es la utilización de servicios web que puede formar parte de una solución de almacén de datos a la hora de realizar la integración de fuentes de información.

Un servicio web es un conjunto de protocolos y estándares que sirven para intercambiar datos entre aplicaciones. Distintas aplicaciones de software desarrolladas en lenguajes de programación diferentes, y ejecutadas sobre cualquier plataforma, pueden utilizar los servicios web para intercambiar datos en redes de ordenadores como Internet. La interoperabilidad se consigue mediante la adopción de estándares abiertos. Para mejorar la interoperabilidad entre distintas implementaciones de servicios Web se han desarrollado diversos perfiles para definir de manera más exhaustiva estos estándares.

Las ventajas de la utilización de Servicios Webs son las siguientes:

- ✓ Aportan interoperabilidad entre aplicaciones de software independientemente de sus propiedades o de las plataformas sobre las que se instalen.
- ✓ Los servicios Web fomentan los estándares y protocolos basados en texto, que hacen más fácil acceder a su contenido y entender su funcionamiento.
- ✓ Al apoyarse en HTTP, los servicios Web pueden aprovecharse de los sistemas de seguridad firewall sin necesidad de cambiar las reglas de filtrado.
- ✓ Permiten que servicios y software de diferentes compañías ubicadas en diferentes lugares geográficos puedan ser combinados fácilmente para proveer servicios integrados.
- ✓ Permiten la interoperabilidad entre plataformas de distintos fabricantes por medio de protocolos estándar y abiertos. Las especificaciones son gestionadas por una organización abierta, la W3C[16],

[16] Consorcio World Wide Web (mas conocida como W3C)

por tanto no hay secretismos por intereses particulares de fabricantes concretos y se garantiza la plena interoperabilidad entre aplicaciones.

Los servicios webs complementan toda una metodología de modelado y diseño para aplicaciones SOA[17], formadas por servicios de aplicación débilmente acoplados y altamente interoperables. Esta arquitectura tiene muchas bondades por la cual se tiene en cuenta para resolver problemas como que se quiere resolver pero su gran limitación es que debe tenerse una organización muy bien consolidada basada en acuerdos entre Entidades las que deben contar con estructuras y plataformas tecnológicas bien establecidas. Esto es un factor crítico ante la necesidad de selección de este tipo de tecnología.

1.9.2 SMTP

Simple Mail Transfer Protocol (SMTP), o Protocolo Simple de Transferencia de Correo es un protocolo de red basado en texto utilizado para el intercambio de mensajes de correo electrónico entre computadoras u otros dispositivos (PDA's, teléfonos móviles, etc.). Está definido en el RFC 2821 y es un estándar oficial de Internet.

SMTP se basa en el modelo cliente-servidor, donde un cliente envía un mensaje a uno o varios receptores. La comunicación entre el cliente y el servidor consiste enteramente en líneas de texto compuestas por caracteres ASCII.

A pesar de que se definen e implementan extensiones como por ejemplo el SMTP-AUTH para lidiar con las limitaciones del SMTP original y facilitar aspectos como la autenticación a los emisores, entre otras, estas limitaciones continúan existiendo, mayormente con la seguridad, los spams, etc. Además de ello, no es posible la transferencia de grandes bloques de datos por correos, debido a las limitaciones en las cuotas de los buzones en la mayoría de los servidores de SMTP.

1.9.3 FTP

El protocolo de transferencia de archivos FTP (File Transfer Protocol) es para el traspaso de archivos entre sistemas conectados a través de la red, sin importar en absoluto donde están localizados estos ordenadores y si usan o no el mismo sistema operativo: basta con que estén conectados a Internet. Este

[17] **A**rquitectura **O**rientada **a** **S**ervicios (en inglés **S**ervice **O**riented **A**rchitecture).

protocolo esta basado en la arquitectura cliente-servidor. Desde un equipo cliente se puede conectar a un servidor para descargar o enviarles archivos, facilitando la copia o el traslado de ficheros, sin correr ningún tipo de riesgo de perdida de información; y de una manera rápida y a la vez muy sencilla.

Aunque existen otros protocolos que permiten el intercambio de ficheros por la red, el FTP es el servicio especial dedicado a la transferencia de ficheros entre dos ordenadores. El servicio FTP es ofrecido por la capa de aplicación del modelo de capas de red TCP/IP al usuario, utilizando normalmente el puerto de red 20 y el 21.

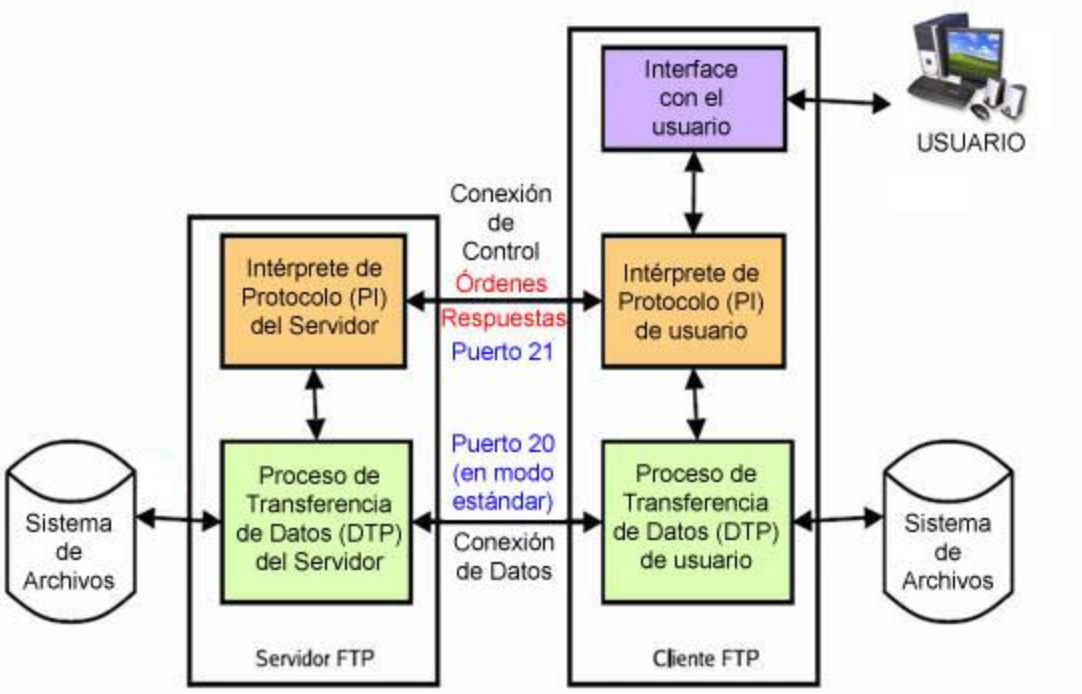

Fig. 2 Diagrama del modelo FTP

Uno de los problemas primordiales del servicio FTP es que está diseñado para brindar la máxima velocidad en la conexión, pero no la máxima seguridad. Esto se debe a que todo el intercambio de información, incluyendo la transferencia al servidor de las credenciales del usuario y de los archivos, se realiza en texto plano sin ningún tipo de cifrado, por lo que cualquier agresor puede capturar este tráfico, acceder al servidor, o apropiarse de los archivos transferidos. Es por ello que son de gran utilidad

aplicaciones como SCP[18] y SFTP[19], incluidas en el paquete SSH[20], que permiten transferir archivos pero cifrando todo el tráfico.

FTP es un protocolo con conexión y con estado: es imprescindible realizar una operación de "entrada" (login) antes de cualquier otra, y se conserva nuestro estado en el servidor. Esto permite aplicar ciertas restricciones, como permisos especiales a ciertos usuarios, cuota de utilización en disco, etc. Para la realización de un almacén de datos eta tecnología se puede utilizar para realizar la integración de datos ya que se puede utilizar para la transferencia de archivos entre las fuentes de datos y el almacén.

1.10 Archivos XML para la Transferencia de Datos

XML, al igual que el SGML, es lo que se conoce como un metalenguaje, o sea un lenguaje (de marcas) capaz de generar otros lenguajes (de marcas) muy simple y flexible. Es un lenguaje con una importante función en el proceso de intercambio, estructuración y envío de datos en la Web. Describe los datos de tal manera que es posible estructurarlos utilizando para ello etiquetas. Originalmente diseñado para satisfacer los desafíos de las publicaciones electrónicas (Web, 2008)

XML es un metalenguaje extensible de etiquetas desarrollado por el World Wide Web Consortium (W3C), que permite definir la gramática de lenguajes específicos (de la misma manera que HTML es a su vez un lenguaje definido por SGML). XML se propone como un estándar para el intercambio de información estructurada entre diferentes plataformas. Se puede usar en bases de datos, editores de texto, hojas de cálculo y casi cualquier cosa imaginable.

XML es una tecnología sencilla que tiene a su alrededor otras que la complementan y la hacen mucho más grande y con unas posibilidades mucho mayores. Tiene un papel muy importante en la actualidad ya que permite la compatibilidad entre sistemas para compartir la información de una manera segura, fiable y fácil.

[18] **S**ecure **C**opy es un medio de transferencia segura de archivos informáticos.

[19] **S**ecure **F**ile **T**ransfer **P**rotocol (Protocolo de Transferencia de Archivos Seguro)

[20] Intérprete de comandos seguro, es el nombre de un protocolo y del programa que lo implementa (en inglés **S**ecure **SH**ell)

Entre las principales ventajas del uso de XML están las siguientes:

- ✓ **Es extensible**: Después de diseñado y puesto en producción, es posible extender XML con la adición de nuevas etiquetas, de modo que se pueda continuar utilizando sin complicación alguna.
- ✓ **El analizador es un componente estándar**, no es necesario crear un analizador específico para cada versión de lenguaje XML. Esto posibilita el empleo de cualquiera de los analizadores disponibles. De esta manera se evitan errores y se acelera el desarrollo de aplicaciones.
- ✓ Si un tercero decide usar un documento creado en XML, es sencillo entender su estructura y procesarla. **Mejora la compatibilidad entre aplicaciones**.

Existen muchos aspectos a considerar en la incorporación de la tecnología XML en una solución, aspectos como la estructura: prólogo, cuerpo, elementos, atributos; el concepto de validez que implica que el documento XML en cuestión esté “bien formado” o “mal formado”. Por las características más notables de esta tecnología explicada anteriormente es muy utilizada para almacenar gran cumulo de información sin importar cuán rápido o diversa esta sea, ni con que frecuencia ésta llegue a cambiar. Estos elementos se tuvieron en cuenta para elegir el uso de este lenguaje para la transferencia de datos.

1.11 Herramientas de ETL

El procesos ETL es uno de los pasos más importantes para garantizar el éxito de un proyecto de Inteligencia de Negocio, ya que para tener el impacto esperado en la organización es necesario que la información del almacén de datos represente la realidad del negocio, sea confiable y este disponible en el momento que los usuarios y la organización la necesitan, por lo que es crítico contar con una herramienta ETL que permita reducir tiempo y costos de desarrollo y el mantenimiento de los procesos existentes. Seleccionar software de integración no es tarea fácil, para ello es imprescindible entender cuáles son sus principales funciones.

1.11.1 Oracle Warehouse Builder

Oracle Warehouse Builder 10g es la herramienta líder del sector en el diseño de los procesos de ETL. Brinda calidad de datos, auditoría de datos, modelado dimensional y relacional totalmente integrado, gestión de todo el ciclo de vida de datos y metadatos de Oracle Database. Oracle Warehouse Builder 10g

ofrece esta funcionalidad en un entorno gráfico y fácil de usar que permite el rápido diseño, implementación y gestión de los proyectos de integración de datos y sistemas de Inteligencia de Negocio.

Incluye un repositorio de metadatos de múltiples usuarios y listo para empresas, capacidades para el modelado de datos y una amplia variedad de técnicas de transformación y extracción. Las particulares principales de Oracle Warehouse Builder (OWB) se incluyen como característica de base de datos sin costo, y tiene adicionalmente tres opciones para requerimientos de integración específicos (Lumpkin, 2007):

Opción Enterprise ETL: Esta opción está específicamente diseñada para aumentar el desempeño y la productividad, e incluye las siguientes características:

- ✓ Opciones avanzadas para la carga de datos
- ✓ Productividad del desarrollador mediante componentes reutilizables
- ✓ Administración de dimensiones lentamente cambiantes
- ✓ Análisis de impacto y linaje completo de datos
- ✓ Soporte de administración avanzada de configuración

Opción de Calidad de Datos: Como parte totalmente integrada del producto, una ventaja muy distinta de las demás herramientas en el mercado, la opción de calidad de datos ofrece soporte para la proliferación de datos, reglas de datos (esencialmente, reglas de negocio) y características del cumplimiento de la información.

Conectores: Los conectores ofrecen acceso optimizado para aplicaciones operacionales líderes. Oracle brinda conectores para Oracle e-Business Suite, Peoplesoft Enterprise, Siebel (CRM) y SAP R/3.

Una ventaja clave de OWB es la gran variedad de funcionalidad que ofrece dentro de una sola herramienta. El modelado de datos, el cumplimiento de datos y la calidad de datos son características centrales que cualquier herramienta para la integración de datos empresariales debe tener. No obstante, una ventaja estructural clave de Oracle Warehouse Builder es la integración de los componentes. Oracle

Warehouse Builder proporciona todas sus capacidades dentro de un repositorio común y una interface de usuarios.

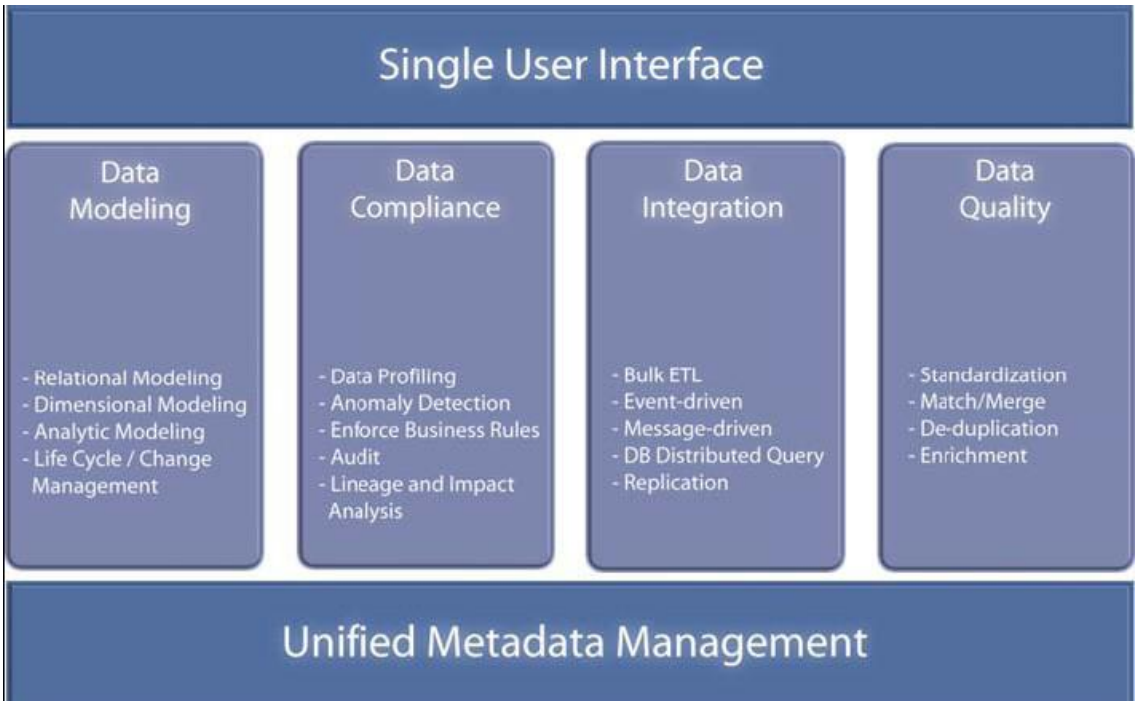

Fig. 3 Integración de funcionalidades de Oracle Warehouse Builder

Al brindar todas estas capacidades en una sola herramienta sobre un repositorio único, OWB resuelve un antiguo desafío de la integración de datos. Muchas soluciones de integración brindan herramientas separadas para estas capacidades diferentes. No obstante, resulta terriblemente ineficiente realizar el modelado de datos en una sola herramienta, y luego el mapeo de ETL en otra herramienta, y luego la proliferación de datos incluso en otra herramienta. OWB brinda un repositorio de metadatos y una interfaz de usuario para todo el proceso de integración.

OWB realiza toda su transformación dentro de una base de datos Oracle aprovechando la escalabilidad y el desempeño de la plataforma de base de datos.

1.11.2 Pentaho Data Integration.

Muchas organizaciones tienen información disponible en aplicaciones y base de datos separados. Pentaho Data Integration abre, limpia e integra esta valiosa información y la pone en manos del usuario. Provee una consistencia, una sola versión de todos los recursos de información, que es uno de los más grandes desafíos para las organizaciones de la informática y las telecomunicaciones hoy en día. Pentaho Data Integration permite una poderosa ETL (Extracción, Transformación y Carga).

El uso de Kettle permite evitar grandes cargas de trabajo manual frecuentemente difícil de mantener y de desplegar.

La arquitectura de Pentaho Data Integration viene representada por la fig.4 (Pentaho, 2008)

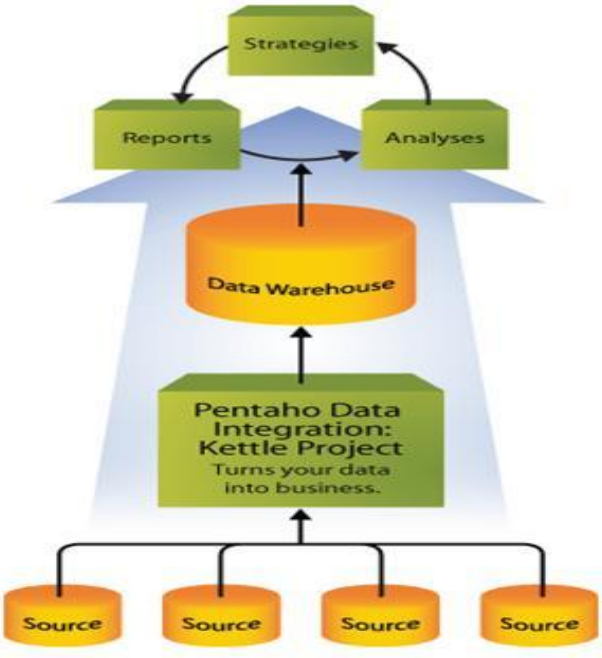

Fig. 4 Arquitectura de Pentaho Data Integration

Propiedades básicas

A parte de ser Software Libre (*Open Source)* y sin costes de licencia, las características básicas de esta herramienta son:

- ✓ Entorno gráfico de desarrollo
- ✓ Uso de tecnologías estándar: Java, XML, Java Script
- ✓ Fácil de instalar y configurar
- ✓ Multiplataforma: Windows, Macintosh, Linux
- ✓ Basado en dos tipos de objetos: Transformaciones (colección de pasos en un proceso ETL) y trabajos (colección de transformaciones)

Incluye cuatro herramientas:

1. Spoon: para diseñar transformaciones ETTL usando el entorno gráfico
2. PAN: para ejecutar transformaciones diseñadas con Spoon
3. CHEF: para crear trabajos
4. Kitchen: para ejecutar trabajos

1.11.3 Data Integration Suite. Sybase

Data Integration Suite proporciona a las organizaciones una aproximación más inteligente y más racional mediante la cual distribuir diferentes tipos de datos a lo largo de toda la organización. Así, se podrá realizar una toma de decisiones más rápida y contando con más información, además de reducir los costes operativos y la complejidad.

El paquete Data Integration Suite ofrece un completo conjunto modular de tecnologías que incorporan una serie de herramientas avanzadas de administración, desarrollo y modelo de datos totalmente integrados, las cuales permiten acelerar el flujo de datos y así poder afrontar los retos de integración complejos (figura.5). Este paquete proporciona las técnicas fundamentales para la integración de datos que ofrecen la consolidación, la distribución y la federación, todo ello a través de un completo conjunto de

componentes tecnológicos. Dando apoyo a esas tecnologías se encuentra una capa proporcionada por WorkSpace, la cual integra el Modelo de datos, el desarrollo y los metadatos.

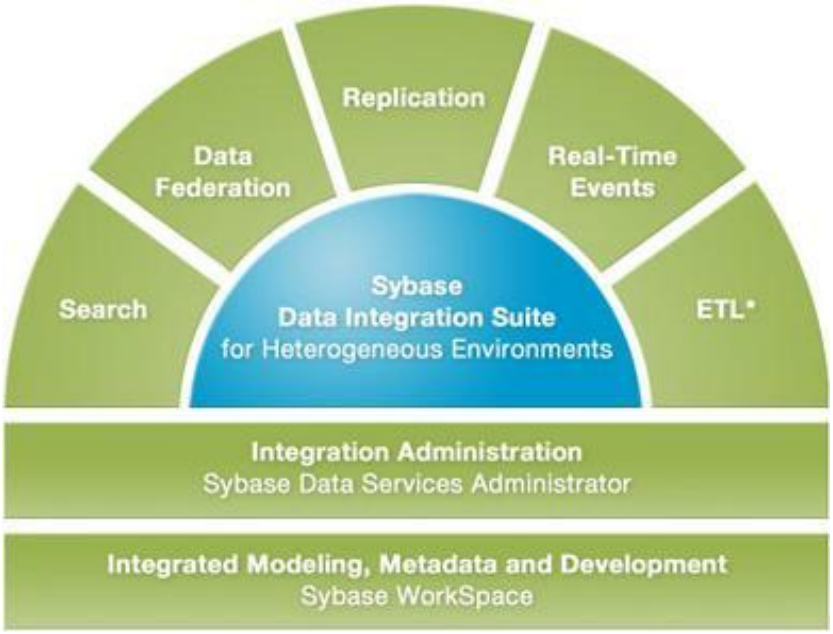

Fig. 5 Paquetes de tecnologías que ofrece Pentaho Data Integration

Las ventajas más significativas de Data Integration Suite son (Sybase, 2008):

Reduce los costes operativos y la complejidad: es posible reducir los costes operativos y la complejidad gracias a la consistencia del modelo de datos y el desarrollo de todas las tecnologías de integración de datos, así como a la posibilidad de reutilizar en gran medida el modelo y los metadatos. Data Integration Suite también puede ejecutarse en plataformas estándar de bajo coste con un alto grado de transparencia: todo un seguro para el futuro.

Acelera el diseño, el desarrollo y la distribución de los flujos de datos permanentes: gracias al modelo de datos, a la administración de los metadatos, a las herramientas de desarrollo, a la administración y a los demás servicios de datos comunes, los arquitectos podrán modelar rápidamente los flujos de datos de extremo a extremo que involucren a múltiples fuentes, comprender la descendencia de la información y el impacto que supondrían los cambios, así como poder realizar modificaciones con mayor control y en menos tiempo.

Completo y flexible: los clientes pueden elegir las técnicas de integración de datos que mejor se ajusten a sus necesidades (o bien una combinación de varias técnicas), entre las que se incluyen ETL, la replicación, la federación o la integración basada en eventos para construir flujos de datos flexibles. Dado que Data Integration Suite es modular, es posible comenzar con los proyectos actuales y después escalar la solución para afrontar los futuros retos que surjan en cuanto a integración de datos, sin importar el tamaño o la complejidad.

1.12 Conclusiones del Capítulo

En este capítulo se han expuesto brevemente conceptos relacionados con la seguridad ciudadana, con el Centro de Tratamiento y Análisis de la Información de Seguridad Ciudadana, la diferencias entre inteligencia de negocio y la inteligencia Institucional destacando su relación con técnicas de integración, se define además que es un Almacén de Datos y los procesos que antevienen.

Se analizó a demás los diferentes métodos de integración, quedando justificada finalmente el uso de **ETL** como técnica idónea para la integración de datos del Almacén de Datos del SINSEC, a fin de dar soporte al tratamiento y análisis de la información de seguridad ciudadana.

Se muestra la importancia del uso de archivos basados en lenguaje **XML** para el intercambio de datos entre los Órganos de Seguridad y el CTAISC, al ser una opción libre del contexto muy apropiadas, dada la heterogeneidad de las fuentes de datos.

Por tales motivos se define como protocolo de comunicación el **FTP**, como solución de integración por lote, basada en el uso de archivos de intercambio de datos en formato XML, como base que garantiza la escalabilidad y la simpleza de toda la solución; desechándose la opción de utilizar una arquitectura SOA ya que el ministerio de no cuenta con tecnología consistente y bien establecida que le permita soportar este tipo de solución.

Mediante el estado del arte de las tecnologías que son más utilizadas para la realización de almacenes de datos, se utilizó como metodología de desarrollo la de Kimball la cual fue seleccionada por su robustez y organización, conduce a una solución completa en una cantidad muy pequeña de tiempo, ya que el MPPRIJ necesitaba lo antes posible el sistema del **SINSEC** funcionando. Además, la metodología genera una gran cantidad de documentación por lo se podía mantener al usuario informado en poco tiempo de los

resultados. Esta metodología esta ampliamente documentada lo que permitía que el equipo de desarrollo (no especializado en el tema) pudiera evacuar todas las dudas que surgieran en el proceso de ETL.

En este capítulo además se caracterizó las herramientas que el mundo se consideran líderes en temas de integración de datos y se decidió utilizar el Oracle Warehouse Builder (**OWB**).Para tomar esta decisión, se tuvo en cuenta que el almacén de datos esta montado en una base de datos Oracle por lo que la herramienta tenía una gran integración con el gestor de base de datos lo que permitía reducir el tiempo y costos de desarrollo, además OWB brinda calidad de datos, auditoría de datos, modelado dimensional y relacional totalmente integrado y gestión de todo el ciclo de vida de datos y metadatos de la base de datos.

Capítulo 2: Técnicas de Extracción, Transformación y Carga de Datos

2.1 Introducción

En este capítulo se realizará una investigación sobre las técnicas de ETL como mecanismo de integración para un almacén de datos. Se exponen cómo desarrollar las diferentes fases del proceso de integración y permite definir los diferentes componentes que se deben incluir en esta solución. Para ello se utilizará como referencia la Metodología propuesta por Kimball.

2.2 Arquitectura de una Solución ETL

Los procesos ETL pueden ser muy complejos. Un sistema ETL mal diseñado puede provocar importantes problemas operativos por lo que para su mejor diseño se debe regir por su arquitectura.

La arquitectura que define este tipo de solución se presenta en la siguiente figura 6. A continuación se describe la arquitectura y se toma de base para el resto del trabajo.

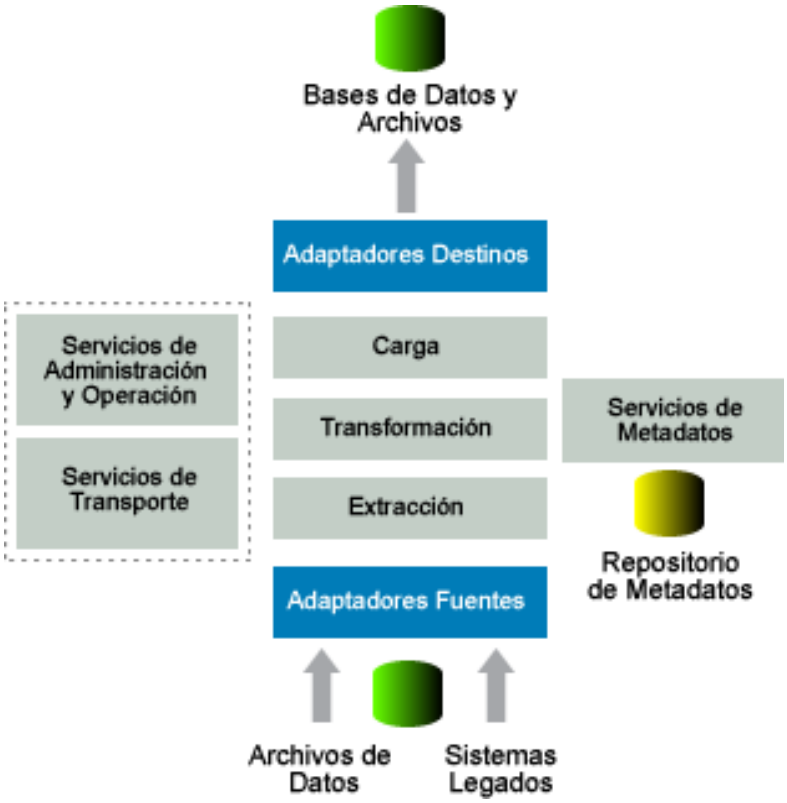

Fig. 6 Arquitectura de una solución ETL

Las componentes de la arquitectura mostrada son los siguientes:

I. **Servicios de Administración y Operaciones**: estos servicios aseguran la utilización efectiva de los recursos en el ambiente de sincronización. Aseguran una administración efectiva mediante la planificación y seguimiento de tareas, gestión de metadatos, recuperación de errores, etc.

II. **Servicios de Transportación**: procesos que garantizan el movimiento de la información cruda o transformada desde una fuente hasta un repositorio destino.

III. **Servicios de Metadatos**: Los Metadatos son información descriptiva sobre los datos y otras estructuras, como objetos, reglas de negocio, y los procesos que manipulan los datos. Los metadatos pueden ser muy técnicos y los usuarios en mucha de las ocasiones no lo entienden, sin embargo es fundamental documentar ya que se debe determinar el verdadero linaje de los datos

que y estos a su ves pueden ser utilizados por los usuarios finales o por el equipo del almacén de datos, demostrando la credibilidad de los datos y su integridad.

Los metadatos pueden ser agrupados en 3 categorías: (Kimball & Caserta, 2004)

a. **Metadatos Técnicos**: enfocado a los diseñadores, desarrolladores y administradores durante el desarrollo, el mantenimiento y la gestión del Ambiente TI utilizado. Este es el punto técnico que agrupa las herramientas, aplicaciones y sistemas, para que juntos constituyan la solución. Ejemplo de Metadatos Técnicos: el diseño del Esquema de un Almacén de Datos típicamente se almacena en un Repositorio como Metadatos el cual es utilizado para generar el script que construye las tablas del DW.

b. **Metadatos del Negocio**: por otra parte, brindan una imagen clara del servicio del ambiente de trabajo a los usuarios finales. Ejemplos de metadatos del Negocio: requerimientos del negocio, timelines, métricas del negocio, flujos de procesos del negocio, terminología del negocio.

c. **Metadatos de Proceso**: Es la presentación de las estadísticas sobre los resultados de la ejecución del propio proceso de ETL, incluyendo medidas tales como filas cargado con éxito, las filas rechazadas, y la cantidad de tiempo de carga, particularmente es importante en el proceso de limpieza de metadatos.

2.3 Identificación de las Fuentes de Datos

Según la necesidad de información y el problema a resolver, es lo que permite definir cuales son las fuentes de datos que se van a integrar, por ello la importancia de saber identificar que órganos de seguridad asumir para la integración.

Para la Identificación de Fuentes de Información en este caso es necesario responderse las siguientes preguntas:

- ¿Alguno de los Órganos de Seguridad proveedores de información ya integrados, contiene la información que se necesita?

- ¿Ya se está recibiendo la Información de alguna de estos Órganos de Seguridad, y demás fuentes pero no está integrada al Almacén?
- ¿Existen en los Órganos de Seguridad sistemas que permitan obtener la información digitalizada, o habría que crearlos?
- ¿Existen condiciones objetivas para que la Información esté disponible en tiempo y espacio (Conectividad, Latencia de Digitalización, otras)?

La aplicación de un Almacén de Datos en el Sector Social, tiene características propias que lo diferencian del resto de las aplicaciones de Almacenes de Datos Empresariales (comúnmente utilizados para hacer estudios comerciales, de clientes, productos, ventas, etc.)

Las fuentes de información para el tema de seguridad ciudadana, en su mayoría son autónomas en cuanto a la manipulación de algunos valores que aunque deberían ser vistos de una manera integral, no es así en la mayoría de los casos.

En el Sector Público, la integración de Fuentes Autónomas de Información lleva consigo un aumento considerable de la complejidad del proceso de unificación y combinación de todas las diferentes variables gestionadas por las Instituciones.

Es evidente que para que tenga sentido la combinación y el análisis posterior de valores de variables, estás deben de tener un significado único para la persona que realiza en análisis.

Las principales Fuentes de Información en el entorno de Seguridad Ciudadana son (Espinosa, 2007):

- Los **Sectores Justicia y Policía**, que disponen de la más completa información, pero no siempre está organizada y sistematizada.
- El **Sector Salud**, que atiende las lesiones de victimas de acuerdo a su severidad, pero pocas veces identifican la causa o móvil por la cual consultan. Clasifican los diagnósticos por el daño físico u orgánico a través del código internacional de enfermedades.
- **La Policía**, registra los delitos contra la vida e integridad personal, contra la propiedad y contra el estado. Acompaña a los fiscales en los levantamientos de cadáveres y recolectan datos acerca de la victima, las características del modo, tiempo y lugar, y en general las que pueden contribuir a la

investigación criminal. Como son los primeros en llegar al sitio del suceso, *los datos relacionados con la hora y fecha de ocurrencia de los hechos, registrados por la Policía, son los más confiables.*

- Las **Lesiones personales**, son de conocimiento de la Policía, sólo cuando el hecho es denunciado y en algunos casos sólo informan los que generan una incapacidad mayor de 30 días. Es una fuente poco confiable en este evento.
- **Medicina Legal**, prácticas de autopsias a las victimas de muertes violentas (homicidios, suicidios, muertes de tránsito y otras no intencionales) pero no siempre tienen una cobertura adecuada en los países. También realizan peritajes médico-legales cuando son ordenados por una autoridad judicial. La cobertura de otras formas de violencia esta supeditada a las denuncias y requerimientos de los fiscales.
- **La Fiscalía**, tiene la función de dirigir, realizar y coordinar las investigaciones en materia legal. Su principal objetivo es enjuiciar a los presuntos responsables de los hechos, mediante la recolección de pruebas y la investigación alrededor de los mismos. El sistema de información esta orientado a distribuir los casos denunciados a los fiscales para avanzar en las investigaciones.
- **Institutos de Estadísticas**, que contienen la información más fiable acerca de la demografía de las Distribuciones Político Administrativas que se abordan, en base a Censos de Población y Viviendas y otro amplio conjunto de variables importantes para el tema de la Seguridad Ciudadana, que combinadas con los demás Órganos de Seguridad pueden dar un resultado excelente.
- **Otros órganos:** Se debe tener en cuenta todos los órganos que de alguna forma contienen información importante para la solución.

2.4 Clasificación de las Fuentes de Datos

Para poder realizar el proceso de integración es necesario clasificar las fuentes de datos atendiendo a sus condiciones, de las Fuentes de Información para en base a la clasificación de cada una, planificar y desarrollar los mecanismos de integración en cada caso.

Las categorías propuestas son las siguientes (Microsoft, 2007):

1. **No cooperativas:** Exportan archivos de intercambio o permiten consultas directas con SQL o a través de Servicios Web, no se garantiza que el destino de la información la integre.

- ✓ **Snapshots:** Copia completa de la información congelada en un instante de tiempo.
- ✓ **Fuentes Específicas:** Ej. Sistemas Legados, Autónomos, etc. En este conjunto se encuentran las fuentes que brindan la información a partir de archivos intermedios, u otros mecanismos que no implican funcionalidades internas puntuales en base al receptor de la información.
- ✓ **Fuentes Consultables:** Suministra Interfaces para Consultas (SQL, Servicios Web, etc.).

2. **Fuentes Cooperativas** que a través de mecanismos de replicación, u otros mecanismos, se establecen intercambios mucho más fiables, seguros y responsables.
   - ✓ **Fuentes de Replicación**: Mecanismos de Publicación/Suscripción.
   - ✓ **Fuentes Callback**: Se invocan códigos externos de ETL cuando ocurren cambios en la información.
   - ✓ **Fuentes de Cambios Internos**: Se activan acciones internas cuando ocurren los cambios (Triggers).

		Fuente 1	Fuente 2	Fuente 3
NO COOPERATIVA	Snapshot			
	Fuentes Específicas (archivos, etc.)			
	Fuentes Consultables (SQL, WS)			
COOPERATIVA	Fuentes de Replicación			
	Fuentes de Call Back			
	Fuentes de Cambios Internos			

Fig. 7 Clasificación de las fuentes de datos

Luego de identificadas las fuentes, debe analizarse si es necesario cambiar de categoría alguna fuente, producto de la oportunidad con que se necesite la información. Los tipos de fuentes cooperativas permiten obtener resultados con latencia mínima. Clasificar las fuentes de datos es un paso importante para definir las características del proceso de integración de la fuente en cuestión.

2.5 Condiciones para la Integración de las Fuentes de Datos

Cuando se realiza una solución de almacenes de datos y se desea realizar el proceso de integración hay que tener bien claro que condiciones deben tener las fuente de datos para el éxito de la solución.

Estas condiciones deben ser (Espinosa, 2007):

1. **Voluntad Política** de los gobernantes, autoridades locales y nacionales, de utilizar la información par a la toma de decisiones.
2. **Coordinación con las Instituciones**, fuentes de Información (Policía, Medicina Forense, Fiscalía, Salud Pública, otras).
3. **Tecnología Apropiada** (Formación de Capital Humano y acompañamiento técnico y pedagógico), Software, análisis de información, estrategias de intervención, evaluación de estrategias, etc.

La ausencia de algunas de estas condiciones influye de forma negativa en el éxito de una solución de este tipo y atenta contra el cumplimiento de los objetivos básicos de la solución.

2.6 Conceptualización

Es necesario establecer los conceptos que utilizan las fuentes de datos y los que utilizarán el almacén antes de realizar la integración, generando preguntas que contribuyen a determinar los conceptos a manipular, ejemplo

¿Que se entiende por Delito, o Denuncia, o Imputado, Agraviado, Víctima, o sean cuáles sean los nuevos conceptos a integrar?

Luego de la conceptualización, podrían entonces, definirse reglas de transformación que traduzcan los términos y valores de las variables a conceptos conocidos, manejados y con un sentido realmente práctico.

2.7 Extracción de Datos

Para integrar las fuentes de datos al almacén el primer proceso que se realiza es el de extracción. En estos casos, lo normal es que los datos necesarios pertenezcan a diferentes organizaciones, a distintos

departamentos dentro de una misma institución e incluso puede ocurrir que algunos datos necesarios para el almacén nunca hayan sido recolectados en el ámbito de la organización por no ser necesarios para sus aplicaciones.

Frecuentemente deben adquirirse datos externos desde bases de datos públicas (como censos, datos demográficos o climatológicos) o desde bases de datos privadas (como los datos de compañías de pagos, bancarias, automovilística, etc.). Esto representa un reto, ya que cada fuente de datos usa diferentes formatos de registro, diferentes grados de agregación de los datos, diferentes claves primarias, diferentes tipos de error, etc.

En esta parte del proceso de integración juega un papel importante definir los medios de comunicación, entre la fuente de datos y el área donde se realizará la limpieza y transformación. Todo este proceso se debe efectuar en las noches o fines de semana que es cuando las fuentes no están en plena labor, para no perjudicar el rendimiento de sus aplicaciones; siendo este principio fundamental para realizar el proceso de extracción.

2.8 Transformación de Datos

El proceso de transformación es el que más influye en la calidad de los datos por lo que el costo de tiempo y recursos aumenta en esta parte de la solución. El tipo de las transformaciones y el costo que esta generan depende enormemente de la calidad de los datos de las fuentes de información. La transformación de los datos es el proceso más importante que se realiza en las ETL.

La transformación de datos engloba cualquier proceso que modifique la forma original de los datos. Pueden existir transformaciones que convierten un conjunto de atributos en otros, o deriven nuevos atributos. Existen otras que afectan a toda la tabla o varias tablas. Prácticamente todos los procesos de preparación de datos entrañan algún tipo de transformación. En esta parte del proceso se diseñan las reglas de transformación, se ponen en práctica los *diccionarios de datos*[21] y donde tiene más impacto la utilización de metadatos, además de que se genera una gran cantidad de información a ser monitoreada, sobre el éxito o fallo de los procesos de ETL.

[21] Lista de todos los elementos de datos de las fuentes y sus descripciones.

2.9 Calidad de Datos

El concepto de "calidad de datos" se asocia con mucha frecuencia a los sistemas de información. Frecuentemente al procesar grandes volúmenes de datos se encuentran datos incompletos e inconsistentes. Estos problemas se acentúan cuando realizamos la integración de distintas fuentes. No obstante, mientras la cantidad de datos erróneos aumentan de manera lineal respecto al tamaño de los datos recopilados, los datos inconsistentes se multiplican; varias fuentes diferentes pueden afirmar cosas distintas sobre el mismo objeto.

La integración también produce una disparidad de formatos, nombres, rangos, etc., que podría no existir, en las fuentes originales. Para resolver esta disparidad se presentan una serie de consejos para la integración, para convertir los datos en otros más apropiados.

La calidad de datos es un término que abarca tanto el estado de los datos, como el conjunto de procesos para lograr dicho estado. El objetivo es disponer de datos libre de duplicados, errores, omisiones, variaciones e innecesario, y que los datos se ajusten a la estructura definida.

Cuando de habla de calidad de datos se debe tener en cuenta que la información que se procese debe ser (Kimball & Caserta, 2004):

I. **Correcta:** Los valores y las descripciones de los datos deben describir su verdadera definición. Un ejemplo de esto es si estamos guardando la ciudad de una persona en particular, entre los datos precisos de esta persona debe incluirse esa cuidad y no otra, la información debe ser correcta.

II. **Inequívoco**: Los valores y las descripciones de los datos sólo pueden tener un solo significado. Un ejemplo de esto es que en el mundo hay varias ciudades que se llaman Santiago por lo tanto es preciso cuando se vaya a dar la dirección de esta ciudad se debe mencionar Santiago y el país a que pertenece (en este caso Cuba) siendo la combinación inequívoca.

III. **Coherente:** Los valores y las descripciones de datos deben usar una notación constante para transmitir su verdadero significado. Ejemplo la provincia de Santiago de Cuba se podría expresar como “*Sgto. de Cuba*”, “*StgCuba*” o “*Santiago de Cuba*”. Para mantener la coherencia de los datos se debe utilizar solo una nomenclatura.

IV. **Completa:** Esta opción se agrupa en dos definiciones:

1. El primero es garantizar que los valores individuales y las descripciones de los datos, se definan para cada caso, permitiendo identificar que valores posibles puede tomar cada dato. *Ejemplo:* hay que velar por que todos los registros que deben tener direcciones, la tengan.

2. El segundo aspecto es que se debe asegurar que el número total de registros completados después que se realice el proceso de integración debe ser del 100% completo (la información que se integra debe ser la misma información que se almacena) o por lo menos asegurarse de que no se pierda información en alguna parte del flujo de datos.

Para garantizar la calidad de los datos en todo el proceso de integración, se definen varios procesos (Kimball & Caserta, 2004):

Perfilado de Datos (Data Profiling) es el proceso de examinar los datos que existen en una organización y recopilar estadísticas e información sobre los mismos. El propósito de dichas estadísticas es:

- ✓ Conseguir métricas de calidad de datos que determinan si los datos cumplen los estándares de la organización.
- ✓ Reducir el riesgo de integrar información a nuevas aplicaciones dado que conocemos su estado.
- ✓ Permitir hacer un seguimiento de la calidad de datos.
- ✓ Tener capacidad de entender problemas derivados de los datos en proyectos que hagan uso intensivo de los mismos.
- ✓ Tener una visión global de los datos de la organización.

Limpieza de Datos (Data Cleansing) es el proceso de detectar o descubrir, y corregir datos corruptos, incoherentes o erróneos de un conjunto de datos. Este proceso permite detectar entradas duplicadas, incompletas y establecer reglas para corregirlas. El objetivo no es borrar información sino mejorar la calidad de los datos construyendo un proceso de mejora continua.

Auditoría de Datos (Data Auditing) es el proceso de gestionar cómo los datos se ajustan a los propósitos definidos por la organización. Se establecen políticas para gestionar los criterios de datos para la organización. No es suficiente con actuar sino que se debe vigilar.

La calidad de los datos depende fundamentalmente de plantarse estrategias para predecir, accionar y velar por la calidad de la información.

2.10 Carga de Datos

El proceso de carga es la fase donde ya los datos han sido extraídos de la fuente y se le han aplicado ya las transformaciones correspondientes, en ese momento los datos están preparados para ser cargados al almacén de datos o cualquier otro destino. Dependiendo de los requerimientos de la organización, este proceso puede abarcar una amplia variedad de acciones diferentes.

Existen dos formas básicas de desarrollar el proceso de carga (William H., 2005)**:**

Acumulación simple: La acumulación simple es la más sencilla y común, y consiste en realizar un resumen de todas las transacciones comprendidas en el período de tiempo seleccionado y transportar el resultado como una única transacción hacia el almacén de datos, acumulando un valor calculado que consistirá típicamente en un sumatorio o un promedio de la magnitud considerada.

Rolling: El proceso de Rolling por su parte, se aplica en los casos en que se opta por mantener varios niveles de granularidad. Para ello se almacena información resumida a distintos niveles, correspondientes a distintas agrupaciones de la unidad de tiempo o diferentes niveles jerárquicos en alguna o varias de las dimensiones de la magnitud almacenada (por ejemplo, totales diarios, totales semanales, totales mensuales, etc.).

La fase de carga interactúa directamente con la base de datos de destino. Al realizar esta operación se aplicarán todas las restricciones y triggers que se hayan definido en ésta (por ejemplo, valores únicos, integridad referencial, campos obligatorios, rangos de valores). Estas restricciones y triggers contribuyen a que se garantice la calidad de los datos en el proceso ETL, y deben ser tenidos en cuenta.

2.11 Área de Almacenamiento Intermedio

Existen situaciones en las que en el momento de construir un almacén de datos, es lícito plantearse el uso de un **área de almacenamiento intermedio**[22].

Se entiende por área temporal, al medio de almacenamiento intermedio que permanece entre las fuentes de datos y el almacén de datos con el objetivo de:

- ✓ Facilitar la extracción de datos (los procesos ETL) desde las fuentes de origen de carácter múltiple realizando un pre-procesamiento.
- ✓ Realizar limpieza de datos.
- ✓ Mejorar la calidad de datos.
- ✓ Ser usado como caché de datos operacionales con el que posteriormente se realiza el proceso de Almacén de Datos.
- ✓ Usar la misma para acceder en detalle a información no contenida en el Almacén de Datos

¿Cuándo es conveniente usar un área de almacenamiento intermedio en un proyecto de **Almacenes de Datos**?

La capacidad de desarrollar procesos ETL eficientemente depende, en parte, ser capaz de determinar el equilibrio adecuado entre almacenamiento físico y el procesamiento en memoria. La decisión de utilizar un área de almacenamiento intermedio varía en función del entorno y los requisitos de la institución que se tengan.

Según Kimball para definir la utilización del área de almacenamiento intermedia se debe tener las siguientes razones:

[22] También se encuentra en la bibliografías como Área de Almacenamiento Temporal y en inglés se conoce como **Staging Area o STA**

Recuperación: Es una buena practica almacenar los datos que son extraídos de las empresas en una base de datos o sistema de archivos como un punto de recuperación, permitiendo que si el proceso de carga falla, no tener la necesidad de volver a transformar. Otra de las causas puede ser también que la fuente de datos sobre escriben sus propios datos, por lo que una falla en cualquier parte del proceso imposibilitaría volverá gestionar la información de las fuente nuevamente. Por lo que la decisión del diseño de esta área puede ser exclusivamente para la recuperación.

Copias de Seguridad: Muy a menudo el enorme volumen de almacenamiento de las copias de seguridad impiden que se realice esta acción tan importante, sin embargo puede ser muy útil mantener, si no se tiene una base de datos de respaldo al menos tener un sistemas de archivos comprimidos que almacenen la información después de cada carga.. Muchas veces por problemas internos de la empresa es preciso realizar el llenado del almacén nuevamente, sin embargo si se tiene una copia de los archivos cargados, bastaría con descomprimir los archivos y realizar nuevamente la carga del almacén sin tener ningún riesgo de perdida de la información.

Auditoría: Muchas veces los procesos de ETL producen incoherencias al comparar los datos provenientes de las fuentes y los resultantes, luego de la extracción y la transformación. Cuando llegue el momento de realizar una auditoría de los procesos de ETL, es mucho más sencillo porque los auditores (o programadores) pueden simplemente comparar el archivo de entrada original, con las reglas lógicas de transformación, contra el archivo de salida.

Diseñar el área de datos correctamente es más importante que el diseño de las aplicaciones habituales, debido a la gran cantidad de datos que se acumula en esta área (a veces más grande que el almacén de datos en sí). Por ello el uso o no del área de almacenamiento intermedio depende de las necesidades y condiciones que se tienen para dar solución al problema en cuestión.

2.12 Conclusiones del Capítulo

En este capítulo mostraron los principales conceptos que componen los procesos de ETL, así como las pautas más importantes que permiten el desempeño satisfactorio de la solución. Estos procesos van más allá de las tres acciones por las cuales se conforma las ETL sino también actividades que contribuyen o que definen la calidad de la solución, que abarca desde la identificación de las fuentes de información a integrar, hasta los procesos de ETL.

Se definió la arquitectura que se utilizó en la solución, las características a tener en cuenta para seleccionar las fuentes de datos, así como las clasificaciones según la vía, el medio y el formato con que nos brindan la información. Se abordaron las condiciones necesarias e imprescindibles para utilizar una fuente de datos así como las peculiaridades de los procesos de ETL, explicando y abordando el tema de la calidad de datos para asegurar la integridad de los datos manejados en estos procesos. Adicionalmente se fundamenta la necesidad de utilizar un área de almacenamiento intermedio para sustentar la solución informática.

Capítulo 3: Descripción de la Solución

3.1 Introducción

En este capítulo se detallan las especificaciones de la capa de integración de datos. La elección de las componentes o tecnologías, descritas en este capítulo, se fundamentan en los capítulos anteriores. Para el modelado de la solución se utilizan las herramientas de Oracle, en específico el Oracle Warehouse Builder 10g R2; el FTP como protocolo de comunicación y el uso de archivos de intercambio de datos con formato XML para la transferencia de datos. Se especifican las peculiaridades de cada componente del proceso ETL, y en específico, el diseño y la creación del área intermedia para desarrollar la integración de las fuentes de datos. También se presentan elementos relacionados con la configuración, la implantación y despliegue de la capa de integración de datos del SINSEC.

3.2 Metodología de Diseño

La metodología de diseño utilizada se basa en la definición y uso de técnicas y de extracción, transformación y carga de datos como herramienta de integración de datos del Almacén de Datos de Seguridad Ciudadana del SINSEC, a fin de poner a disposición de los analistas toda la información disponible en los Órganos de Seguridad y complementar el requerimiento de análisis y tratamiento de la información relacionada con Incidencia Delictiva y Actuar Policial en la República Bolivariana de Venezuela.

Como procesos básicos para aplicar las técnicas de integración de datos, primeramente se debe conocer las necesidades de información, las fuentes que las cubren y por último la realidad de estas fuentes de información y la posibilidad y factibilidad de utilizarlas debidamente, para luego combinar las estructuras de datos bases que darán origen al Sistema de Información. La información debe ser depurada e integrada para generar la información que debe ser utilizada por los analistas, funcionarios o empresarios.

Para la solución se utilizó un modelo iterativo e incremental de 4 pasos que constituye las bases del desarrollo del SINSEC, enfocado en la asimilación paulatina de la información y de la propia tecnología por parte de los analistas y del equipo de soporte del MPPRIJ, así como en el incremento del dominio de análisis del propio almacén. (ALBET.S.A., 2008).

Dicho modelo prevé la mejora continua de la organización en cuanto a la calidad de la gestión.

Fig. 8 Lógica de Desarrollo del SINSEC

1. Se requiere de nueva información para ampliar el dominio de análisis.
2. Identificación de fuentes fiables de obtención de la información necesaria.
3. Proceso de integración
4. Análisis y explotación sistemática, por parte de los analistas, del nuevo dominio informativo ampliado que puede conllevar a la necesidad de adquisición de nueva información.

El éxito de una solución de este tipo para contribuir a la seguridad ciudadana, y que también se evidencia en otros tipos de aplicaciones similares viene dado, en gran medida, por un estudio amplio de las instituciones y órgano de seguridad involucradas en el tema, a partir del conocimiento y de la experiencia de los responsables.

De esta forma el SINSEC, se encarga de:

1. *El Proceso de Extracción, Transformación y Carga de Datos, factor clave para lograr que los datos extraídos se integren finalmente en un entorno homogéneo y estandarizado (el cual se explica en el presente trabajo).*
2. La creación y puesta en marcha del Almacén de Datos, donde residen en Modelos Multidimensionales, con la información integrada de los Órganos de Seguridad, y preparados para la consulta, por parte de los Analistas y demás Especialistas en estos temas.
3. Facilitar la explotación del Almacén de Datos de Seguridad Ciudadana, a través de herramientas de Inteligencia de Negocios (Inteligencia Institucional), y de un amplio espectro de nuevos modelos de consulta, que organiza la información de acuerdo a temas específicos y facilita su acceso.

El desarrollo de esta solución parte desde que se han dado los primeros pasos para la creación del almacén y se está en la etapa de integración de la fuente de datos.

3.3 Arquitectura del SINSEC

La siguiente imagen muestra la arquitectura que se utilizó para el desarrollo de la solución, teniendo en cuenta la numeración se especificará su significado y su impacto enmarcándose principalmente el trabajo en las numeraciones que forman parte del proceso de integración.

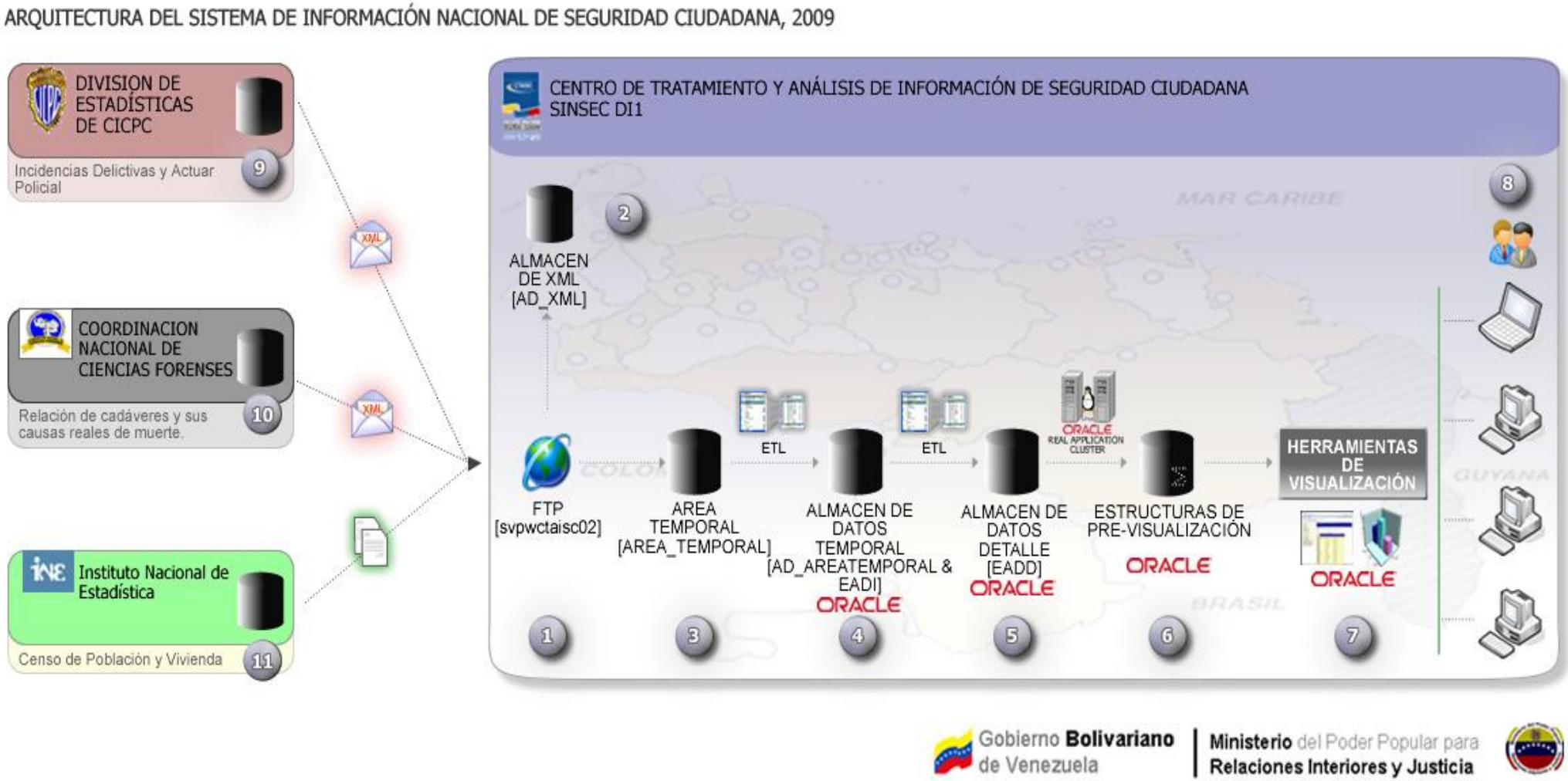

Fig. 9 Arquitectura del SINSEC

3.4 Componentes de la Arquitectura del SINSEC

3.4.1 Componente 1. [FTP]

FTP como vía de comunicación e integración de las fuentes con el CTAISC. Cada Fuente de Datos tiene credenciales para acceder a su Área de Transferencia y de esta forma, hacer llegar la información hacia el Ministerio. La transferencia de información se realiza de forma segura y personalizada. Esta componente forma parte de los Servicios de Transportación referenciados en la arquitectura genérica de soluciones ETL.

3.4.2 Componente 2. [AD_XML]

Almacén de archivos XML (en disco) como mecanismo de Respaldo con vistas a almacenar la historia de los Archivos de Intercambio de Datos XML como mecanismo paralelo de respaldo. En este punto se tendrían todos los archivos de Intercambio provenientes de las fuentes. Un ejemplo de ello se muestra en la figura 3.3.

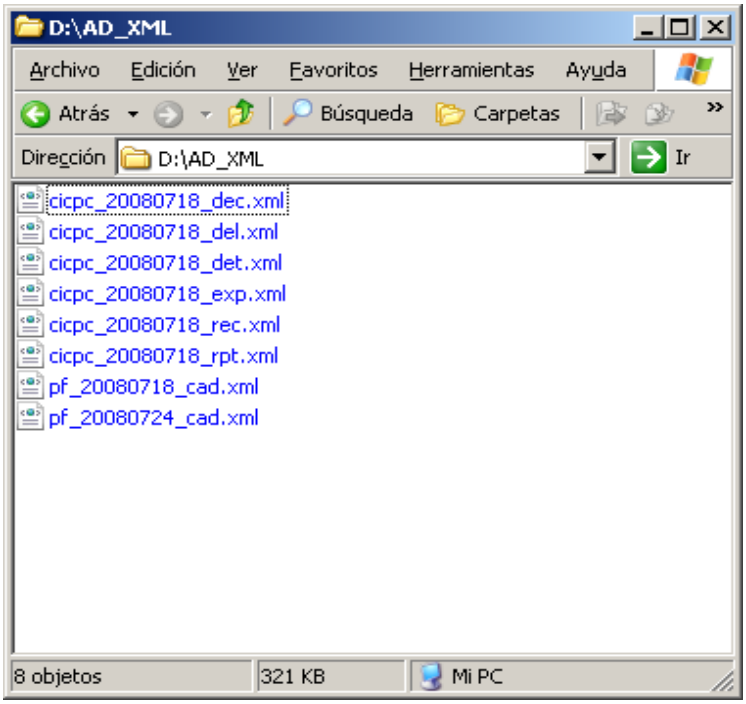

Fig. 10 Almacén de Datos XML [AD_XML]

Se realiza una copia exacta de los Archivos de Intercambio de Datos hacia la Carpeta Histórica de XML.

/home/etl/ad_xml

3.4.3 Componente 3. [AREA_TEMPORAL]

Área temporal (en disco) con vistas a preparar los Archivos de Intercambio de Datos XML antes de su integración. Los archivos recibidos de las fuentes de información vienen estructurados, en su nombre, en base a algunas reglas para poder ser identificados. Ejemplo. El Archivo de Intercambio cicpc_01072008_del.xml representa el archivo proveniente de CICPC, en la fecha 01/07/2008, con todos los delitos (incidentes delictivos o casos abiertos).

En el AREA_TEMPORAL se prepara el archivo para ser integrado. Se elimina la fecha para que la integración no sea relativa, sino absoluta a un nombre único. En el ejemplo anterior el archivo quedaría listo para ser procesado bajo el nombre cicpc_del.xml.

/home/etl/areatemporal

3.4.4 Componente 4. [ALMACÉN DE DATOS TEMPORAL]

Entre las componentes más importantes de la arquitectura del SINSEC se encuentra la Integración de Datos. Esta componente se encarga de:

i. La carga de los archivos de intercambio de datos, provenientes de las fuentes de información.

ii. La extracción de los datos que contienen los archivos de intercambio de datos.

iii. La validación de integridad y limpieza de los datos cargados.

iv. Permite aplicar las reglas de correspondencia, proceso que se basa en un Diccionario de Correspondencia debidamente formado, cuya función es traducir los términos utilizados por las fuentes de datos, a los términos que maneja el SINSEC.

En base a este punto, el SINSEC maneja un lenguaje o terminología que aunque partió de un estudio exhaustivo de los términos de negocio de las fuentes de datos, se generaliza, permitiendo que aunque cambien los valores de las fuentes de información, los analistas del SINSEC puedan continuar haciendo

sus análisis, posteriormente que se reconfigure el Diccionario de Correspondencia para poder interpretar los valores que se refinen de las fuentes.

Esta componente además, luego de la depuración e interpretación de datos anteriormente mencionada, permite integrar los datos de varias fuentes de datos en un repositorio libre de errores con una visión única donde puede tenerse una interrelación entre los datos de una fuente con otra.

Estos procesos se realizan en dos momentos:

- ✓ Componente AD_AREATEMPORAL, encargada de la depuración y la interpretación.
- ✓ Componente Estructura de Almacén de Datos Integrados, que permite combinar y hacer corresponder la información de más de una fuente, por algún concepto de unión. Por ejemplo, en esta componente se integra la información de la policía con la de la morgue, con el criterio de enlace EXPEDIENTE DELICTIVO.

Estas dos partes importantes de la solución forman parte del área de almacenamiento intermedio. Más adelante se describe la configuración de ambas subcomponentes.

Componente 4.1 [AD_AREATEMPORAL]

A continuación se describen las funciones de este esquema. Que es parte del área de almacenamiento intermedio.

Respaldo y Preparación de Archivos de Intercambio de Datos

El respaldo de los archivos de intercambio de datos XML hacia AD_XML (Ver AD_XML), permite la recuperación de estos archivos en el caso que se necesiten.

La preparación de los archivos mediante el ajuste del nombre de los archivos hacia AREATEMPORAL (Ver AREATEMPORAL).

Se realiza una copia exacta de los Archivos de Intercambio de Datos hacia la Carpeta AREA_TEMPORAL donde se acomoda el título de cada XML a una forma absoluta de Integración. Ejemplo: de cicpc_04072008_del.xml a cicpc_del.xml.

Repositorio de Metadatos

Como ya se mencionó anteriormente los Servicios de Metadatos representan Información descriptiva sobre los datos y otras estructuras, como objetos, reglas de negocio, y los procesos que manipulan los datos.

En el AREA_TEMPORAL se almacenan datos sobre los procesos de ETL. Se define el mecanismo de adquisición de cada archivo de integración en formato XML, así como la estructura de cada uno de sus nodos.

ETL inicial desde los archivos XML

Se realiza la extracción, y carga (ETL) de los datos contenidos en los archivos de intercambio de datos hacia un esquema de Base de Datos Oracle basando el proceso en los valores del Repositorio de Metadatos anteriormente descrito. En este punto no se realiza “transformación” de los datos, sino que son cargados exactamente como aparecen en los archivos.

Traducción basada en Diccionarios de Correspondencias

La traducción, valor a valor, de las variables que provienen de las fuentes de datos según un diccionario de correspondencia previamente configurado. Como principal ventaja del uso de diccionarios de correspondencia se tiene que este mecanismo permite separar la terminología del SINSEC, y del CTAISC, como órgano superior integrador, de las terminologías heterogéneas de las fuentes de datos.

El diccionario de correspondencia descrito en el repositorio de metadatos (ETL_TRANSLATION_DICTIONARY) contiene los valores de entrada y valores de salida para posible valor de cada una de las variables, provenientes de cada fuente de datos. Esto quiere decir, que se traducen todos los elementos de entrada.

Existen casos en el que la traducción no es posible, producto de valores que no aparecen NULL, o valores no previstos. Para estos casos se aplica una traducción de sustitución que indica la transformación a realizar para el caso de valores omitidos, o valores no previstos.

Es importante destacar en este punto, que cada traducción realizada durante los procesos de extracción, transformación y carga de datos, queda registrada dentro de las trazas del sistema, específicamente dentro del repositorio de metadatos (ETL_LOG). La capa de visualización desarrollada, a través de técnicas y herramientas de Inteligencia de Negocios, tiene como principal objetivo el tratamiento y análisis de la información del CTAISC, y como segundo objetivo: el seguimiento del proceso de integración de datos, que constituye el principal Mecanismo de Monitoreo de los Procesos de ETL.

La principal desventaja o elemento que aumenta considerablemente la complejidad de la solución consiste en que debe tenerse un personal con suficiente dominio y potestad que se encargue de mantener actualizados los diccionarios de correspondencias.

El ajuste del diccionario de correspondencia se realiza por medio del paquete DBMS_SINSEC. Este paquete se implementó para la invocación manual de las principales operaciones del proceso de integración. Se dedica una sección a este tema.

Componente 4.2 [ESTRUCTURAS DE ALMACÉN DE DATOS INTEGRADO]

Este esquema almacena la información de las fuentes de datos antes de pasar al Almacén de Datos Histórico Detallado.

Existen estructuras de almacenamiento para el proceso de integración de datos.

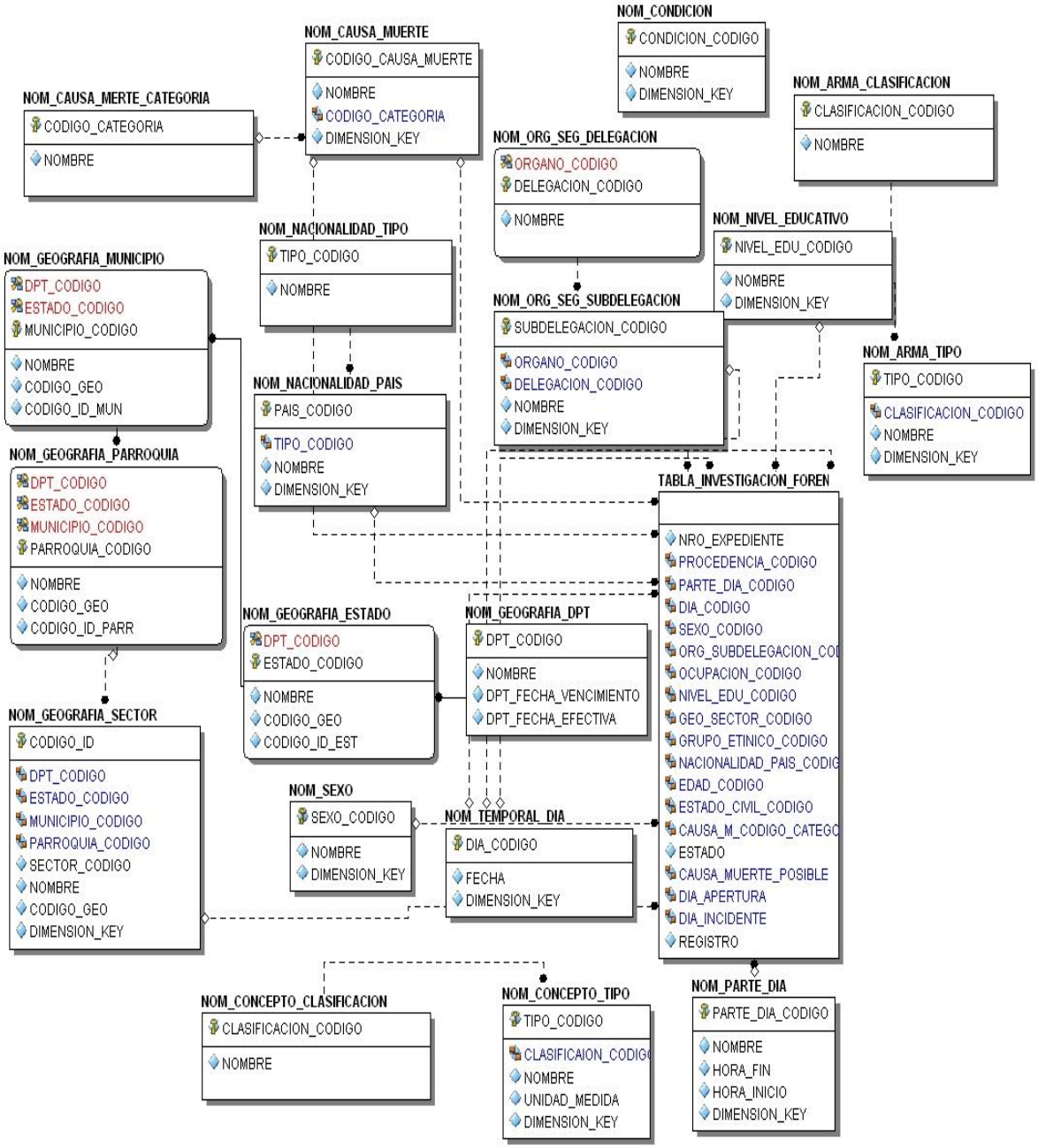

Fig. 11 Módulo de Medicina Forense del esquema estructuras de almacenamiento de datos integrados.

En este esquema que forma parte del área de almacenamiento intermedio se realiza las transformaciones que permiten medir la calidad de los datos. Además de dejar un registro de las principales transformaciones realizadas como política de respaldo, así como auditar el proceso de transformación.

3.4.5 Componente 5. [EADD]

Esta componente constituye explícitamente, el Almacén de Datos del SINSEC. En ella se almacena, en forma de cubos, los datos provenientes de las fuentes de información, depurados e interpretados al lenguaje del SINSEC, ya listos para el análisis.

3.4.6 Componente 6. [ESTRUCTURAS DE PREVISUALIZACIÓN]

En el transcurso del tiempo, los datos del Almacén suelen aumentar substancialmente y por tal motivo se hace necesario preparar el sistema para que pueda dar resultados en tiempos lógicos y oportunos, de lo contrario recuperar información del almacén sería una tarea muy difícil debido al tiempo en que tardaría este tipo de actividad. Una consulta a un almacén de datos que no considere este punto, podría tardar horas y horas, e incluso días, para mostrar el resultado deseado, en el mejor de los casos.

Esta componente permite preparar la información para ser consultada, por medio de vistas materializadas que pueden programarse para ser actualizadas en horarios en que la inactividad en el centro lo permita.

3.4.7 Componente 7. [ORACLE BI]

Suite de Herramientas de Inteligencia de Negocios (Oracle Business Intelligence Standard Edition) para la preparación y presentación de los datos a los usuarios finales. Sus componentes facilitan el análisis de los datos, en forma dinámica, a los analistas y expertos del CTAISC.

Esta componente se estructura a partir de la concepción de:

a. Áreas de Análisis

b. Libros de Trabajo

c. Horas de Datos o Reportes.

Estos conceptos no forman parte del presente trabajo.

3.4.8 Componente 8. [CLIENTES]

Los clientes forman parte de la solución. Cada rol o comportamiento de personas ante el sistema fue debidamente concebido. El CTAISC como Centro de Análisis de Información de Seguridad Ciudadana tiene responsabilidades ante otros entes a los cuales debe entregar información periódicamente en muchos casos. Asimismo, la información puede ser consultada directamente a partir de las herramientas de Inteligencia de Negocios implantadas. Los roles creados para el SINSEC permiten la diferenciación de los comportamientos dado el tipo de accesos a la Información en cuestión. La seguridad es el punto clave para lograr tal objetivo.

3.4.9 Componente 9. [CICPC]

CICPC, brinda información de los Hechos Delictivos y el Actuar Policial. Para esta fuente se configuró el Diccionario de Correspondencia de la Componente 4 que traduce toda la terminología de este Órgano de Seguridad hacia el Dialecto del SINSEC. Actualmente se reciben de CICPC 5 Archivos de Intercambio, utilizando un protocolo de comunicación FTP (File Transfer Protocol).

3.4.10 Componente 10. [PF]

Patología Forense, brinda información de los Cadáveres que se reciben cada día. A cada cadáver se le realiza una autopsia dando como resultado un Dictamen (también llamado Protocolo). Para esta fuente se configuró el Diccionario de Correspondencia de la Componente 4 que traduce toda la terminología de este ente del Estado hacia el Dialecto del SINSEC. Actualmente se recibe de Patología Forense un Archivo de Intercambio, utilizando un protocolo de comunicación FTP (File Transfer Protocol).

3.4.11 Componente 11. [INE]

Si bien el concepto de seguridad, dentro del marco constitucional, rebasa la cuestión delictiva, esta última representa un aspecto importante de la seguridad y un desafío para la sociedad venezolana actual, tanto en términos de costos de vidas humanas, problemas económicos y disminución de la calidad de vida, como en términos de erosión de la confianza ciudadana en las instituciones del Estado, cuando no hay respuestas efectivas y adecuadas para enfrentarla.

Para determinar los factores asociados a la delincuencia se requiere información estadística sobre la población, y sus características a diferentes niveles espaciales (geográficos). De esta forma se formularán medidas y estrategias más ajustadas a las características regionales y locales. Se le realizó una solicitud de Información (SOLICITUD-INE-JUNIO-08) al Instituto Nacional de Estadística (INE).

Se describen los resultados de la integración luego de analizar la información brindada por el INE.

3.5 Descripción de las Fuentes de Datos

El SINSEC en su primer Dominio Informativo integrará la Información del CICPC, Patología Forense y en Instituto Nacional de Estadísticas para complementar la demanda de información de:

- Hechos Delictivos (Homicidios, Robos, Hurtos, etc.)
- Agraviados
- Imputados
- Decomisos
- Recuperaciones Policiales
- Expedientes de Casos Culminados
- Dictámenes Forenses que corroboren o desmientan las estadísticas de Homicidios
- Información Demográfica por Distribución Político Territorial, con vistas a obtener relaciones entre las víctimas y la población total por región.

Las características de las fuentes de datos seleccionadas se muestran en la tabla siguiente:

Fuente	Tecnología Utilizada	Medio de Transferencia	Conectividad	Información Histórica Digitalizada	Clasificación
CICPC	BD-Access	XML	no	si	Fuentes Específicas
PF	BD-Access	XML	no	no	Fuentes Específicas
INE	Oracle	CD	si	si	Snapshot

Tabla 1 Clasificación de las Fuentes de Datos del SINSEC

3.5.1 CICPC

Los archivos de intercambio de datos, que se reciben de CICPC son los siguientes*:*

cicpc_*_del.xml Incidentes delictivos, agraviados, imputados, vehículos solicitados, armas solicitadas.

cicpc_*_det.xml Detenciones de las subdelegaciones en un rango de fechas.

cicpc_*_dec.xml Decomisos de las subdelegaciones en un rango de fechas.

cicpc_*_rec.xml Recuperaciones de las subdelegaciones en un rango de fechas.

cicpc_*_exp.xml Expedientes remitidos por las subdelegaciones en un rango de fechas.

cicpc_*_rpt.xml Reporte de las subdelegaciones que han brindado información en un rango de fechas.

NOTA: El Asterisco (*) en el nombre de los archivos se sustituye cada día por la fecha. Un ejemplo de archivo sería: cicpc_20080704_del.xml para un archivo de intercambio del día 04 de Julio del 2008.

A continuación se describen estos archivos

1. **CICPC_DEL.XML**

Archivo XML con el listado de todas las incidencias ocurridas en un rango de tiempo determinado. El incidente tiene asociado todos los objetos o conceptos incluidos lógicamente en el incidente (Agraviados, Imputados, Vehículos Denunciados, Armas Denunciadas, Otros montos).

Nombre del archivo XML: cicpc_FECHAACTUAL_del.xml (cicpc_20080406_del.xml)

Campo	Descripción
incidencia	Listado de incidentes delictivos ocurrido durante el tiempo que se solicita.
agraviado	Listado de agraviados de los incidentes delictivos ocurridos durante el tiempo que se solicita.
imputado	Listado de imputados de los incidentes delictivos ocurridos durante el tiempo que se solicita.
vehiculo_denunciado	Listado de vehículos denunciados en los incidentes delictivos ocurridos durante el tiempo que se solicita.
arma_denunciada	Listado de armas denunciadas en los incidentes delictivos ocurridos durante el tiempo que se solicita.
monto_denunciado	Listado de montos por concepto denunciados en los incidentes delictivos ocurridos durante el tiempo que se solicita.

Tabla 2 Contenido del XML: CICPC_DEL.XML

Detalle técnico

Listado serializado de objetos que contendrán la siguiente información:

<incidente>

…

<agraviado>…</agraviado>

<imputado>…</imputado>

<vehiculo_denunciado>…</ vehiculo_denunciado >

<arma_denunciada>…</ arma_denunciada >

<monto_denunciado>…</ monto_denunciado >

</incidente>

Las Especificaciones del XML se describen en los anexos.

2. CICPC_DET.XML

Archivo XML con el listado de todos los detenidos y sus datos.

Nombre del archivo XML: cicpc_FECHAACTUAL_det.xml (*cicpc_20080406_det.xml*)

Campo	Descripción
detención	Listado de detenciones efectuadas durante el tiempo que se solicita.

Tabla 3 Contenido del XML: CICPC_DET.XML

Detalle Técnico

Listado serializado de objetos que contendrán la siguiente información:

<detencion>…</detencion>

Las Especificaciones del XML se describen en los anexos.

3. CICPC_DEC.XML

Archivo XML con el listado de todos los decomisos (solamente de drogas).

Nombre del archivo XML: cicpc_FECHAACTUAL_dec.xml (cicpc_20080406_dec.xml)

Campo	Descripción
Decomiso	Listado de decomisos de drogas ocurridos durante el tiempo que se solicita.

Tabla 4 Contenido del XML: CICPC_DEC.XML

Detalle Técnico

Listado serializado de objetos que contendrán la siguiente información:

<decomiso>

…

<droga>…</droga>

</decomiso>

Las Especificaciones del Archivo se describen en los anexos.

4. CICPC_EXP.XML

Listado de todos los expedientes remitidos a fiscalía y que fueron concluidos policialmente o no.

Nombre del archivo XML: cicpc_FECHAACTUAL_exp.xml (*cicpc_20080406_exp.xml*)

Campo	Descripción
Expediente	Listado de expedientes remitidos a fiscalía durante el tiempo que se solicita.

Tabla 5 Contenido del XML: CICPC_EXP.XML

Detalle Técnico

Listado serializado de objetos que contendrán la siguiente información:

<expediente>

…

</expediente>

Las Especificaciones del Archivo se describen en los anexos.

5. **CICPC_REC_XML**

Archivo XML con el listado de todas las recuperaciones llevadas a cabo y los objetos recuperados. Una recuperación tiene asociado uno o varios elementos clasificados en Vehículos, Armas, Montos por Concepto, Drogas; además se incluyen las vidas rescatadas.

Nombre del archivo XML: cicpc_FECHAACTUAL_rec.xml (cicpc_20080406_rec.xml)

Campo	Descripción
Recuperación	Listado de recuperaciones efectuadas durante el tiempo que se solicita.
vida_rescatada	Listado de vidas rescatadas producto de las denuncias y casos abiertos de secuestros, etc.
vehiculo_recuperado	Listado de vehículos recuperados durante el período de tiempo que se solicita.
arma_recuperada	Listado de armas recuperadas durante el período de tiempo que se solicita.
monto_recuperado	Listado de montos por concepto recuperados durante el período de tiempo que se solicita.
droga_recuperada	Listado de drogas por tipo recuperadas durante el período de tiempo que se solicita.

Tabla 6 Contenido del XML: CICPC_REC.XML

Detalle Técnico

Listado serializado de objetos que contendrán la siguiente información:

<recuperacion>

…

<vida_rescatada>…</vida_rescatada>

<vehiculo_recuperado>…</vehiculo_recuperado>

<arma_recuperada>…</arma_recuperada>

<monto_recuperado>…</ monto_recuperado >

</recuperacion>

Las Especificaciones del Archivo se describen en los anexos.

3.5.2 PF

Los Archivos de Intercambio de datos son Archivos XML. Se reciben de Patología Forense:

pf_*_cad.xml Cadáveres recibidos por Patología Forense con sus causas de muerte, móviles, y datos generales del cadáver.

NOTA: El Asterisco (*) en el nombre de los archivos se sustituye cada día por la fecha. Un ejemplo de archivo seria: pf_20080704_cad.xml para un archivo de Intercambio del día 04 de Julio del 2008.

PF_CAD.XML

Archivo XML con el listado de todos los cadáveres.

Nombre del archivo XML: pf_FECHAACTUAL_cad.xml (pf_20080406_cad.xml)

Campo	Descripción
Cadáver	Listado de cadáveres recibidos durante el tiempo que se solicita.

Tabla 7 Contenido del XML: PF_CAD.XML

Detalle Técnico

Listado serializado de objetos que contendrán la siguiente información:

<cadaver>

…

</cadáver>

Las Especificaciones del Archivo se describen en los anexos.

3.5.3 INE

De la información obtenida del Instituto Nacional de Estadísticas (INE), fueron recuperados los siguientes elementos:

Demografía

- ✓ Población Total
- ✓ Distribución por Sexo

Composición del Empleo

- ✓ Cantidades de Desocupados.
- ✓ Cantidad de Personas con Ocupación.
- ✓ Población total laboralmente activa.

Geografía

- ✓ Distribución de las zonas según tipo área Rural o Urbana.

La información el INE fue entregada en diferentes formatos digitales almacenados en CD.

3.6 Diseño del Área Almacenamiento Intermedio

El diseño del área almacenamiento intermedio esta conformado por los Esquemas, AD_AREATEMPORAL] y las Estructuras de Almacén de Datos Integrados. En estos dos esquemas están los distribuidos los metadatos de negocio y los metadatos técnicos. En la siguiente figura se muestra el

diseño del área almacenamiento intermedio. Para mejor entendimiento se dividieron los metadatos de negocio de los técnicos.

3.6.1 Metadatos de Negocio

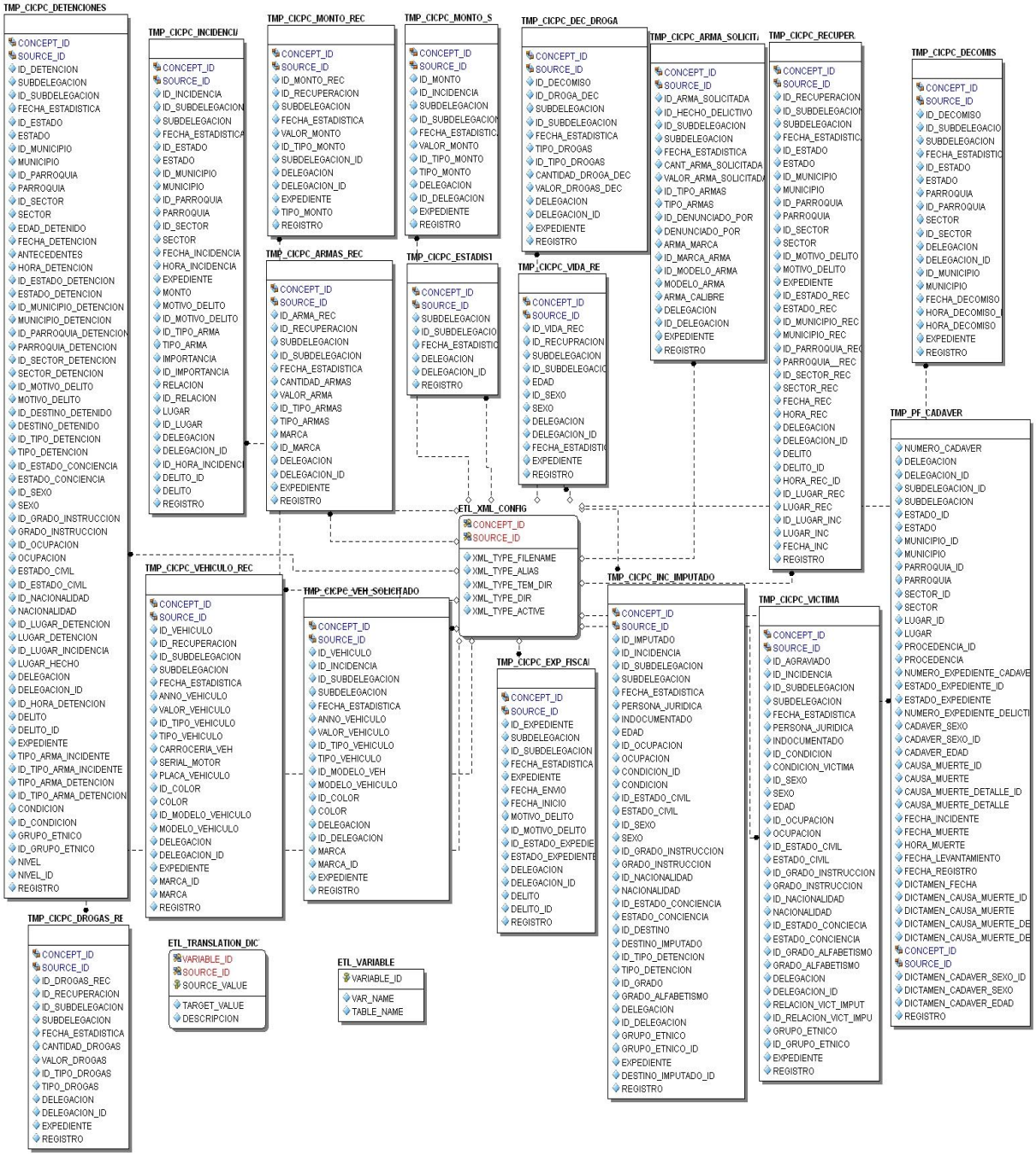

Fig. 12 Modelo de datos de los metadatos de negocio

Entre los metadatos de negocio necesarios para tener información de las fuentes de datos se pueden mencionar algunos (para más detalles ver anexos):

Nombre	Descripción
ETL_AUDIT_TRANSLATION	Mantiene un registro de todas las transformaciones que no se aplicaron con el diccionario de correspondencia. Se tiene un registro de los valores que no existen o que vienen nulos en el proceso de integración.
ETL_INTEGRATION_CONCEPT	Contiene los diferentes tipos de conceptos con que se trabaja (Hecho Delictivo, Agraviado, Imputado, Decomiso, etc.).
ETL_INTEGRATION_SOURCE	Define las Fuentes de Datos que se están integrando al Almacén de Datos.
ETL_TRANSLATION_DICTIONARY	Contiene todas las transformaciones que se le aplican a los datos de las diferentes fuentes. Este es el llamado Diccionario de Correspondencia.
ETL_VARIABLE	Define todas las variables que existen asociadas a un concepto (Ejemplo. Sexo, Edad, Delito, Causa de Muerte).

Tabla 8 Descripción de las estructuras de los metadatos de negocio.

3.6.2 Metadatos de Técnicos

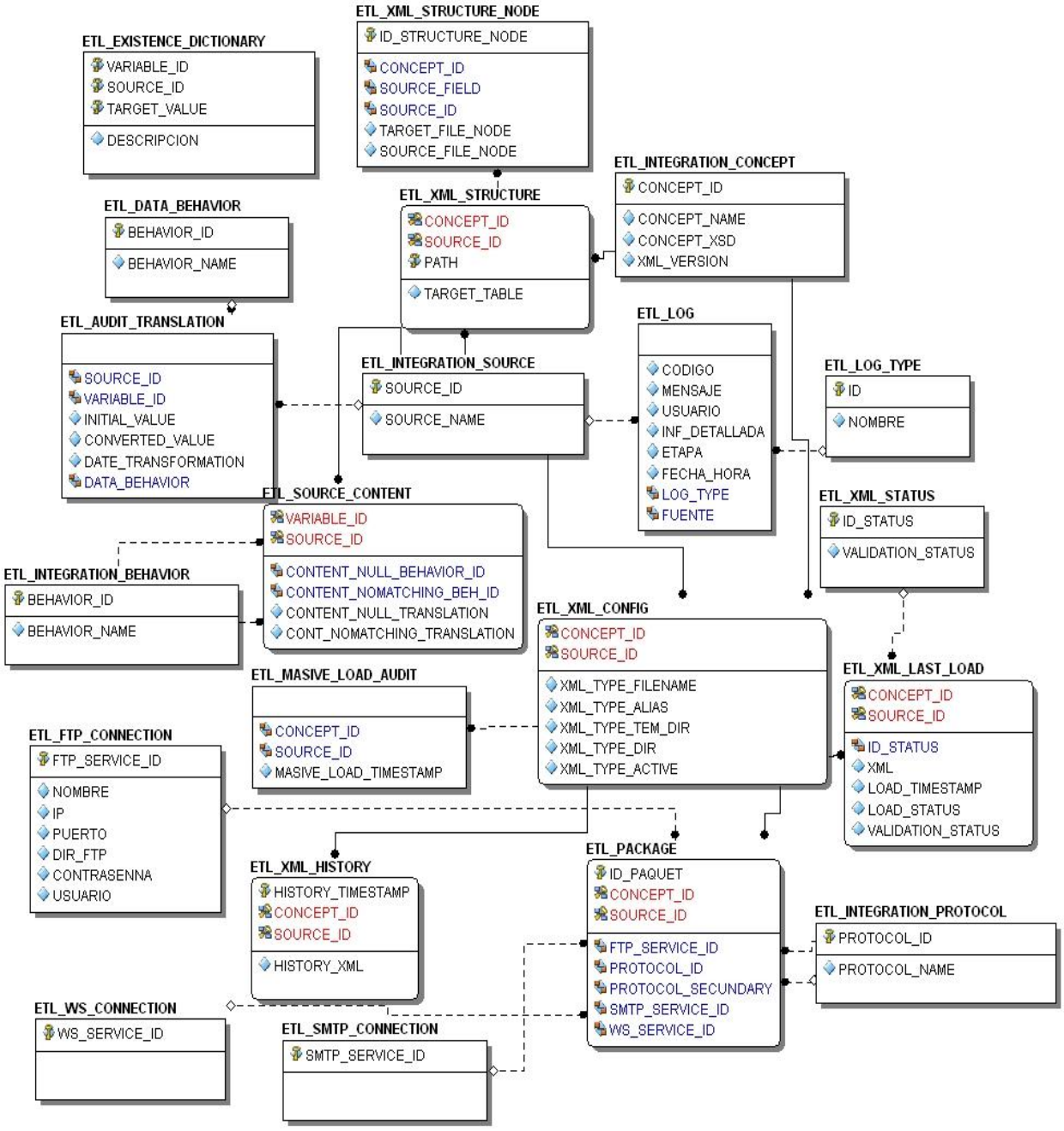

Fig. 13 Modelo de datos de los metadatos técnicos

Entre los metadatos técnicos necesarios para tener información de la ejecución del proceso ETL y de su configuración, se pueden mencionar algunos (para más detalles ver anexos):

Nombre	Descripción
ETL_XML_CONFIG	Almacena la configuración cada Archivo de Intercambio de Datos.
ETL_XML_STRUCTURE	Define la estructura de los Archivos de Integración de Datos XML.
ETL_LOG	Almacena las trazas del proceso de integración relativo a errores, alertas y casos de éxito.
ETL_FTP_CONNECTION	Configuración de las conexiones al Protocolo de Transferencia de Ficheros (FTP).
ETL_DATA_BEHAVIOR	Almacena los tipos de comportamientos o transformaciones (Inferencia, Traducción, Rechazo) que se le aplican a los datos. Solamente se implementa el comportamiento de traducción.
ETL_AUDIT_TRANSLATION	Mantiene un registro de todas las transformaciones que no se aplicaron con el diccionario de correspondencia. Se tiene un registro de los valores que no existen o que vienen nulos en el proceso de integración

Tabla 9 Descripción de las estructuras de los metadatos técnicos

3.7 Servicios de Transportación

Se utilizaron 2 servidores para dar soporte a la solución de integración:

- ✓ Un servidor FTP.
- ✓ Un servidor de ETL (Oracle Database, Warehouse Builder Data Quality).

A continuación se muestran los metadatos técnicos almacenados con las especificaciones de la conexión. En este caso se utiliza una conexión FTP por las peculiaridades del problema y las condiciones tecnológicas de la fuente de información.

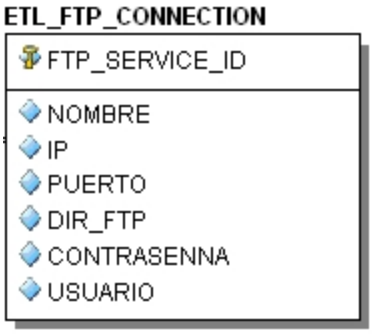

Fig. 14 Metadatos Técnicos para la conexión FTP.

La información relativa al servidor FTP es la siguiente:

Nombre del Atributo	Descripción del atributo
FTP_SERVICE_ID	Identificador de la conexión al FTP
Nombre	Nombre relativo al FTP
IP	IP por el cual se realizará la conexión
DIR_FTP	Nombre relativo por el cual se puede conectar al servidor FTP
CONTRASENNA	Contraseña por la que se realizará la conexión.
USUARIO	Usuario por el cual se realizará la conexión. Este usuario debe ser utilizado solamente para realizar el proceso de ETL

Tabla 10 Descripción de la tabla FTP_CONNECTION

3.8 Diseño del Proceso de Transformación

El OWB, como herramienta seleccionada para las ETL, proporciona un ambiente visual que permite diseñar los movimientos o transformaciones de datos.

Un Mapping[23] define una macro-transformación desde una o varias tablas hacia una o varias tablas. En el proceso de ejecución del Mapping se realizan transformaciones que no son más que funciones que reciben como entrada un conjunto finito de parámetros y devuelven como salidas un conjunto de valores que pueden ser utilizados a su vez como mismo se utiliza una tabla, por ejemplo.

Una transformación podría ser la función SUMAR que recibe como entrada dos elementos a y b y devuelve como salida un valor c. la transformación descrita tendría la siguiente forma:

TRANSFORMACIÓN: sumar (número a, número b): número c

Las transformaciones utilizadas para el proceso de integración del los datos provenientes de los Órganos de Seguridad del Estado Venezolano fueron las siguientes:

Descripción de las transformaciones utilizadas
TRANSFORMACIÓN: fecha_hora_cicpc ():
Transforma la hora proveniente de CICPC a un formato (datetime) conocido por la base de datos Oracle. El formato de fecha proveniente de CICPC es el siguiente AAAA-MM-DD THH: MM: SS.
TRANSFORMACIÓN: transformación ():
Esta transformación se encarga de definir las transformaciones según el tipo de fuente que invoca el diccionario de transformación, modificando el valor que viene de la fuente al valor se utilizará para realizar la búsqueda. Los parámetros son el identificador de la fuente, el valor que envía la fuente y a que variable hace referencia.
TRANSFORMACIÓN: transf_arma_tipo ():
Convierte el identificador del tipo de arma enviado por la fuente de datos a su traducción en el proceso de transformación permitiendo estandarizar la variable.

[23] En las bibliografías consultadas se muestra además como correspondencia.

TRANSFORMACIÓN: transf_cantidad_imputado ():
Devuelve la cantidad de imputados que están asociados a un delito en específico.
TRANSFORMACIÓN: transf_cantidad_victima ():
Devuelve la cantidad de victimas que están asociados a un delito en específico.
TRANSFORMACIÓN: transf_causa_muerte ():
Convierte el identificador de la causa de muerte enviado por la fuente de datos a su traducción en el proceso de transformación permitiendo estandarizar la variable para luego integrar con el almacén de datos.
TRANSFORMACIÓN: transf_concepto ():
Convierte el identificador de los conceptos enviado por la fuente de datos a su traducción en el proceso de transformación permitiendo definir las variables que se trabajan.
TRANSFORMACIÓN: transf_condicion ():
Convierte el identificador de la condición del imputado y el agraviado enviado por la fuente de datos a su traducción en el proceso de transformación permitiendo estandarizar la variable.
TRANSFORMACIÓN: transf_destino ():
Convierte el identificador del destino del los expedientes que son enviados por fiscalía enviado por la fuente de datos a su traducción en el proceso de transformación.
TRANSFORMACIÓN: transf_detencion ():
Convierte el identificador del tipo de detención que fue enviado por la fuente de datos a su traducción en el proceso de transformación.

Tabla 11 Descripción de las transformaciones utilizadas

En el anexo se muestran las restantes transformaciones.

3.8.1 Mappings

Para el diseño de los mappings se utilizaron diferentes tipos de transformaciones, algunas incluidas en la herramienta y otras creadas manualmente para apoyar el flujo de datos entre la fuente y el destino. Para entender en qué consisten estas transformaciones se presentan a continuación algunos ejemplos de mappings.

Entre los componentes que conforman un mapping se encuentran: las tablas fuentes y destinos, además de las transformaciones utilizadas. Toda esta información se realiza en términos de base de datos, esquemas, tablas, entre otros. A continuación se muestran los elementos que componen un mapping.

 **MAPPING:** TEM_MAPP_DECOMISO

Tablas Fuentes

- ✓ TMP_CICPC_DECOMISO
- ✓ TMP_CICPC_DEC_DROGA

Tablas Destino

- ✓ MAPPING_DECOMISO

Transformaciones utilizada:

- ✓ transformacion()
- ✓ verificar_concepto()
- ✓ verificar_organo()
- ✓ trans_fecha()
- ✓ verificar_geo()
- ✓ fecha_hora_cicpc()

Transformaciones de la herramienta

- ✓ to_date()

- ✓ joiner()
- ✓ constant()

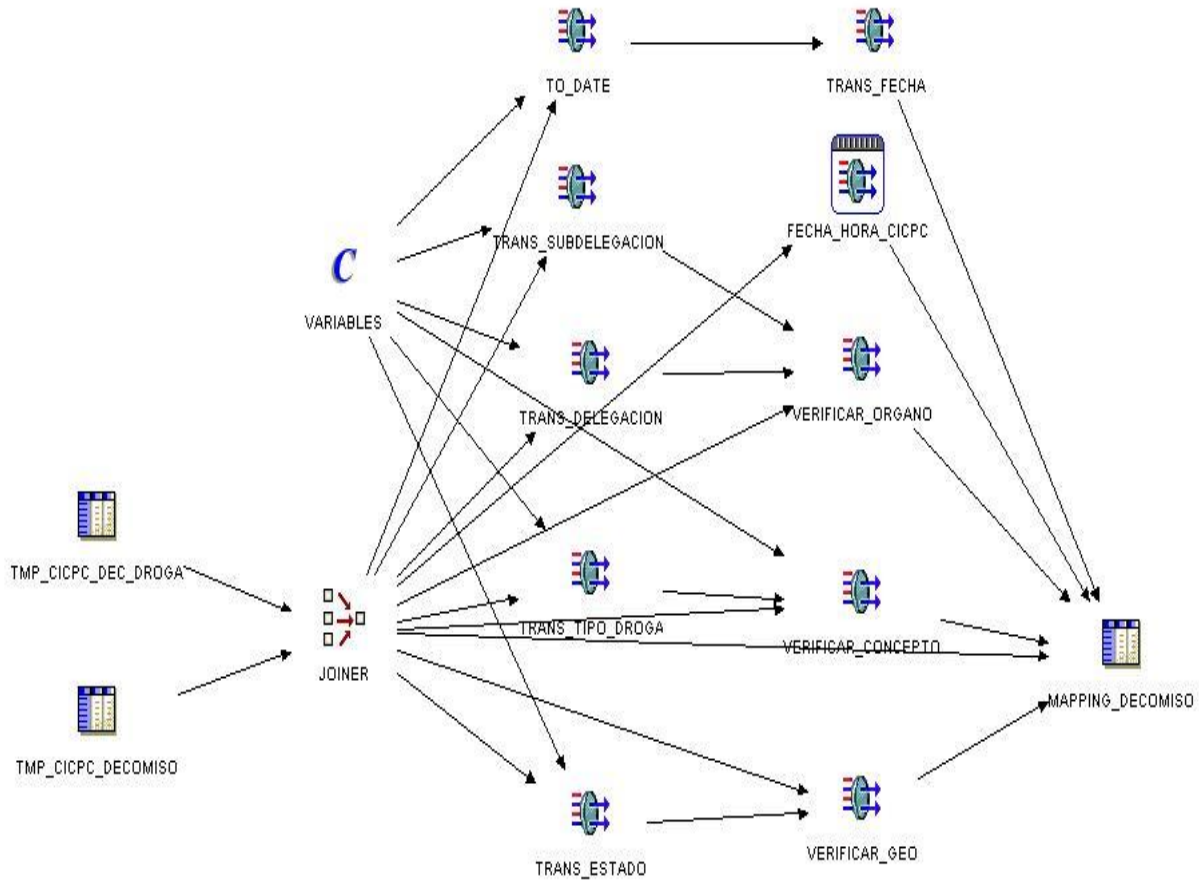

Fig. 15 Mapping: TEM_MAPP_DECOMISO

MAPPING: TEM_MAPP_DECOMISO

Tablas Fuentes:

- ✓ TMP_CICPC_DECOMISO
- ✓ TMP_CICPC_DEC_DROGA

Tabla Destino:

- ✓ MAPPING_DECOMISO

Funciones involucradas:

- ✓ transformacion()
- ✓ verificar_concepto()
- ✓ verificar_organo()
- ✓ trans_fecha()
- ✓ verificar_geo()
- ✓ fecha_hora_cicpc()

Transformaciones de la herramienta

- ✓ constant()
- ✓ joiner()
- ✓ to_date()

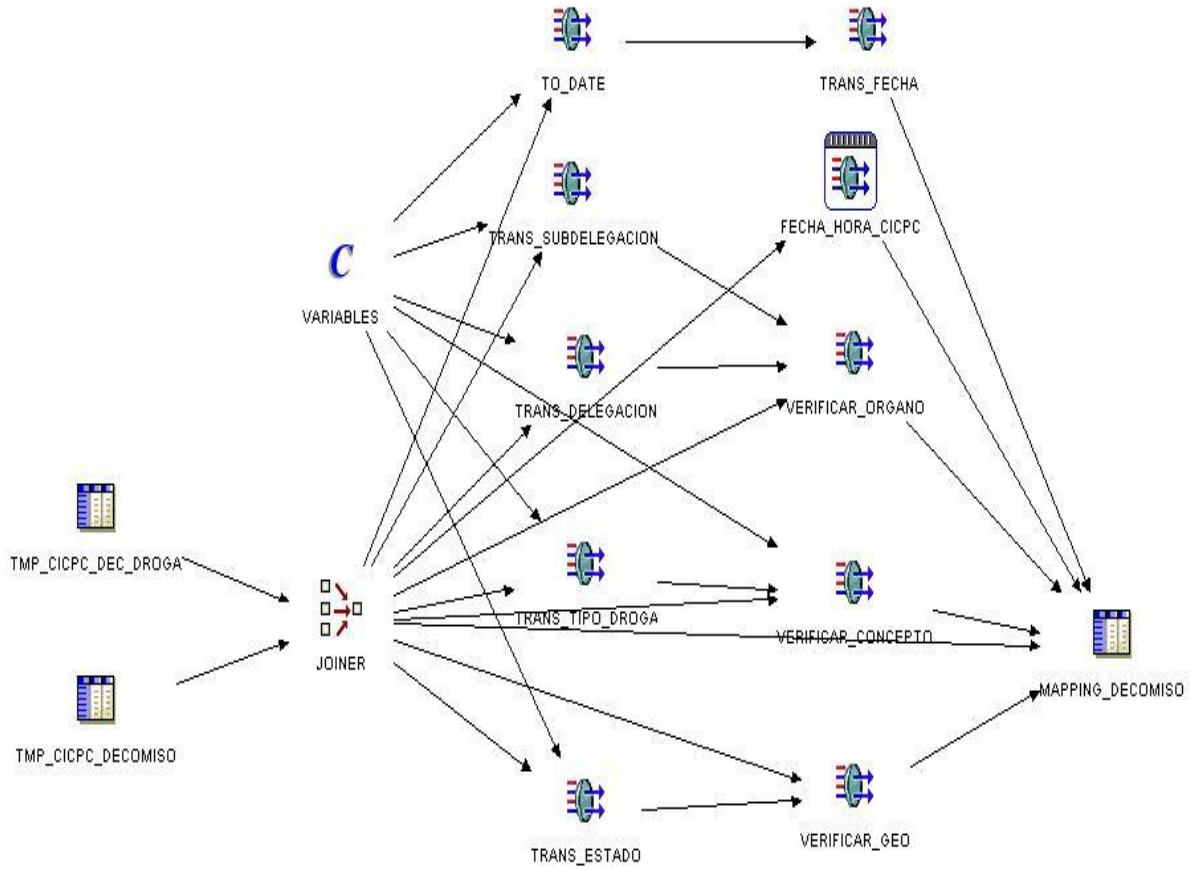

Fig. 16 Mapping: TEM_MAPP_DECOMISO

MAPPING: MAPPING_DETENCION_EADI

Tablas Fuentes:
- ✓ MAPPING_DETENCION

Tabla Destino:
- ✓ TABLA_ DETENCION

Transformaciones de la herramienta

- ✓ Imput_parameter()

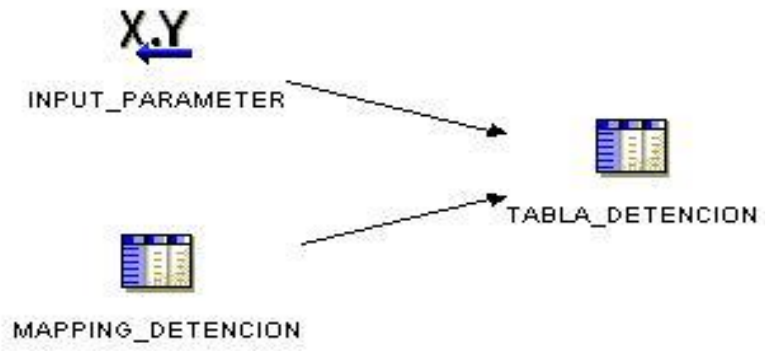

Fig. 17 MAPPING_DETENCION_EADI

El listado de todos los Mappings se muestra en los anexos.

3.9 Servicios de Administración y Operaciones

3.9.1 Activación Manual del Proceso de Integración

Para realizar el proceso de ETL de forma manual se desarrolló un paquete que brinda una interfaz cómoda y muy intuitiva para la interacción con el SINSEC. En el paquete se implementaron diferentes funcionalidades que brindan la posibilidad de verificar el proceso de ETL, ejecutarlo, auditar todo el

proceso, entre otras cosas. Se brinda la posibilidad de habilitar mensajes que se generan a medida que se va realizando la activación manual del proceso ETL. Para habilitar o deshabilitar los mensajes cuando utilice el paquete debe ejecutar:

set serveroutput on; Para habilitar los mensajes de salida a la consola.

***set serveroutput off*;** Para deshabilitar los mensajes de salida a la consola.

Nombre del Paquete: DBMS_SINSEC_ETL

Los procedimientos que contiene este paquete se describen a continuación:

Procedimiento 1:

dbms_sinsec_etl.diccionario_adicionar (*variable_id varchar2, fuente_id varchar2, valor_fuente varchar2, valor_destino varchar2, descripción varchar2*)

Esta función adiciona una nueva correspondencia al diccionario de transformaciones. Puede provocar una violación referencial y valores con tipos de datos incorrectos. Siempre se insertará en la tabla de LOG la información referente al error que ocurrió.

Descripción de los Parámetros:

VARIABLE_ID: Es el tipo de la variable a la que se le quiere adicionar una correspondencia.

FUENTE_ID: La fuente que va enviar los valores de la correspondencia.

VALOR_FUENTE: El valor en cuestión, el valor no debe exceder más de los 25 caracteres entre letras y números.

VALOR_DESTINO: Traducción que se le va a dar al valor de la fuente.

DESCRIPCION: Una pequeña descripción sobre el significado del valor.

Procedimiento 2:

dbms_sinsec_etl.diccionario_eliminar (*variable_id varchar2, fuente_id varchar2, valor_fuente varchar2*)

Esta función elimina una nueva correspondencia al diccionario de transformaciones. Como problema pueden aparecer datos no encontrados y valores con tipos de datos incorrectos. Siempre se insertará en la tabla de LOG la información referente al error que ocurrió y no se realizará la transacción.

Descripción de los Parámetros:

VARIABLE_ID: Es el tipo de la variable a la que se le quiere eliminar una correspondencia.

FUENTE_ID: La fuente que le corresponde los valores.

VALOR_FUENTE: El valor en cuestión que se desea suprimir

Procedimiento 3:

dbms_sinsec_etl.limpiar_diccionario()

Esta función elimina todas las tuplas del diccionario. Es de importancia saber que sólo los administradores podrán realizar esta operación, en caso de no estar configurado el diccionario de correspondencia, quedaría inconsistente el proceso de carga de la información. Como principal problema puede no tenerse todos los privilegios necesarios. Siempre se insertará en la tabla de LOG la información referente al error que ocurrió y no se realizará la transacción. Es importante destacar que en caso de no restablecerse los datos, no se hará ninguna transformación por lo que el proceso de carga no se realizará correctamente.

Procedimiento 4:

dbms_sinsec_etl.cargar_diccionario (*directorio varchar2, nombre_xml varchar2*)

Esta función adiciona los nuevos valores al diccionario de correspondencia contenidos en XML determinado. Entre los posibles problemas están: Falta de privilegios, Violación referencial y Valores con tipos de datos incorrectos. Se insertará en la tabla de LOG la información referente al error que ocurrió y no se realizará la transacción. Es importante destacar que la codificación del Archivo XML deberá ser UTF-8, no deberá ser modificado manualmente y debe tener la estructura definida para la carga

Descripción de los Parámetros:

DIRECTORIO: Directorio donde se encuentra el Archivo XML.

NOMBRE_XML: Nombre del Archivo en el disco.

Procedimiento 5:

dbms_sinsec_etl.activar_contenido (fuente varchar2, concepto varchar2)

Esta función activa un concepto de una fuente de datos. Entre los problemas posibles están: Violación referencial y Valores con tipos de datos incorrectos. Se insertará en la tabla de LOG la información referente al error que ocurrió y no se realizará la transacción.

Descripción de los Parámetros

FUENTE: La fuente a la cual se le activará el contenido.

CONCEPTO: El concepto que se desea activar.

Procedimiento 6:

dbms_sinsec_etl.desactivar_contenido (*fuente varchar2, concepto varchar2*)

Esta función desactiva un concepto de una fuente de datos. Entre los posibles problemas están: Violación referencial y Valores con tipos de datos incorrectos. Se insertará en la tabla de LOG la información referente al error que ocurrió y no se realizara la transacción.

Descripción de los Parámetros

FUENTE: La fuente a la cual se le desactivará el contenido.

CONCEPTO: El concepto que se desea desactivar.

Procedimiento 7:

dbms_sinsec_etl.cargar (fuente varchar2, fecha varchar2:= sysdate)

Esta función inicializa el proceso de carga de todos los archivos que se esperan de una fuente de datos. Ejemplo: dbms_sinsec_etl.cargar ('CICPC', '04072008'); para la fecha 04 de Julio del 2008. El

formato de fecha es 'ddmmyyyy'. Es importante que los Archivos de Intercambio de Datos se encuentren en el FTP, exactamente en la carpeta que corresponde según la fuente de datos. Entre los problemas posibles están: Violación referencial, Valores con tipos de datos incorrectos y Valores incorrectos. Se insertará en la tabla de LOG la información referente al error que ocurrió y no se realizará la transacción.

Descripción de los Parámetros

FUENTE: Fuente de Datos que desea cargar.

FECHA: Fecha de la carga. La fecha es importante para buscar los Archivos de Intercambio de Datos adecuados según sea la necesidad de integración. Vale notar que el nombre de los archivos está dado por la fuente, la fecha y el contenido a cargar. Si se intenta cargar un Archivo indicando una fecha incorrecta la carga fallará.

Procedimiento 8:

dbms_sinsec_etl.sys_borrar_log (dias varchar2)

Esta función elimina el registro de log. Es posible mantener intactos los registros de los N días anteriores a la fecha actual mediante el parámetro días. Se elimina de la tabla de LOG la información referente a los mensajes de los procesos

Descripción de los Parámetros

DIAS: Días anteriores que se desean mantener.

Procedimiento 9:

dbms_sinsec_etl.sys_borrar_auditoria (dias varchar2)

Esta función elimina el registro de las transformaciones (Por valor nulo y por no correspondencia). Es posible mantener intactos las auditorias de los N días anteriores a la fecha actual mediante el parámetro días. Se elimina de la tabla de Auditorias la información referente a los las transformaciones

Descripción de los Parámetros

DIAS: Días anteriores que se desean mantener.

Procedimiento 10:

dbms_sinsec_etl.actualizar_vistas ()

Esta función actualiza las Vistas Materializadas de la capa de Pre-Visualización (Punto 6.). Actualiza todas las Vistas Materializadas (VMA_*)

Procedimiento 11:

dbms_sinsec_etl.sys_borrar_xml_historicos (dias varchar2)

Esta función elimina el registro de los Archivos de Intercambio de Datos XML históricos. Es posible mantener intactos los registros de los N días anteriores a la fecha actual mediante el parámetro días. Se eliminan los XML históricos de la base de datos.

Descripción de los Parámetros

DIAS: Días anteriores que se desean mantener.

Procedimiento 12:

dbms_sinsec_etl.sys_restaurar_sistema (fecha varchar2:= sysdate)

Esta función verifica y garantiza la integridad del proceso de integración. Si ocurre un error inesperado en el proceso de integración, se aconseja llamar a esta función. Además está habilitada una Tarea que invoca a sys_restaurar_sistema a una hora determinada. Elimina rastros de datos del proceso de integración.

Descripción de los Parámetros

FECHA: Es posible verificar y restaurar solamente una fecha de integración.

3.9.2 Activación Programada del Proceso de Integración

El proceso de ejecución de las tareas que activan las transformaciones, se debe realizar en un horario que su impacto sea mínimo tanto a la fuente como al destino, es por ello la importancia de determinar el horario adecuado para realizar esta ejecución. En este caso las fuentes de datos subían los archivos XML al FTP a la 1:00 AM, por lo que a partir de las 2:00 AM y hasta la 7:00 AM el impacto del proceso de integración era ínfimo.

La activación del Proceso de Integración a través de Jobs configurados bajo el esquema AD_AREATEMPORAL (Componente 4.1) tiene las características siguientes:

JOB 42:	Garantiza la integridad de los procesos de integración del SINSEC, eliminando los registros de trazas históricos.
Esquema:	AD_AREATEMPORAL
Comandos:	dbms_sinsec_etl.SYS_BORRAR_AUDITORIA;
	dbms_sinsec_etl.SYS_BORRAR_REGISTROS;
	dbms_sinsec_etl.SYS_BORRAR_HISTORICO;
Fecha ejecución:	Diario, 2:00:00 AM

Tabla 12 Eliminación de los registros de traza histórica

JOB 142:	Activa el proceso de Integración de Datos.
Esquema:	AD_AREATEMPORAL
Comandos:	dbms_sinsec_etl.CARGAR ('PF');
	dbms_sinsec_etl.CARGAR ('CICPC');
Fecha ejecución:	Diario, 2:00:00 am

Tabla 13 Activación programada del proceso de ETL

3.10 Conclusiones del Capítulo

En este capítulo se presentaron los componentes desarrollados para lograr la integración de los datos como parte de la solución, para garantizar el tratamiento y análisis de información, así como las estructuras que quedaron implantadas dentro del SINSEC. Se describen las transformaciones utilizadas, servicios de administración de operaciones, las estructuras del área de almacenamiento intermedio así como la descripción de cada una de las fuentes de datos.

Capitulo 4: Resultados de la Investigación

4.1 Introducción

Una vez que se han desplegado todas las estructuras de la capa de ETL, se da paso a la validación de la solución, verificando que se cumplan todos los requerimientos que hacen confiable la información que se brinda al usuario funcional del sistema.

En este capítulo se presentan las pruebas y validaciones que se le realizaron a la capa de Integración de Datos.

4.2 Validación de los Procesos de ETL

Al terminar la implantación del sistema, se pasó a la etapa de pruebas, en la que se tuvo el apoyo de los usuarios finales, quienes acompañaron este proceso durante todas las actividades de prueba. Todos los que participaron en la etapa de pruebas junto al equipo de desarrollo, fueron analistas de información de Seguridad Ciudadana. Ellos fueron:

- Abg. Rocksolin Cabrera.
- Lic. Franklin Orellana.
- Ing. Johan Caldera.
- Ing. Rubén Silva.
- Lic. Francy Moreno
- Lic. Marianela Marrero.

Como resultados de estas pruebas, surgieron un conjunto de no conformidades, las que se resolvieron satisfactoriamente, quedando finalmente aceptado el producto.

Dentro del Registro de Fallas y Dificultades detectadas estaban:

1. Los usuarios del sistema no poseen un manual de usuario para el trabajo con la aplicación, por lo que deben inferir la mayoría de las funcionalidades del sistema (*Referente a la Capa de Inteligencia de Negocios*).

2. Cuando se exporta una consulta a Excel la misma presenta fallas al abrir con Microsoft office Excel 2003 (*Referente a la Capa de Inteligencia de Negocios*).
3. Cuando se seleccionan varios datos después de realizado el reporte, tiende a no actualizarse la información, se congela (*Referente a la Capa de Inteligencia de Negocios*).
4. Cuando un registro no tiene información, debe aparecer como cero “0”. Ejemplo. Si se solicita el reporte de delitos por todos los Estados, deben aparecer aún aquellos estados que no presenten delitos (*Referente a la Capa de Inteligencia de Negocios*).
5. Los gráficos del informe no se pueden realizar por separado, generando por ende, que el mismo no se adapte a las necesidades del CTAISC (*Referente a la Capa de Inteligencia de Negocios*).
6. No permite realizar reportes múltiples por regiones (*Referente a la Capa de Inteligencia de Negocios*).
7. Al momento de realizar determinado reporte, el sistema se presenta excesivamente lento y constantemente se congela, presentando errores aleatoriamente y se hace necesario cerrar la aplicación para corregir dichos errores (*Referente a la Capa de Inteligencia de Negocios*).
8. El logo del MPRIJ de la plantilla prediseñada debe estar en la parte superior de la misma (*Referente a la Capa de Inteligencia de Negocios*).
9. Además, surgieron dudas respecto a la nomenclatura utilizada, y el uso de funciones estadísticas dentro de la herramienta.

En respuesta a las No Conformidades detectadas, se trabajó en la identificación del problema y su ajuste o solución, según el caso correspondiente. Algunos de los puntos señalados se debían a desconocimiento de las herramientas, otros fueron temas de configuración y programación, los que se resolvieron.

Como parte de la Capacitación y Transferencia Tecnológica se tomaron un conjunto de datos de prueba, mediante archivos de intercambio de datos en formato XML para cada uno de las fuentes de datos integradas. Los datos integrados mostraron una gran variedad de situaciones en el amplio espectro de valores para cada variable dentro de cada archivo. Al culminar los procesos de integración, se validó que los datos contenidos en los archivos fueron depurados, ajustados según las reglas de transformación configuradas, y almacenados en las estructuras multidimensionales, quedando listos para el análisis, a

través de las herramientas de Inteligencia de Negocios. Se anexa un ejemplo de las salidas por consola, a través del mecanismo manual de invocación del proceso de integración (DBMS_SINSEC_ETL).

4.3 Carta de Aceptación del Producto

Después de ser desplegada la propuesta realizada para darle solución a las necesidades informativas del SINSEC las partes implicadas en el desarrollo del almacén de datos estuvieron involucradas.

El Acta de aceptación de la conclusión del desarrollo, implantación y capacitación de la Capa de Integración de Datos, referente a las herramientas de ETL, en el CTAISC, con Referencia CT-SW-CC-012, fue aprobada en agosto del 2008, quedando constancia de ello en el Acta de Entrega con Referencia CT–CC-022, en agosto del 2008.

albet INGENIERÍA Y SISTEMAS | PROYECTOS

ACTA DE ACEPTACION

Entro **la Republica Bolivariana de Venezuela**, actuando por organo del **Ministerio del Poder Popular para Relaciones Interiores y Justicia de la Republica Bolivariana de Venezuela**, representado en este acto por la ciudadana venezolana Salda Vielma Angulo, mayor de edad, portador de la cedula de identidad V.8.042.985 en su condicion de Gerente General, quien a los fines y efectos derivados del presente documento se denominara **"Parte Venezolana"**, de una parte y de otra parte, la **Republica de Cuba** por medio de la **Sociedad Mercantil ALBET Ingeniería y Sistemas**, conocida de forma abreviada como **ALBET, S.A.**, representada en este acto por la ciudadana cubana **Mailin Ochoa Calzadilla**, mayor de edad, portador de pasaporte Nº 0847531 en su condicion de Gerente General, quien a los fines y efectos derivados del presente documento se denominara **"Parte Cubana"**.

AMBAS PARTES, considerando que en cumplimiento del **CONTRATO E06-005-000 de Diseño e Implementacion del Centro de Comando y Control de Seguridad Ciudadana y Modelo para la Implementacion de Centros de Atencion de Emergencias 171, del Ministerio del Interior y Justicia de la República Bolivariana de Venezuela**, ha quedado concluido, en el Centro de Tratamiento y Análisis de Información de Seguridad Ciudadana (CTAISC), el desarrollo, implantacion y una parte de la transferencia tecnológica de la Capa de Extraccion, Transformación e Integración de Datos (ETL) referente a la administración y configuración de la misma. Además ha sido entregada la documentacion correspondiente.

UNICA: Considerando que ha quedado debidamente concluido lo antes mencionado y **además**, ha sido entregada la documentacion que antes se relaciona, **Las Partes** acuerdan:

1. Formalizar mediante la presente **Acta** la aceptacion del desarrollo, implantacion y capacitacion y una parte de la transferencia tecnologica de la Capa de Extraccion, Transformacion e Integración de Datos y la documentacion correspondiente, asi como su contenido en cuestion.

Y a todos los efectos legales procedentes, **Ambas Partes** suscriben la presente, en dos (2) ejemplares a un mismo tenor y efectos, en la ciudad de Caracas a los 27 días del mes de Julio del año 2008.

POR LA PARTE VENEZOLANA

Salda Vielma Angulo
Gerente General

POR LA PARTE CUBANA

Iván Maykel Cárdenas Tandron
Jefe Sistema SINSEC

Mailin Ochoa Calzadilla
Gerente General

ALBET, S.A.
Centro de Negocios Miramar, Edificio Barcelona, Oficina 322, Avenida 5ta e/176 y 78, Miramar, Playa, Ciudad Habana, Cuba
Tel/Fax: +53 (7) 837 2407
E-mail: albet@albet.cu

Referencia: CT_SW_CC_012

Fig. 18 Carta de aceptación del producto, en específico el proceso de ETL

4.3 Conclusiones del Capítulo

Al culminar la etapa de pruebas de la Capa de Integración de Datos, donde se valida con un conjunto de datos de prueba que la información integrada se muestra correctamente mediante la Capa de Visualización (BI), se pasó a la firma del Acta de Aceptación del Sistema de Información Nacional de Seguridad Ciudadana.

Conclusiones Generales

Inicialmente se determinaron las necesidades de integración del Almacén de Datos del SINSEC y una vez que se identificaron tales necesidades, se especificaron las condiciones reales de integración de los órganos de seguridad identificados, para cubrir la demanda de información y facilitar el funcionamiento del CTAISC, y se efectuó la migración de los datos de los órganos seleccionados implementando los procedimientos de ETL para cada caso.

Como resultado principal de este trabajo, se aplicaron las técnicas de integración de datos mediante las herramientas existentes logrando integrar los datos de principales Órganos de Seguridad con que cuenta el Gobierno de la República Bolivariano de Venezuela al Almacén de Datos del SINSEC, y se logró establecer un patrón de carga, que será seguido para la Integración de nuevas fuentes de datos.

.

Recomendaciones

Con vistas a lograr una mayor efectividad en el funcionamiento del SINSEC y para darle continuidad al mismo se recomienda:

1. Realizar la integración de nuevas fuentes de datos de Seguridad Ciudadana que complementen la información actual, a fin de investigar la influencia de nuevas variables que puedan incidir de alguna forma en la inseguridad de la población.
2. Mejorar la Calidad de Datos mediante Perfiles de Datos y Reglas de Corrección, una vez que las fuentes de datos tengan datos históricos.
3. Que se mantenga estable el personal que labora con el sistema, para lograr una evolución tanto en la mentalidad de los analistas, como en las necesidades de información.
4. Aplicar otras técnicas de integración de información como EII en el CTAISC para fortalecer y potenciar las capacidades del CTAISC, analizando si se requiere convertir las fuentes de datos existentes, a Fuentes Consultables (Cooperativas) con vistas a facilitar la toma de decisiones.

Referencias Bibliográficas

ALBET.S.A. (2008). *Sistema de Información Nacional de Seguridad Ciudadana.* República Bolivariana de Venezuela.

Armstrong-Smith, D. A. (2006). *Oracle Discoverer 10g Handbook.* San Francisco, California: The McGraw-Hill Companies.

C.Imhoff. (abril de 2005). Intelligent Solutions: Understanding the Three E's of Integration EAI, EII and ETL. *DM Review Magazine* .

Encinosa, L. J. *Apuntes para una historia de la informatica en cuba.*

Espinosa, R. (2007). *Evaluacion de la metodología, indicadores y estadísticas sobre seguridad cuidadana en América Latina.* Santiago Chile: Universidad del Valle.

ILPES. (1998). *Guia para la identificación, preparación y evaluación de proyectos de seguridad ciudadana con enfasis en vigilancia policial.* SANTIAGO, CHILE: (INSTITUCION), ILPES.

INE. (2006). *Encuesta de Victimización y Percepción Policial.* Caracas. Venezuela: Instituto Nacional de Estadísticas.

Inmon, W. H. (2005). *Building the Data Warehouse* (4ta Edicion ed.). Indianapolis, Indiana: Wiley Publishing, Inc.

Kimball, R., & Caserta, J. (2004). *The Data Warehouse ETL Toolkit* (1ra ed.). 10475 Crosspoint Boulevard Indianapolis: Wiley Publishing, Inc.

Kimball, R., & Ross, M. (2002). *The Data Warehouse Toolkit* (2da ed.). (R. Elliott, Ed.) Canada: Wiley Computer Publishing.

Lumpkin, G. (2007). Oracle Database 11g para Data Warehousing e Inteligencia de Negocios. *Informe Ejecutivo de Oracle* , 4.

Microsoft. (2007). Microsoft ETL Guide. *Microsoft ETL Guide* .

Morgenthal. (2000). *Enterprise Applications Integration with XML and Java. Prentice Hall PTR.* Bk&CD Rom edition.

Morgenthal. (2005). *Enterprise Information Integration: A Pragmatic Approach.* Bk&CD Rom edition.

Oficial, G. (2002). *ARTÍCULO 55 DE LA CONSTITUCIÓN DE LA REPÚBLICA BOLIVARIANA DE VENEZUELA.* República Bolivariana de Venezuela.

Pentaho. (2008). *Pentaho Commercial open Source Business Intelligence*. Recuperado el 1 de 3 de 2008, de @Pentaho: www.pentaho.com

Rivera Victoria, S. M. (2007). *El Datawarehouse y el Business Intelligence en Gobierno.* España: Universidad de León.

Russo, M. (10 de 2008). *SQLBlog*. Recuperado el 17 de 1 de 2009, de THE SQL Server Blog Spot on the Web: http://sqlblog.com/blogs/marco_russo/archive/2008/09/20/methodology-comparison-kimball-inmon-and-sqlbi.aspx

Sybase. (2008). Recuperado el 2008, de Business Intelligence and Data Management Software system including Data Warehousing: http://www.sybase.com/

Web, W. W. (14 de 10 de 2008). *W3C*. Recuperado el 26 de 1 de 2009, de Consorcio World Wide Web: http://www.w3c.es/

Bibliografía

ALBET.S.A. (2008). *Sistema de Información Nacional de Seguridad Ciudadana.* República Bolivariana de Venezuela.

Armstrong-Smith, D. A. (2006). *Oracle Discoverer 10g Handbook.* San Francisco, California: The McGraw-Hill Companies.

C.Imhoff. (abril de 2005). Intelligent Solutions: Understanding the Three E's of Integration EAI, EII and ETL. *DM Review Magazine* .

Encinosa, L. J. *Apuntes para una historia de la informatica en cuba.*

Espinosa, R. (2007). *Evaluacion de la metodología, indicadores y estadísticas sobre seguridad cuidadana en América Latina.* Santiago Chile: Universidad del Valle.

ILPES. (1998). *Guia para la identificación, preparación y evaluación de proyectos de seguridad ciudadana con enfasis en vigilancia policial.* SANTIAGO, CHILE: (INSTITUCION), ILPES.

INE. (2006). *Encuesta de Victimización y Percepción Policial.* Caracas. Venezuela: Instituto Nacional de Estadísticas.

Inmon, W. H. (2005). *Building the Data Warehouse* (4ta Edicion ed.). Indianapolis, Indiana: Wiley Publishing, Inc.

Kimball, R., & Caserta, J. (2004). *The Data Warehouse ETL Toolkit* (1ra ed.). 10475 Crosspoint Boulevard Indianapolis: Wiley Publishing, Inc.

Kimball, R., & Ross, M. (2002). *The Data Warehouse Toolkit* (2da ed.). (R. Elliott, Ed.) Canada: Wiley Computer Publishing.

Lumpkin, G. (2007). Oracle Database 11g para Data Warehousing e Inteligencia de Negocios. *Informe Ejecutivo de Oracle* , 4.

Microsoft. (2007). Microsoft ETL Guide. *Microsoft ETL Guide* .

Morgenthal. (2000). *Enterprise Applications Integration with XML and Java. Prentice Hall PTR.* Bk&CD Rom edition.

Morgenthal. (2005). *Enterprise Information Integration: A Pragmatic Approach.* Bk&CD Rom edition.

Oficial, G. (2002). *ARTÍCULO 55 DE LA CONSTITUCIÓN DE LA REPÚBLICA BOLIVARIANA DE VENEZUELA.* República Bolivariana de Venezuela.

Pentaho. (2008). *Pentaho Commercial open Source Business Intelligence*. Recuperado el 1 de 3 de 2008, de @Pentaho: www.pentaho.com

Rivera Victoria, S. M. (2007). *El Datawarehouse y el Business Intelligence en Gobierno.* España: Universidad de León.

Russo, M. (10 de 2008). *SQLBlog*. Recuperado el 17 de 1 de 2009, de THE SQL Server Blog Spot on the Web: http://sqlblog.com/blogs/marco_russo/archive/2008/09/20/methodology-comparison-kimball-inmon-and-sqlbi.aspx

Sybase. (2008). Recuperado el 2008, de Business Intelligence and Data Management Software system including Data Warehousing: http://www.sybase.com/

Web, W. W. (14 de 10 de 2008). *W3C*. Recuperado el 26 de 1 de 2009, de Consorcio World Wide Web: http://www.w3c.es/

Anexos

Anexo 1. Descripción de todas las transformaciones utilizadas.

Descripción de las transformaciones utilizadas
TRANSFORMACIÓN: fecha_hora_cicpc ():
Transforma la hora proveniente de CICPC a un formato de Datetime conocido por Oracle. El formato de fecha proveniente de CICPC es el siguiente AAAA-MM-DD THH: MM: SS.
TRANSFORMACIÓN: transformación ():
Esta transformación se encarga de definir las transformaciones según el tipo de fuente que invoca el diccionario de transformación, modificando el valor que viene de la fuente al valor se utilizará para realizar la búsqueda. Los parámetros son el identificador de la fuente, el valor que envía la fuente y a que variable hace referencia.
TRANSFORMACIÓN: transf_arma_tipo ():
Convierte el identificador del tipo de arma enviado por la fuente de datos a su traducción en el proceso de transformación permitiendo estandarizar la variable.
TRANSFORMACIÓN: transf_cantidad_imputado ():
Devuelve la cantidad de imputados que están asociados a un delito en específico.
TRANSFORMACIÓN: transf_cantidad_victima ():
Devuelve la cantidad de víctimas que están asociados a un delito en específico.
TRANSFORMACIÓN: transf_causa_muerte ():
Convierte el identificador de la causa de muerte enviado por la fuente de datos a su traducción en el proceso de transformación permitiendo estandarizar la variable para luego integrar con el

almacén de datos.
TRANSFORMACIÓN: transf_concepto ():
Convierte el identificador de los conceptos enviado por la fuente de datos a su traducción en el proceso de transformación permitiendo definir las variables que se trabajan.
TRANSFORMACIÓN: transf_condicion ():
Convierte el identificador de la condición del imputado y el agraviado enviado por la fuente de datos a su traducción en el proceso de transformación permitiendo estandarizar la variable.
TRANSFORMACIÓN: transf_destino ():
Convierte el identificador del destino del los expedientes que son enviados por fiscalía enviado por la fuente de datos a su traducción en el proceso de transformación.
TRANSFORMACIÓN: transf_detencion ():
Convierte el identificador del tipo de detención que fue enviado por la fuente de datos a su traducción en el proceso de transformación.
TRANSFORMACIÓN: transf_edad ():
Convierte el identificador de la edad enviado por la fuente de datos a los rangos que se trabajan en el proceso de transformación y que es también utilizada por el almacén.
TRANSFORMACIÓN: transf_estado_civil ():
Convierte el identificador del estado civil enviado por la fuente de datos a su traducción en el proceso de transformación.
TRANSFORMACIÓN: transf_estado_conciencia ():
Convierte el identificador del estado de conciencia con que esta una persona en un estado de la investigación o cuando ocurrió un hecho determinado y que es enviado por la fuente de datos

para realizar su traducción en el proceso de transformación.
TRANSFORMACIÓN: transf_estado_exp ():
Convierte el identificador del estado del expediente abierto por la fiscalía que envió la fuente de datos a su traducción en el proceso de transformación.
TRANSFORMACIÓN: transf_fecha ():
Convierte una fecha enviado por la fuente de datos a su traducción en el proceso de transformación permitiendo que el almacén de datos entienda su significado.
TRANSFORMACIÓN: transf_grupo_etnico ():
Convierte el identificador del grupo étnico enviado por la fuente de datos a su traducción en el proceso de transformación.
TRANSFORMACIÓN: transf_lugar ():
Convierte el identificador del lugar donde se quiera hacer referencia y que ha sido enviado por la fuente de datos a su traducción en el proceso de transformación.
TRANSFORMACIÓN: transf_nacionalidad ():
Convierte el identificador de la nacionalidad enviado por la fuente de datos a su traducción en el proceso de transformación.
TRANSFORMACIÓN: transf_nivel_edu ():
Convierte el identificador del nivel educacional de cualquier persona que se desee (como puede ser los imputados, los agraviados o las víctimas) enviado por la fuente de datos a su traducción en el proceso de transformación.
TRANSFORMACIÓN: transf_ocupacion ():
Convierte el identificador de la ocupación de cualquier persona que se desee (como puede ser

los imputados, los agraviados o las víctimas) enviado por la fuente de datos a su traducción en el proceso de transformación.
TRANSFORMACIÓN: transf_organo ():
Convierte el identificador del Órgano de seguridad (Delegación y Subdelegación) enviado por la fuente de datos a su traducción en el proceso de transformación.
TRANSFORMACIÓN: transf_parte_dia ():
Convierte el identificador de la parte del día en que ocurrió el hecho delictivo enviado por la fuente de datos a su traducción en el proceso de transformación. Esta función recibe la hora del día que se realizó el hecho y se devuelve como resultado la parte del día.
TRANSFORMACIÓN: transf_procedencia ():
Convierte el identificador de la procedencia del imputado enviado por la fuente de datos a su traducción en el proceso de transformación.
TRANSFORMACIÓN: transf_relacion_vi ():
Convierte el identificador de la relación de la víctima con el imputado enviado por la fuente de datos a su traducción en el proceso de transformación.
TRANSFORMACIÓN: transf_sector ():
Convierte el identificador del sector enviado por la fuente de datos a su traducción en el proceso de transformación.
TRANSFORMACIÓN: transf_sexo ():
Convierte el identificador del sexo enviado por la fuente de datos a su traducción en el proceso de transformación.
TRANSFORMACIÓN: transf_vehiculo ():

Convierte el identificador del vehículo enviado por la fuente de datos a su traducción en el proceso de transformación.
TRANSFORMACIÓN: verificar_delito ():
Verifica la existencia de los delitos y devuelve la codificación definida en el diccionario de correspondencia.
TRANSFORMACIÓN: verificar_geo ()
Verifica que una localización (elemento geo-referenciado), dada por las combinaciones de Estados, Municipios, Parroquias, Sectores existen, y tienen coherencia. En caso de existir devuelve una codificación única.
TRANSFORMACIÓN: verificar_organo ():
De las fuente de información existen algunas que tiene una distribución de los órganos de seguridad por el cual fluye la información jerárquicamente por lo que se debe verificar que exista ese órgano con su respectiva jerarquía y en caso que exista devolver la codificación con se almacena en el diccionario de correspondencia.
TRANSFORMACIÓN: verificar_arma ():
Verifica la existencia de un tipo de arma específico, en caso de que exista devuelve una nueva llave que esta conformada por la clasificación del arma y dentro de esa clasificación el tipo de arma específica.

Anexo 1 Ejemplo de invocación del proceso de integración (DBMS_SINSEC_ETL)

```
Cargando Patologia Forense

SQL> set serveroutput on;

SQL> exec dbms_sinsec_etl.cargar('PF');

--

INICIO ... 24/07/08 04:34:06,211834000 PM -04:00

--

220 FTP del CTAISC.

331 Please specify the password.

230 Login successful.

250 Directory successfully changed.

227 Entering Passive Mode (172,16,100,52,135,148)

150 Opening BINARY mode data connection for pf_20080724_cad.xml (14621 bytes).

226 File send OK.

227 Entering Passive Mode (172,16,100,52,98,64)

250 Delete operation successful.

221 Goodbye.

--

--  DESCARGA EXITOSA DE ARCHIVOS HACIA AD_XML [REF.2] --24/07/08

04:34:06,234748000 PM -04:00
```

```
--

220 FTP del CTAISC.

331 Please specify the password.

230 Login successful.

250 Directory successfully changed.

221 Goodbye.

--

--  DESCARGA EXITOSA DE ARCHIVOS HACIA AREA_TEMPORAL [REF.3] --24/07/08

04:34:06,251057000 PM -04:00

--

--

--  CARGA EXITOSA DE ARCHIVOS HACIA LA BD AREA_TEMPORAL [REF.4] --24/07/08

04:34:06,268522000 PM -04:00

--

Audit run id = 25427

Return result = OK

Execution status = COMPLETE

No. task errors = 0

No. task warnings = 0

No. errors = 0
```

```
No. selected = 7

No. inserted = 7

No. updated  = 0

No. deleted  = 0

No. discarded= 0

No. merged   = 0

No. corrected= 0

Audit run id = 25435

Return result = OK

Execution status = COMPLETE

No. task errors = 0

No. task warnings = 0

No. errors = 0

No. selected = 7

No. inserted = 7

No. updated  = 0

No. deleted  = 0

No. discarded= 0

No. merged   = 0

No. corrected= 0
```

```
----

SE HAN EJECUTADO TODOS LOS MAPPINGS, Y SE HA TRANFERIDO DATOS HACIA EL
ESQUEMA

TEMPORAL

----

Audit run id = 25444

Return result = OK

Execution status = COMPLETE

No. task errors = 0

No. task warnings = 0

No. errors = 0

No. selected = 7

No. inserted = 7

No. updated  = 0

No. deleted  = 0

No. discarded= 0

No. merged   = 0

No. corrected= 0

Audit run id = 25452

Return result = OK

Execution status = COMPLETE
```

```
No. task errors = 0

No. task warnings = 0

No. errors = 0

No. selected = 7

No. inserted = 7

No. updated  = 0

No. deleted  = 0

No. discarded= 0

No. merged   = 0

No. corrected= 0

----

SE HAN EJECUTADO TODOS LOS MAPPINGS, Y SE HA TRANFERIDO DATOS HACIA EL
ALMACEN

----

--

FIN  ... 24/07/08 04:34:08,998283000 PM -04:00

--



Procedimiento PL/SQL terminado correctamente.
```

Anexo 2 Procesos de Transformación (Mappings)

A continuación se muestran todas las transformaciones que se modelaron el la herramienta. Con los respectivos componentes que la conforman.

Fuente del Mapping

ESQUEMA: AD_AREATEMPORAL

Destino de los Mapping:

ESQUEMA: EADI

MAPPING: TEM_MAPP_INC_ARMA

Tablas Fuentes:
- ✓ TMP_CICPC_INCIDENCIA
- ✓ TMP_CICPC_ARMA_SOLICITADA

Tabla Destino:
- ✓ MAPPING_INCIDENTE_MONTO

Funciones involucradas:
- ✓ transformación.
- ✓ trans_fecha.
- ✓ verificar_arma.
- ✓ verificar_concepto.
- ✓ verificar_delito.
- ✓ verificar_geo.

- ✓ verificar_organo.

Imagen

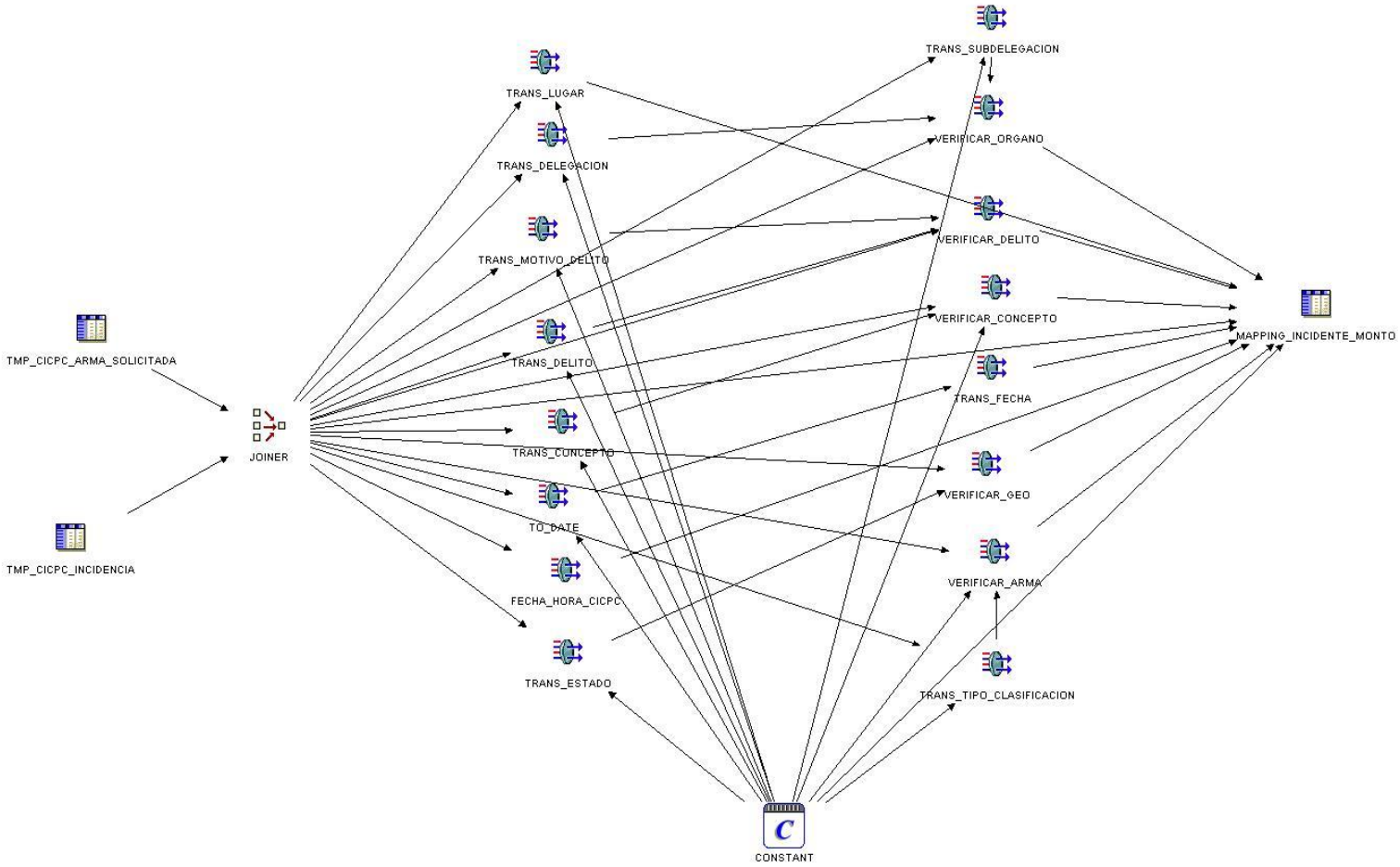

Fig. 19 TEM_MAPP_INC_ARMA

MAPPING: TEM_MAPP_INC_MONTO

Tablas Fuentes:

- ✓ TMP_CICPC_INCIDENCIA
- ✓ TMP_CICPC_MONTO_SOLICITADO

Tabla Destino:

- ✓ MAPPING_INCIDENTE_MONTO

Funciones involucradas:

- ✓ transformacion
- ✓ trans_fecha
- ✓ verificar_arma
- ✓ verificar_concepto
- ✓ verificar_delito
- ✓ verificar_geo
- ✓ verificar_organo
- ✓ verificar_arma
- ✓ fecha_hora_cicpc

Imagen

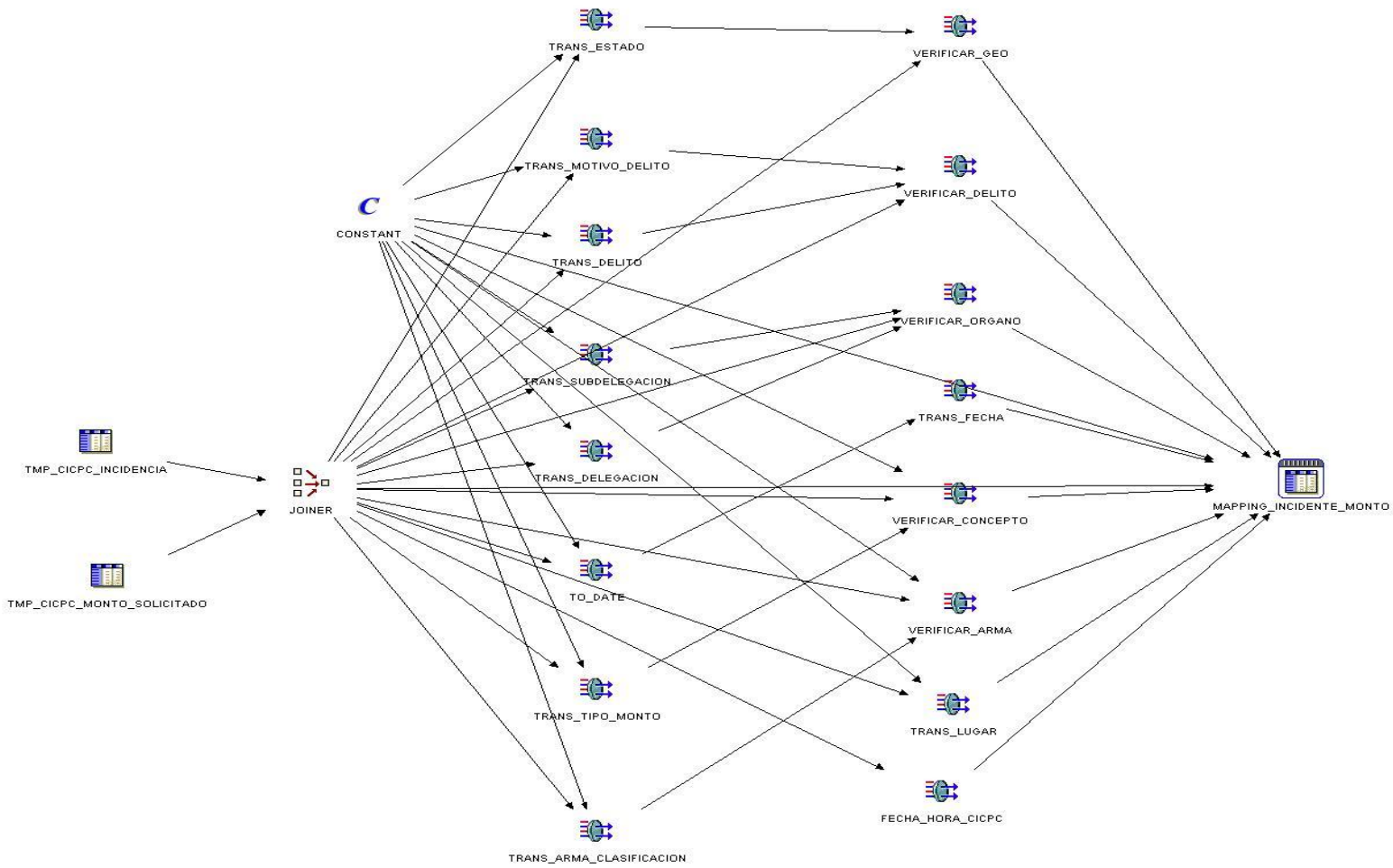

Fig. 20 TEM_MAPP_INC_MONTO

MAPPING: TEM_MAPP_INC_VEHICULO

Tablas Fuentes:

- ✓ TMP_CICPC_INCIDENCIA
- ✓ TMP_CICPC_VEH_SOLICITADO

Tabla Destino:

- ✓ MAPPING_INCIDENTE_MONTO

Funciones involucradas:

- ✓ transformación
- ✓ trans_fecha
- ✓ verificar_arma
- ✓ verificar_concepto
- ✓ verificar_delito
- ✓ verificar_geo
- ✓ verificar_organo
- ✓ verificar_arma
- ✓ fecha_hora_cicpc
- ✓ verificar_vehiculo

Imagen

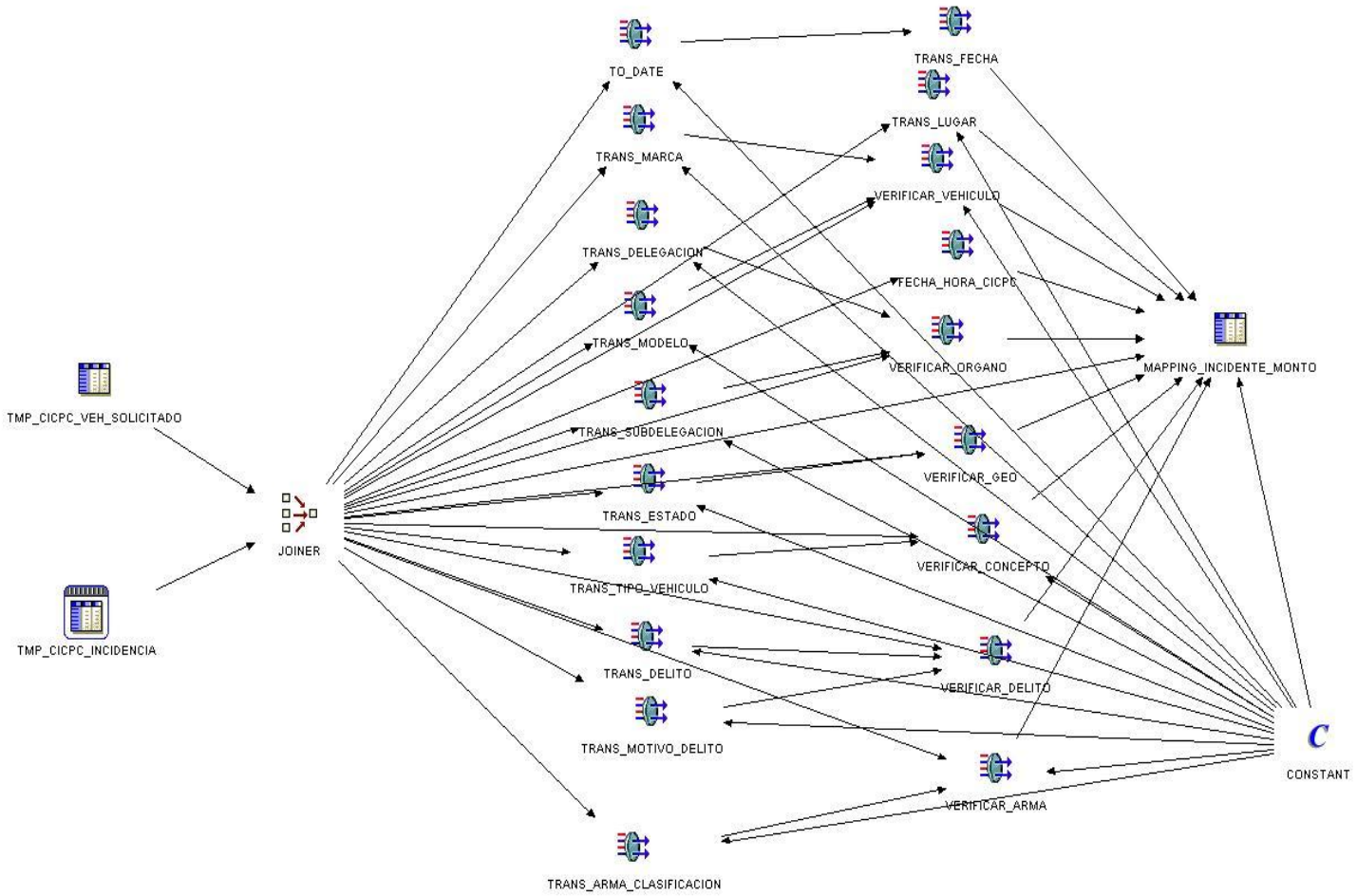

Fig. 21 TEM_MAPP_INC_VEHICULO

MAPPING: TEM_MAPP_INC_VICTIMA

Tablas Fuentes:

- ✓ TMP_CICPC_INCIDENCIA
- ✓ TMP_CICPC_VICTIMA

Tabla Destino:

- ✓ MAPPING_INCIDENTE_VICTIMA

Funciones involucradas:

- ✓ Transformacion
- ✓ trans_fecha
- ✓ verificar_arma
- ✓ verificar_delito
- ✓ verificar_geo
- ✓ verificar_organo,
- ✓ verificar_nacionalidad,
- ✓ verificar_ocupacion,
- ✓ verificar_arma
- ✓ fecha_hora_cicpc
- ✓ verificar_vehiculo
- ✓ verificar_ocupacion,

Imagen

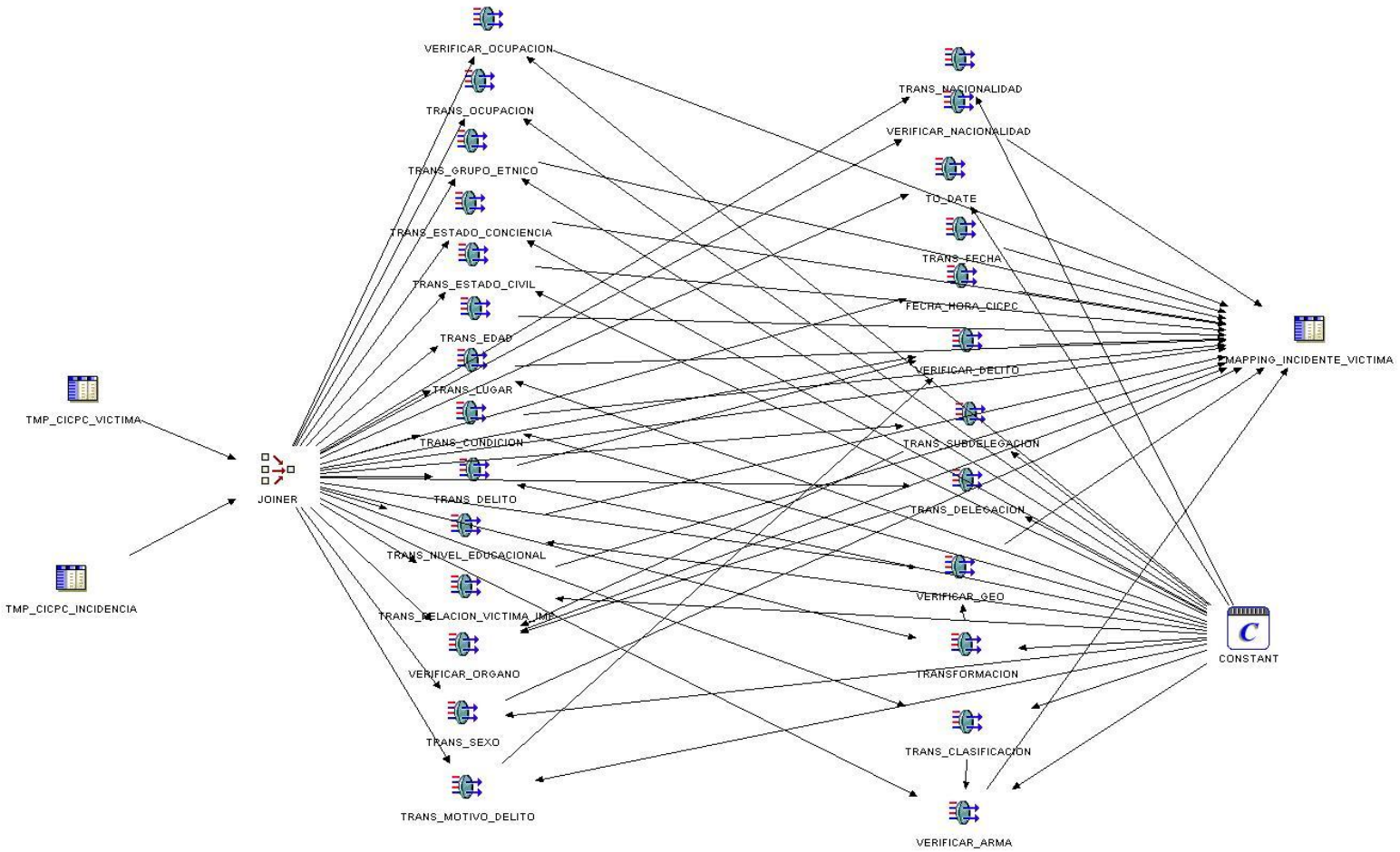

Fig. 22 TEM_MAPP_INC_VICTIMA

MAPPING: TEM_MAPP_CADAVER

Tablas Fuentes:

- ✓ TMP_PF_CADAVER

Tabla Destino:

- ✓ MAPPING_INVESTIGACION_FORENSE

Funciones involucradas:

- ✓ transformacion,
- ✓ verificar_muerte
- ✓ trans_fecha
- ✓ verificar_concepto
- ✓ verificar_geo
- ✓ verificar_organo,
- ✓ verificar_nacionalidad,
- ✓ verificar_ocupacion,
- ✓ fecha_hora_pf
- ✓ verificar_ocupacion,

Imagen

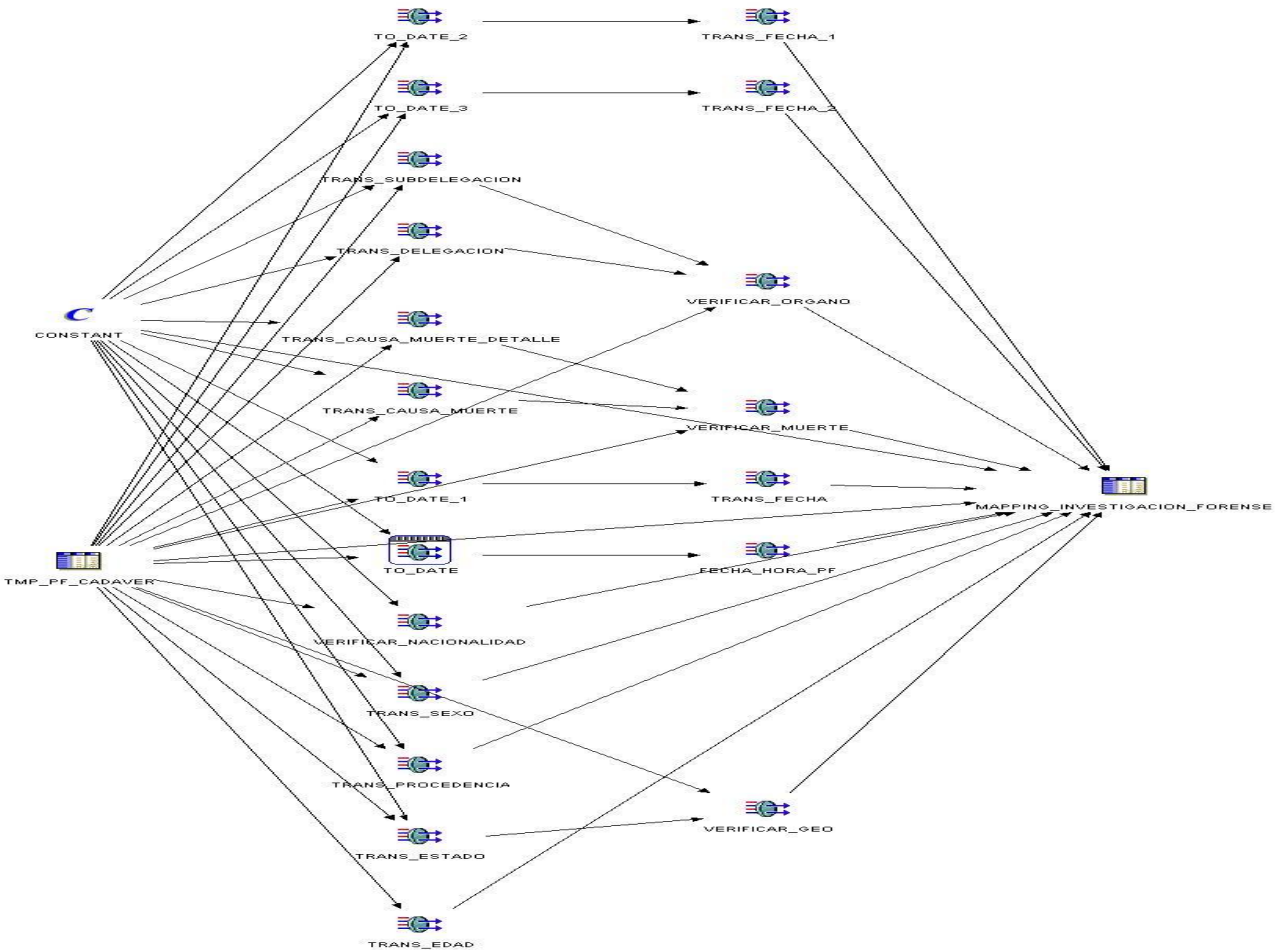

Fig. 23 TEM_MAPP_CADAVER

MAPPING: TEM_MAPP_REC_VEHICULO

Tablas Fuentes:

- ✓ TMP_CICPC_RECUPERACION
- ✓ TMP_CICPC_VEHICULO_REC

Tabla Destino:

- ✓ MAPPING_RECUPERACION

Funciones involucradas:

- ✓ transformacion,
- ✓ trans_fecha
- ✓ verificar_concepto
- ✓ verificar_geo
- ✓ verificar_organo,
- ✓ fecha_hora_cicpc
- ✓ verificar_delito

Imagen

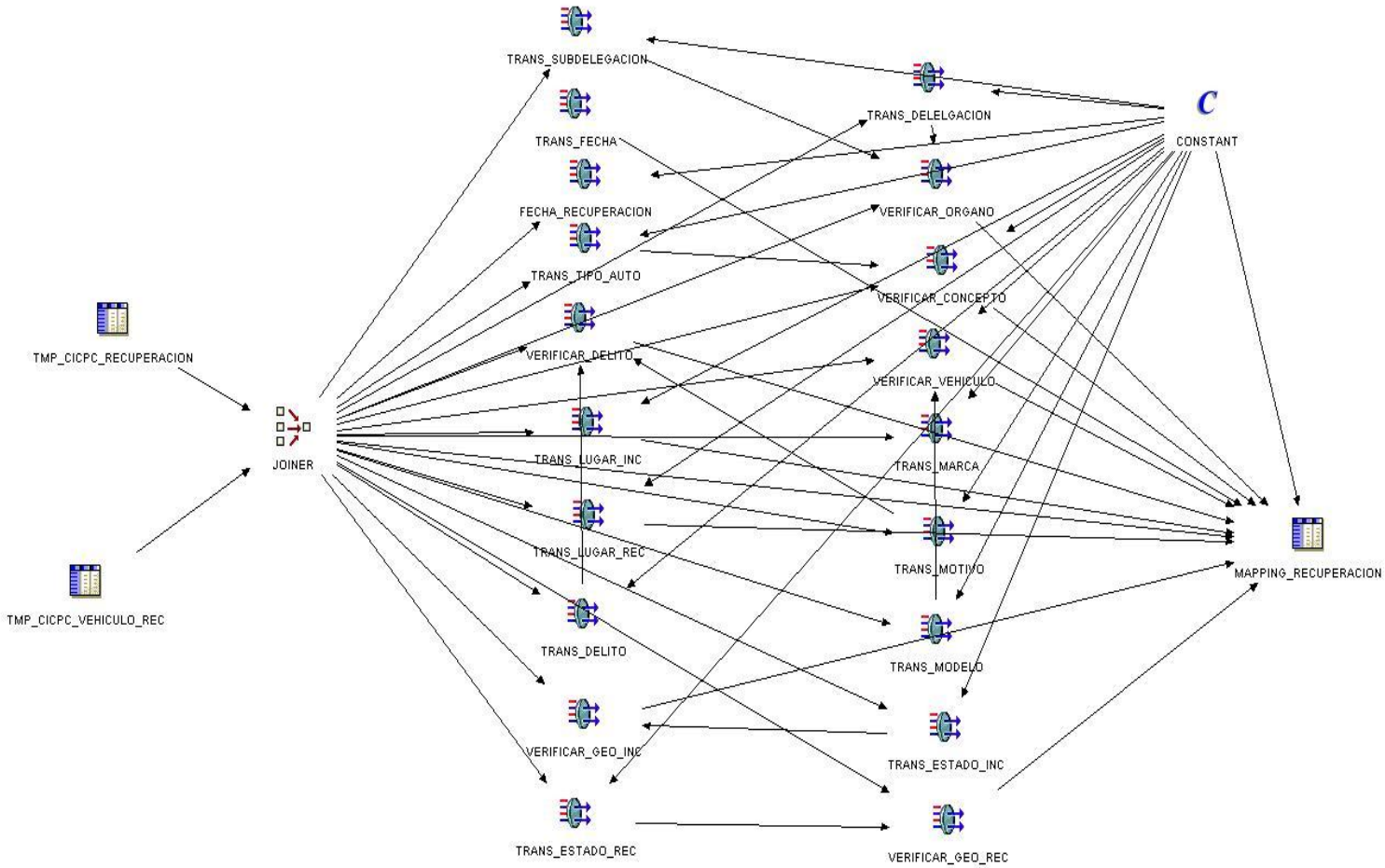

Fig. 24 TEM_MAPP_REC_VEHICULO

MAPPING: TEM_MAPP_REC_DROGAS

Tablas Fuentes:

- ✓ TMP_CICPC_RECUPERACION
- ✓ TMP_CICPC_DROGAS_REC

Tabla Destino:

- ✓ MAPPING_RECUPERACION

Funciones involucradas:

- ✓ transformacion,
- ✓ trans_fech
- ✓ verificar_concepto
- ✓ verificar_geo
- ✓ verificar_organo
- ✓ fecha_hora_cicpc
- ✓ verificar_delito

Imagen

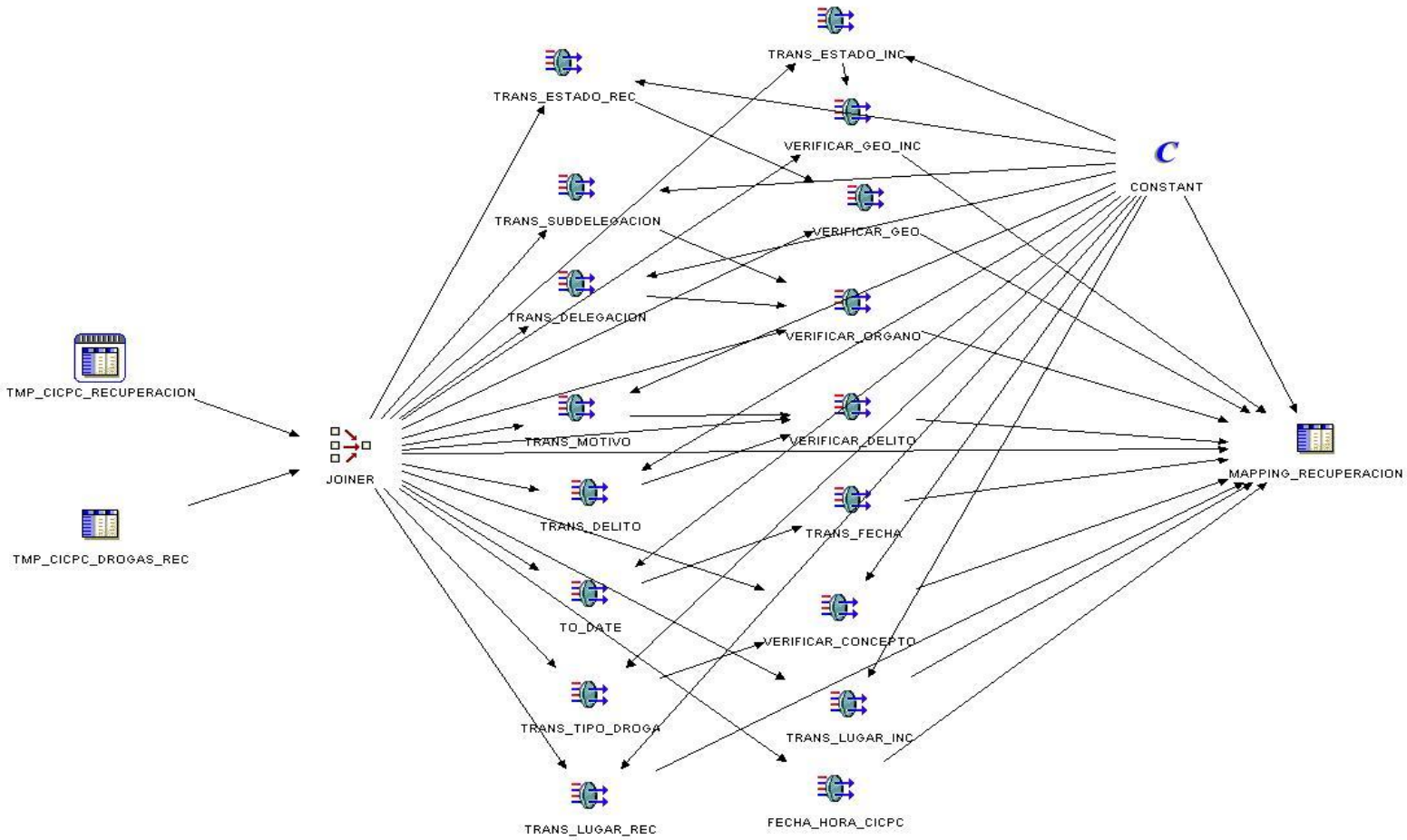

Fig. 25 TEM_MAPP_REC_DROGA

MAPPING: TEM_MAPP_REC_MONTO

Tablas Fuentes:

- ✓ TMP_CICPC_RECUPERACION
- ✓ TMP_CICPC_MONTO_REC

Tabla Destino:

- ✓ MAPPING_RECUPERACION

Funciones involucradas:

- ✓ Transformación
- ✓ trans_fecha
- ✓ verificar_concepto
- ✓ verificar_geo
- ✓ verificar_organo,
- ✓ fecha_hora_cicpc
- ✓ verificar_delito

Imagen

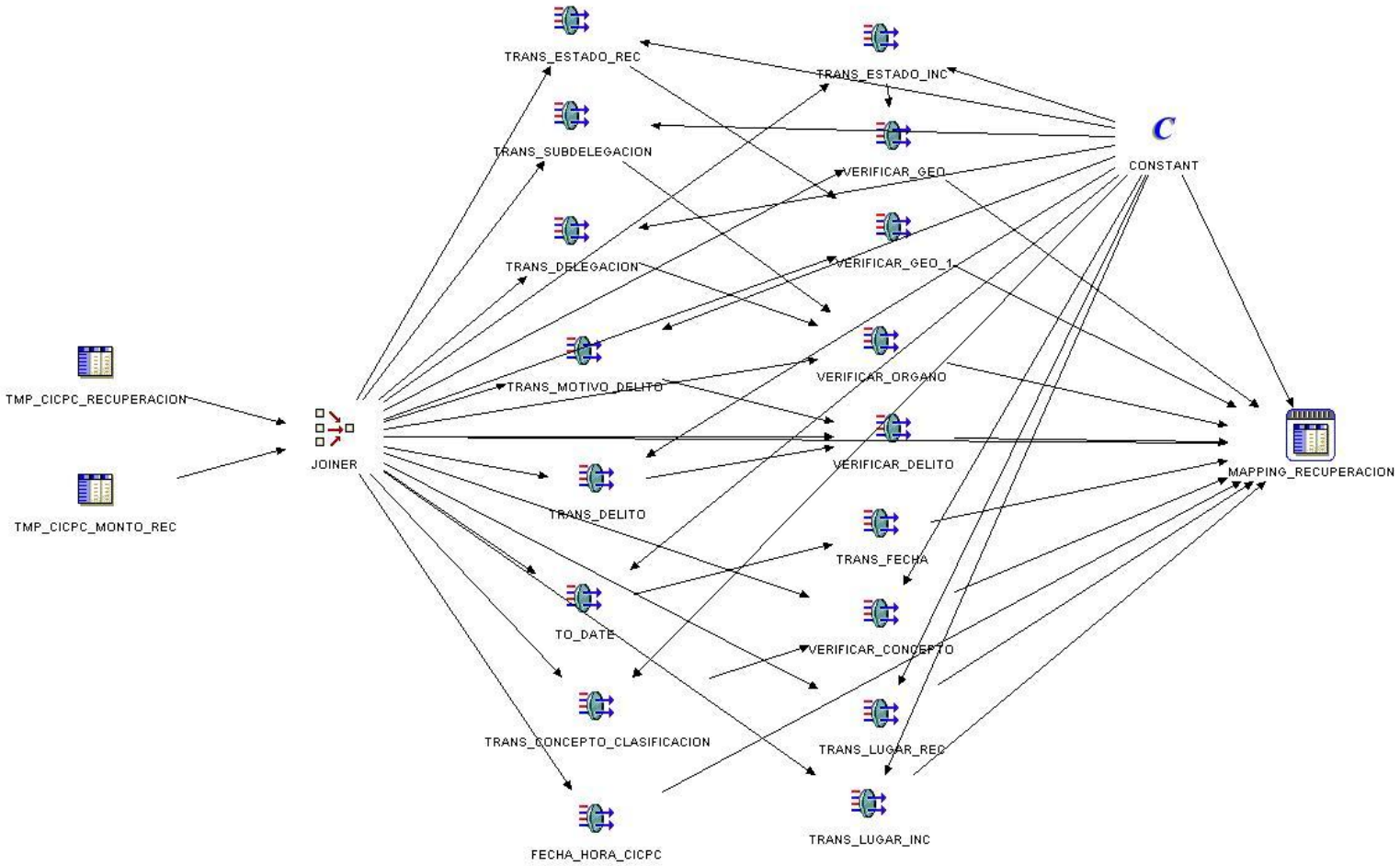

Fig. 26 TEM_MAPP_REC_MONTO

MAPPING: TEM_MAPP_REC_VIDA

Tablas Fuentes:

- ✓ TMP_CICPC_RECUPERACION
- ✓ TMP_CICPC_VIDA_REC

Tabla Destino:

- ✓ MAPPING_RESCATE_PERSONA

Funciones involucradas:

- ✓ transformacion
- ✓ trans_fecha
- ✓ verificar_concepto
- ✓ verificar_geo
- ✓ verificar_organo
- ✓ trans_edad

Imagen:

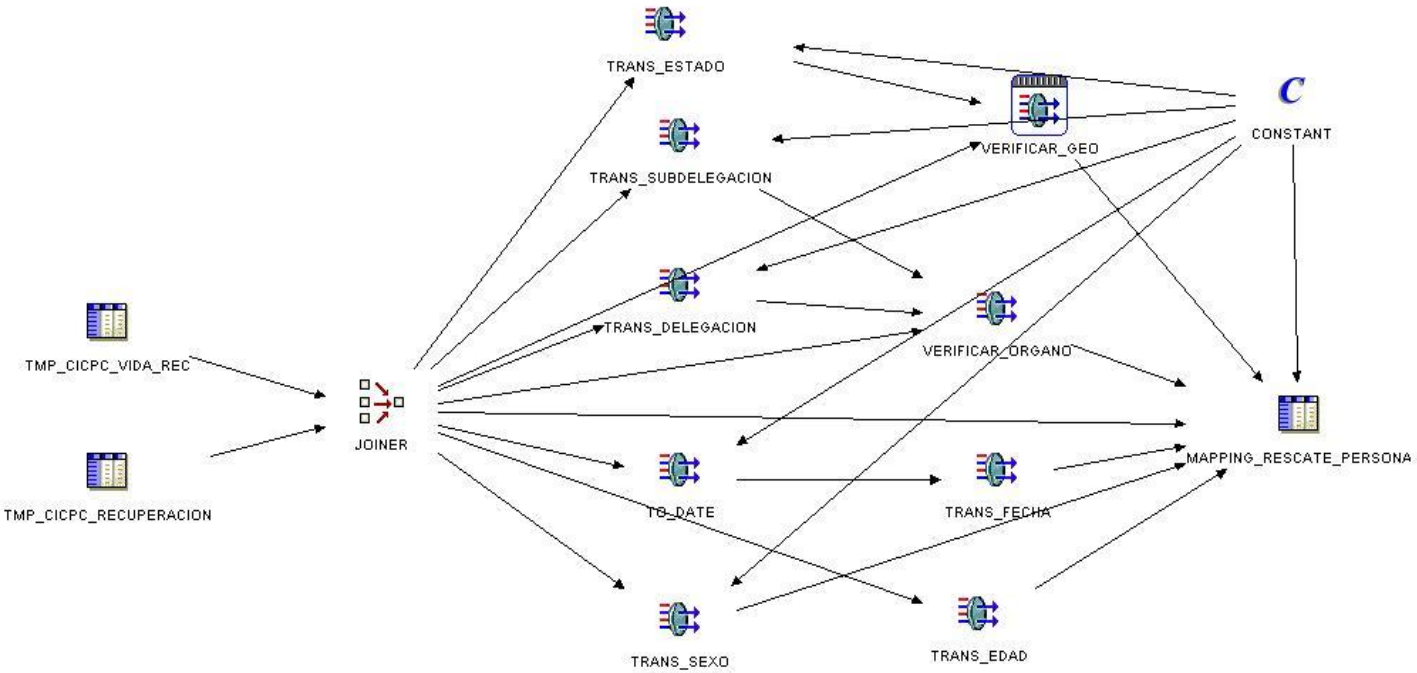

Fig. 27 TEM_MAPP_REC_VIDA

MAPPING: MAPPING_DECOMISO_EADI

Tablas Fuentes:
- ✓ MAPPING_DECOMISO

Tabla Destino:
- ✓ TABLA_DECOMISO

Imagen:

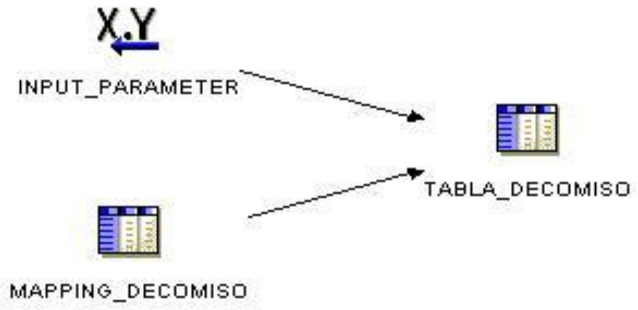

Fig. 28 MAPPING_DECOMISO_EADI

MAPPING: MAPPING_DETENCION_EADI

Tablas Fuentes:
- ✓ MAPPING_DETENCION

Tabla Destino:
- ✓ TABLA_ DETENCION

Imagen:

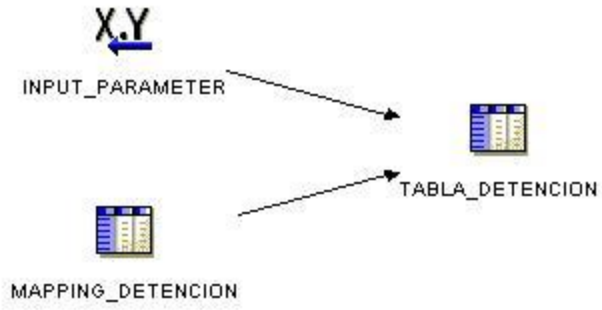

Fig. 29 MAPPING_DETENCION_EADI

MAPPING: MAPPING_EXPEDIENTE_EDAI

Tablas Fuentes:
- ✓ MAPPING_EXPEDIENTE

Tabla Destino:
- ✓ TABLA_ EXPEDIENTE

Imagen:

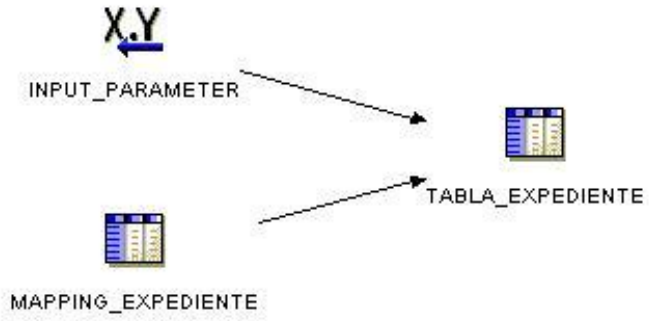

Fig. 30 MAPPING_EXPEDIENTE_EDAI

MAPPING: MAPPING_INCIDENTE_EADI

Tablas Fuentes:

- ✓ MAPPING_INCIDENTE

Tabla Destino:

- ✓ TABLA_ INCIDENTE

Imagen:

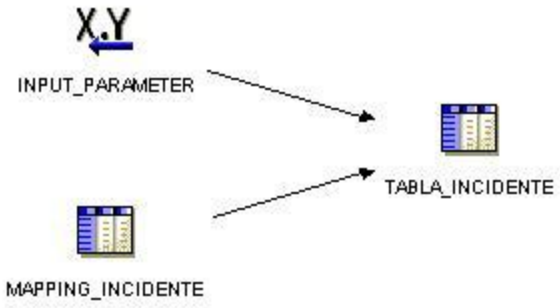

Fig. 31 MAPPING_INCIDENTE_EADI

MAPPING: MAPPING_INC_MONTO_EADI

Tablas Fuentes:

- ✓ MAPPING_INCIDENTE_MONTO

Tabla Destino:

- ✓ TABLA_INCIDENTE_MONTO

Imagen:

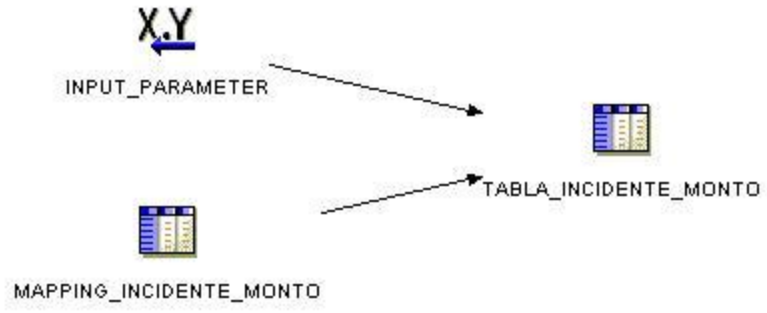

Fig. 32 MAPPING_INC_MONTO_EADI

MAPPING: MAPPING_INC_VICTIMA_EADI

Tablas Fuentes:

- ✓ MAPPING_INCIDENTE_VICTIMA

Tabla Destino:

- ✓ TABLA_ INCIDENTE_VICTIMA

Imagen:

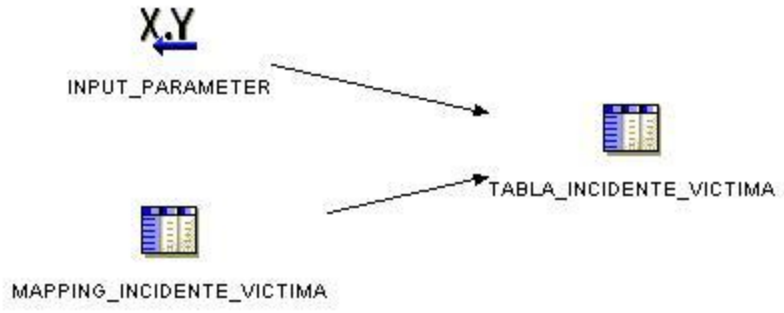

Fig. 33 MAPPING_INC_VICTIMA_EADI

MAPPING: MAPPING_RECUPERACION_EADI

Tablas Fuentes:

- ✓ MAPPING_RECUPERACION

Tabla Destino:

- ✓ TABLA_ RECUPERACIÓN

Imagen:

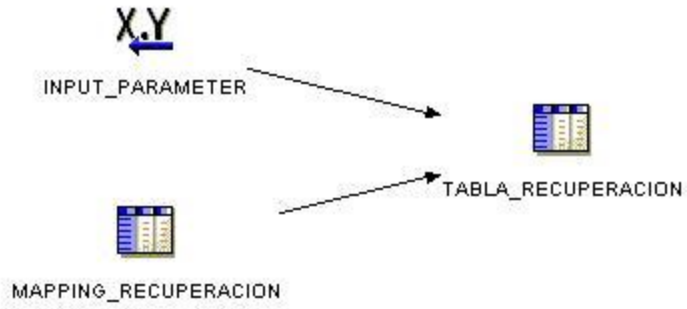

Fig. 34 MAPPING_RECUPERACION_EADI

MAPPING: MAPPING_RES_PERSONA_EADI

Tablas Fuentes:

- ✓ MAPPING_RESCATE_PERSONA

Tabla Destino:

- ✓ TABLA_ RESCATE_PERSONA

Imagen:

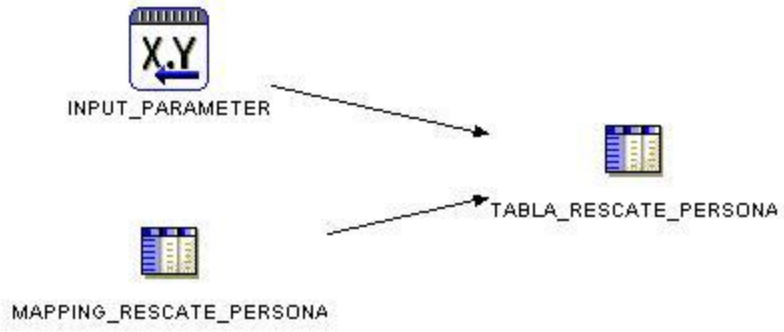

Fig. 35 MAPPING_RES_PERSONA_EADI

Procesos EADI -> EADD[24]

Fuente del Mapping

ESQUEMA: EADI (Estructuras de Almacén de Datos Integrado)

Destino de los Mapping:

ESQUEMA: EADD (Estructuras de Almacén de Datos Detalle)

MAPPING: TEM_MAPPING_DECOMISO

Tablas Fuentes:

- ✓ TABLA_DECOMISO

[24] Las Estructuras del Almacén de Datos Detallado constituye el Almacén de Datos Histórico

Tabla Destino:

- ✓ MAPPING_HECH_DECOMISO

Funciones involucradas:

- ✓ transf_temporal
- ✓ transf_sector
- ✓ transf_parte_dia
- ✓ transf_organo
- ✓ transf_concepto

Imagen:

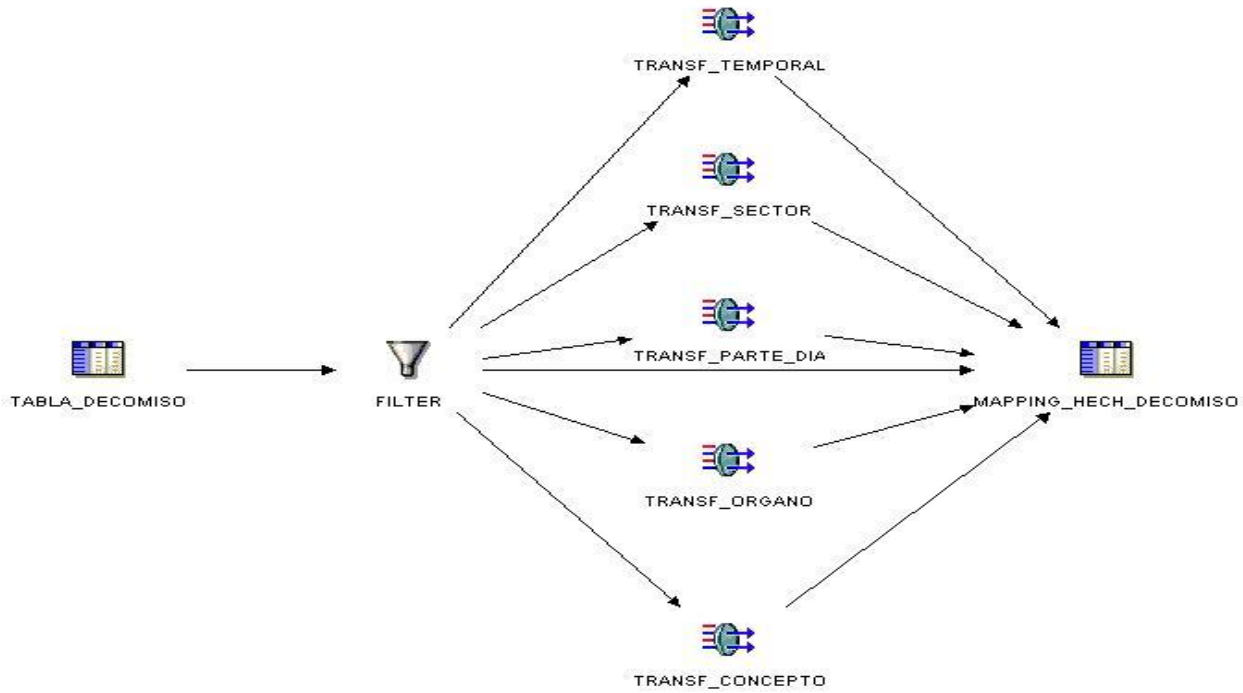

Fig. 36 TEM_MAPPING_DECOMISO

MAPPING: TEM_MAPPING_DETENCION

Tablas Fuentes:

- ✓ TABLA_DETENCION

Tabla Destino:

- ✓ MAPPING_ HECH_DETENCION

Funciones involucradas:

- ✓ transf_arma_tipo
- ✓ transf_causa_muerte
- ✓ transf_concepto
- ✓ transf_condicion
- ✓ transf_delito
- ✓ transf_destino
- ✓ transf_detencion
- ✓ transf_edad
- ✓ transf_estado_civil
- ✓ transf_estado_conciencia
- ✓ transf_estado_exp
- ✓ transf_grupo_etnico
- ✓ transf_lugar
- ✓ transf_nacionalidad
- ✓ transf_nivel_edu
- ✓ transf_ocupacion
- ✓ transf_organo
- ✓ transf_parte_dia
- ✓ transf_procedencia
- ✓ transf_relacion_vi
- ✓ transf_sector
- ✓ transf_sexo
- ✓ transf_temporal
- ✓ transf_vehiculo

Imagen:

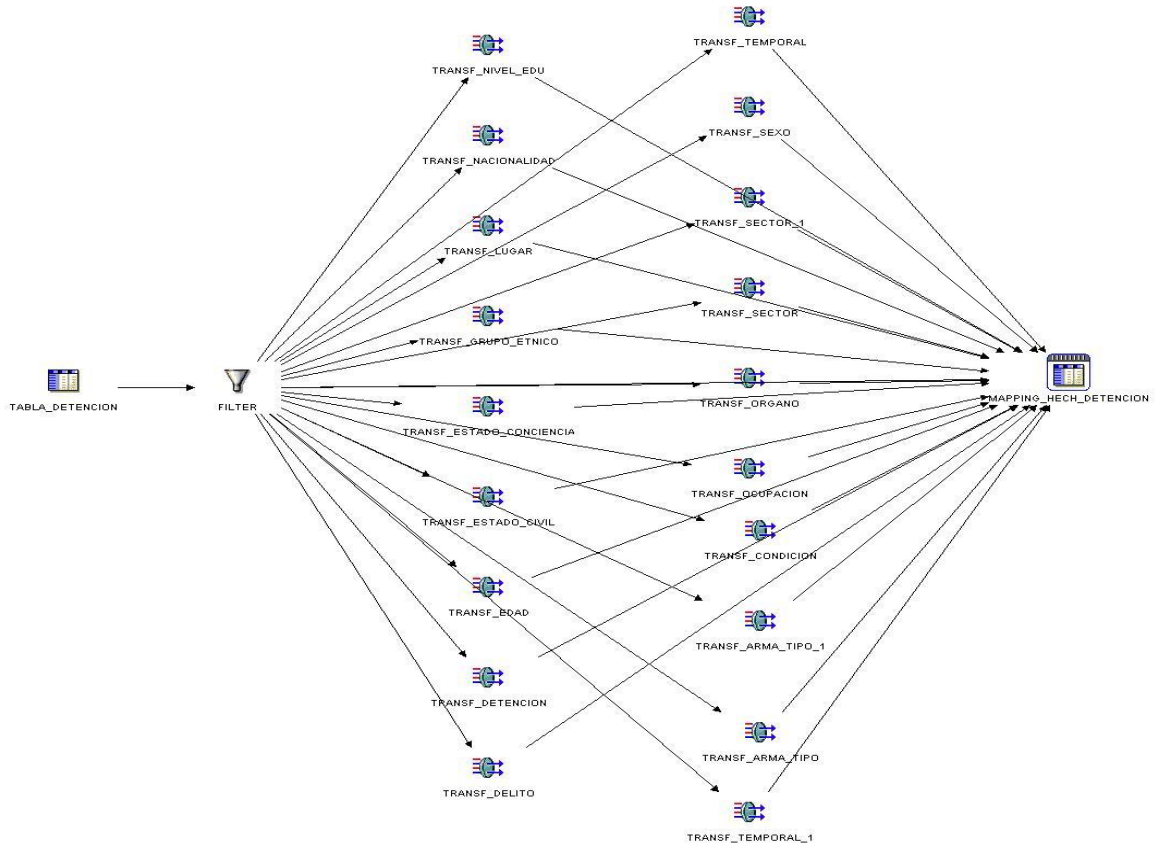

Fig. 37 TEM_MAPPING_DETENCION

MAPPING: TEM_MAPPING_INC_MONTO

Tablas Fuentes:

- ✓ TABLA_INCIDENTE_MONTO

Tabla Destino:

- ✓ MAPPING_HECH_INCIDENTE_MONTO

Funciones involucradas:

- ✓ transf_temporal
- ✓ transf_organo
- ✓ transf_lugar
- ✓ transf_arma_tipo
- ✓ transf_concepto
- ✓ transf_delito
- ✓ transf_sector
- ✓ transf_parte_dia
- ✓ transf_vehiculo

Imagen:

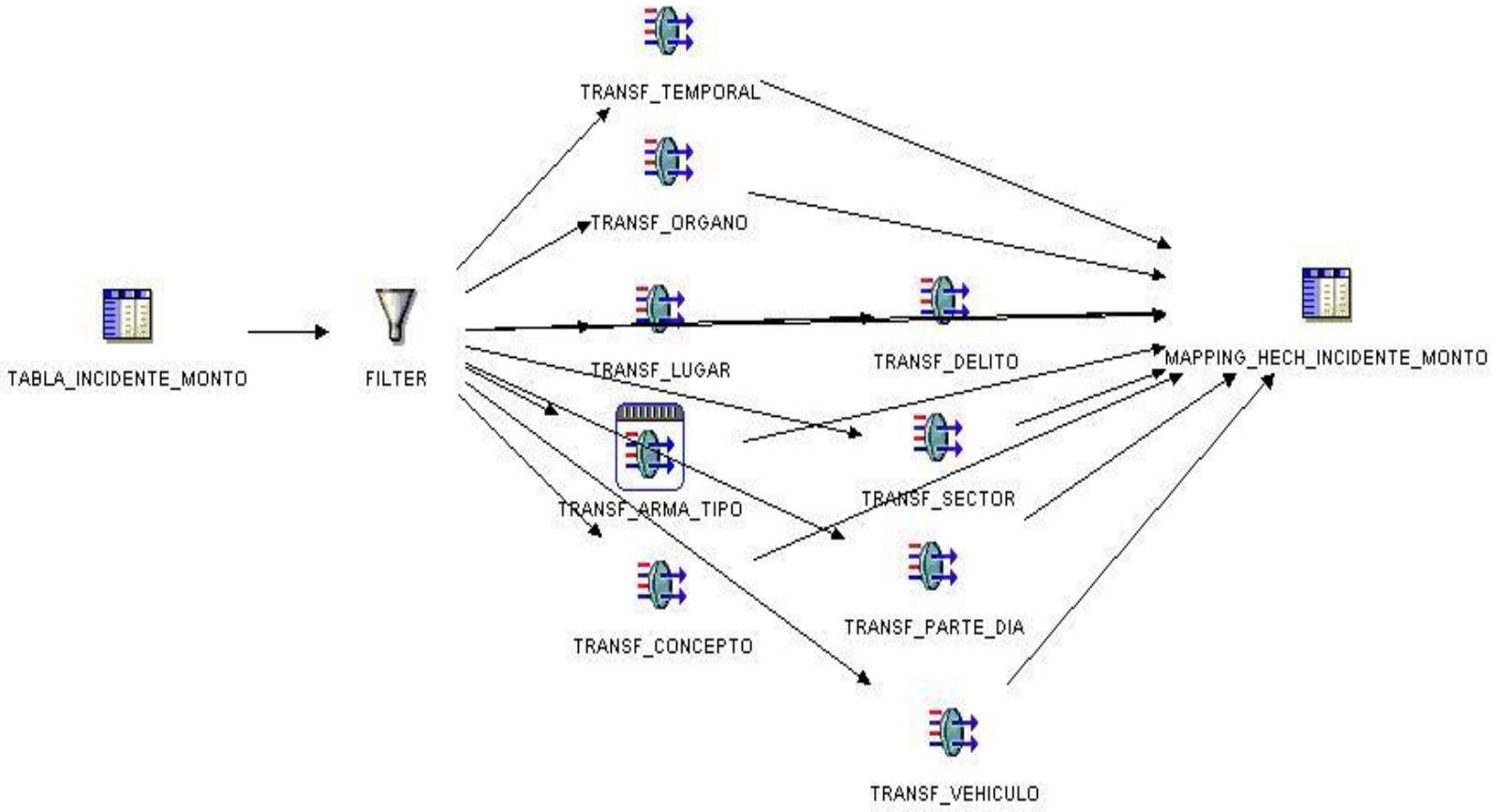

Fig. 38 TEM_MAPPING_INC_MONTO

MAPPING: TEM_MAPPING_EXPEDIENTE

Tablas Fuentes:

- ✓ TABLA_EXPEDIENTE

Tabla Destino:

- ✓ MAPPING_HECH_EXPEDIENTE

Funciones involucradas:

- ✓ fecha_hora
- ✓ transf_delito
- ✓ transf_estado_conciencia
- ✓ transf_organo
- ✓ transf_temporal.

Imagen:

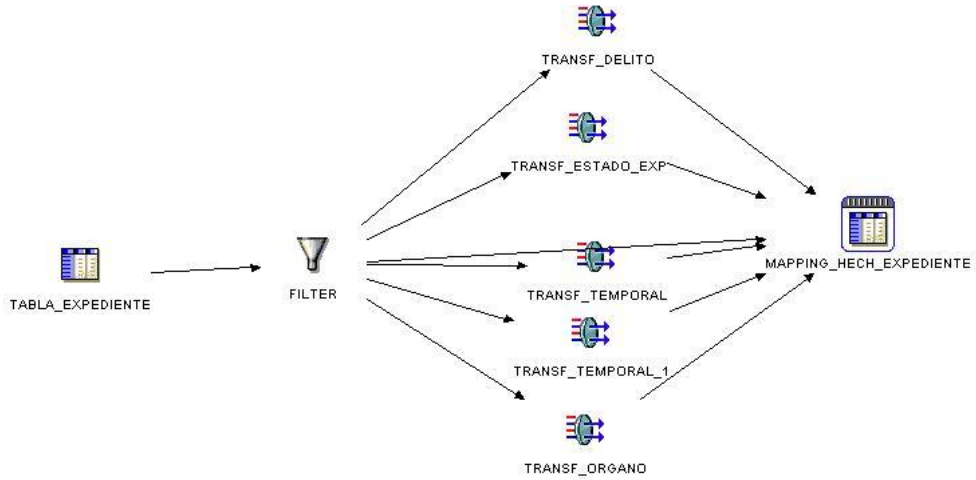

Fig. 39 TEM_MAPPING_EXPEDIENTE

MAPPING: TEM_MAPPING_INCIDENTE

Tablas Fuentes:

- ✓ TABLA_INCIDENTE

Tabla Destino:

- ✓ MAPPING_HECH_INCIDENTE

Funciones involucradas:

- ✓ transf_arma_tipo
- ✓ transf_delito
- ✓ transf_organo
- ✓ transf_parte_dia
- ✓ transf_sector
- ✓ transf_temporal

Imagen:

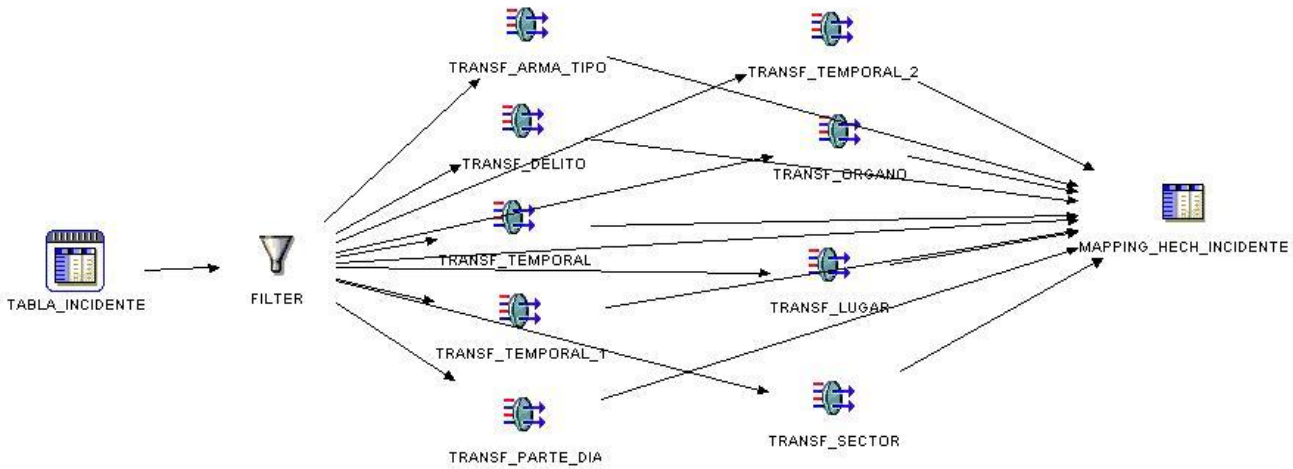

Fig. 40 TEM_MAPPING_INCIDENTE

MAPPING: TEM_MAPPING_RECUPERACION

Tablas Fuentes:
- ✓ TABLA_RECUPERACION

Tabla Destino:
- ✓ MAPPING_HECH_RECUPERACION

Funciones involucradas:
- ✓ fecha_hora
- ✓ transf_concepto
- ✓ transf_delito
- ✓ transf_lugar
- ✓ transf_organo
- ✓ transf_parte_dia
- ✓ transf_sector
- ✓ transf_temporal
- ✓ transf_vehiculo

Imagen:

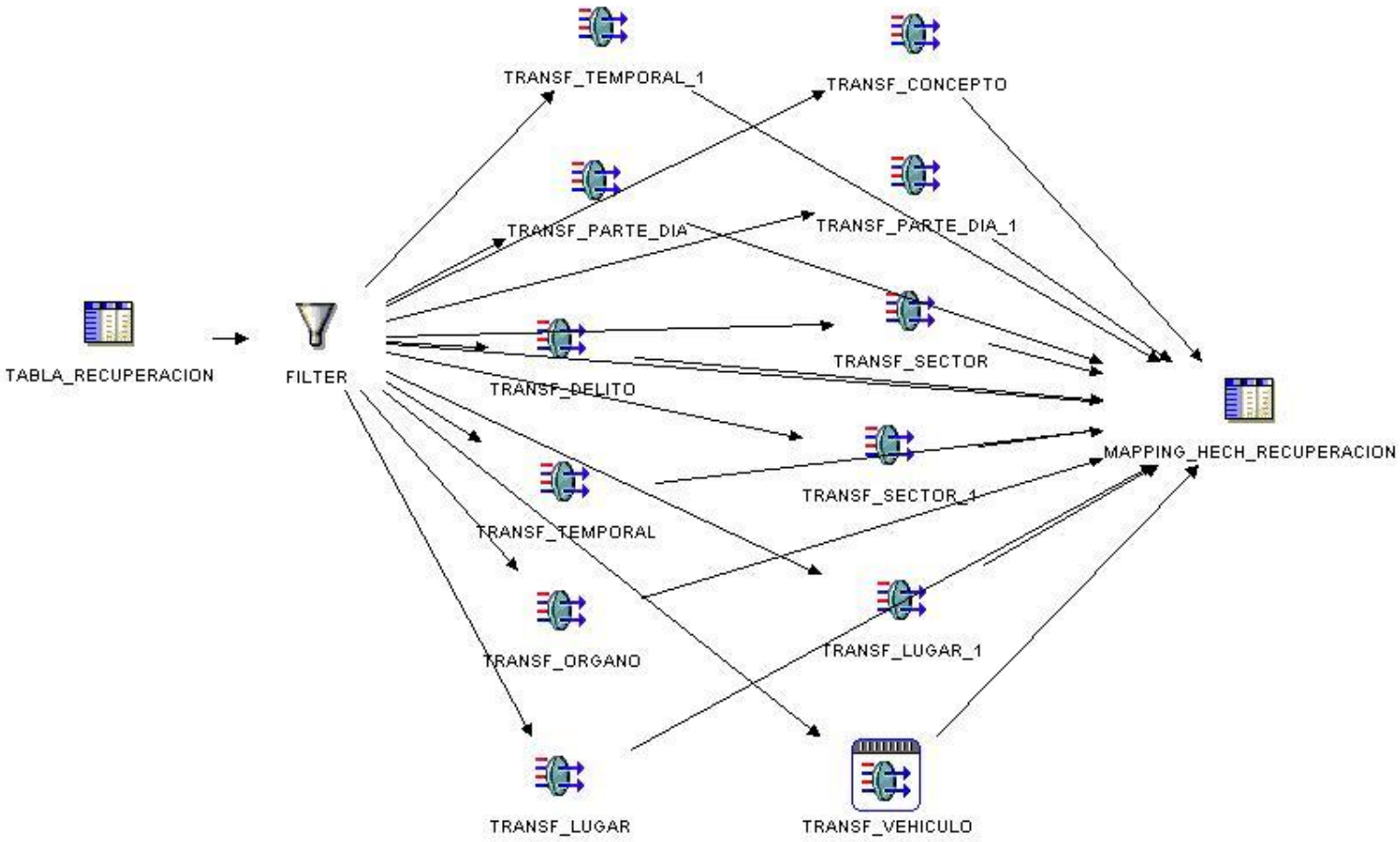

Fig. 41 TEM_MAPPING_RECUPERACION

MAPPING: TEM_MAPPING_RESCATE_PERSONA

Tablas Fuentes:

- ✓ TABLA_RESCATE_PERSONA

Tabla Destino:

- ✓ MAPPING_HECH_RESCATE_PERSONA

Funciones involucradas:

- ✓ transf_edad
- ✓ transf_organo
- ✓ transf_sector
- ✓ transf_sexo
- ✓ transf_temporal

Imagen:

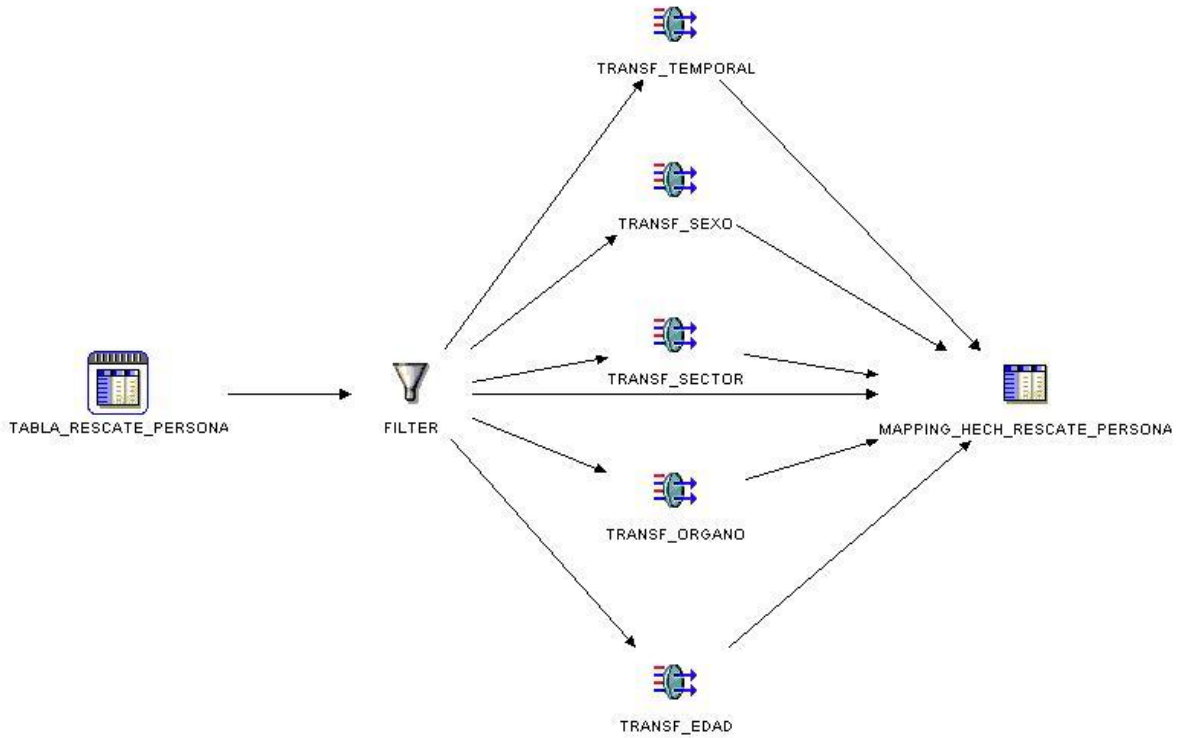

Fig. 42 TEM_MAPPING_RESCATE_PORSONA

MAPPING: MAPPING_DETENCION_EADD

Tablas Fuentes:
- ✓ MAPPING_ HECH_DETENCION

Tabla Destino:
- ✓ HECH_DETENCION

Imagen:

Fig. 43 MAPPING_DETENCION_EADD

MAPPING: MMAPPING_INC_MONTO_EADD

Tablas Fuentes:
- ✓ MAPPING_HECH_INCIDENTE_MONTO

Tabla Destino:
- ✓ HECH_INCIDENTE_MONTO

Imagen:

Fig. 44 MMAPPING_INC_MONTO_EADD

MAPPING: MAPPING_EXPEDIENTE_EADD

Tablas Fuentes:
- ✓ MAPPING_HECH_EXPEDIENTE

Tabla Destino:
- ✓ HECH_EXPEDIENTE

Imagen:

Fig. 45 MAPPING_EXPEDIENTE_EADD

MAPPING: MAPPING_INCIDENTE_EADD

Tablas Fuentes:
- ✓ MAPPING_HECH_INCIDENTE

Tabla Destino:
- ✓ HECH_INCIDENTE

Imagen:

Fig. 46 MAPPING_INCIDENTE_EADD

MAPPING: MAPPING_INC_VICTIMA_EADD

Tablas Fuentes:

- ✓ MAPPING_HECH_INCIDENTE_VICTIMA

Tabla Destino:

- ✓ HECH_INCIDENTE_VICTIMA

Imagen:

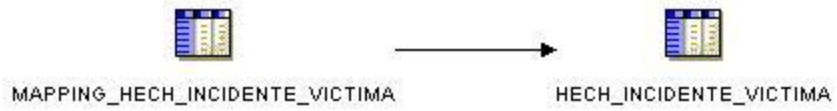

Fig. 47 MAPPING_INC_VICTIMA_EADD

MAPPING: MAPPING_INV_FORENSE_EADD

Tablas Fuentes:
- ✓ MAPPING_HECH_INV_FORENSE

Tabla Destino:
- ✓ HECH_INVESTIGACION_FORENSE

Imagen:

Fig. 48 MAPPING_INV_FORENSE_EADD

MAPPING: MAPPING_RESCATE_PERSONA_EADD

Tablas Fuentes:
- ✓ MAPPING_HECH_RESCATE_PERSONA

Tabla Destino:
- ✓ HECH_RESCATE_PERSONA

Imagen:

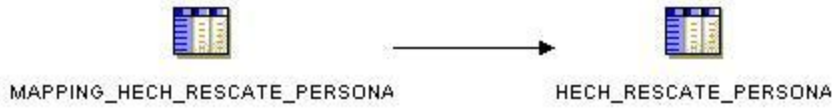

Fig. 49 MAPPING_RESCATE_PERSONA_EADD

Anexo 3 Descripción de las estructuras que conforma el área de almacenamiento intermedio

Nombre	Tipo	Llaves Primarias	# Atrib.
ETL_AUDIT_TRANSLATION	Independiente		6
ETL_DATA_BEHAVIOR	Independiente	BEHAVIOR_ID	2
ETL_EXISTENCE_DICTIONARY	Independiente	VARIABLE_ID,SOURCE_ID, TARGET_VALUE	4
ETL_FTP_CONNECTION	Independiente	FTP_SERVICE_ID	7
ETL_INTEGRATION_BEHAVIOR	Independiente	BEHAVIOR_ID	2
ETL_INTEGRATION_CONCEPT	Independiente	CONCEPT_ID	4
ETL_INTEGRATION_PROTOCOL	Independiente	PROTOCOL_ID	2
ETL_INTEGRATION_SOURCE	Independiente	SOURCE_ID	2
ETL_LOG	Independiente		8
ETL_LOG_TYPE	Independiente	ID	2
ETL_MASIVE_LOAD_AUDIT	Independiente		3
ETL_PACKAGE	Dependiente	ID_PAQUET, CONCEPT_ID, SOURCE_ID	8
ETL_SMTP_CONNECTION	Independiente	SMTP_SERVICE_ID	1
ETL_SOURCE_CONTENT	Dependiente	VARIABLE_ID,SOURCE_ID	6
ETL_TRANSLATION_DICTIONARY	Dependiente	VARIABLE_ID,SOURCE_ID,SOURCE_VALUE	5
ETL_VARIABLE	Independiente	VARIABLE_ID	3

ETL_WS_CONNECTION	Independiente	WS_SERVICE_ID	1
ETL_XML_CONFIG	Dependiente	CONCEPT_ID,SOURCE_ID	7
ETL_XML_HISTORY	Dependiente	HISTORY_TIMESTAMP, CONCEPT_ID,SOURCE_ID	4
ETL_XML_LAST_LOAD	Dependiente	CONCEPT_ID,SOURCE_ID	7
ETL_XML_STATUS	Independiente	ID_STATUS	2
ETL_XML_STRUCTURE	Dependiente	CONCEPT_ID,SOURCE_ID, PATH	4
ETL_XML_STRUCTURE_NODE	Independiente	ID_STRUCTURE_NODE	6

Tabla 14 Listado de las tablas que conforma la parte mas importante del proceso de ETL (Área de Almacenamiento Intermedio)

1. Tabla de Auditoría de las Transformaciones

Nombre	ETL_AUDIT_TRANSLATION
Llaves Primarias	
Definición	Mantiene un registro de todas las transformaciones que no se aplicaron con el diccionario de correspondencia. Se tiene un registro de los valores que no existen o que vienen nulos en el proceso de integración.

Atributos

Atributo	**Tipo de Dato**	**Null**	**Definición**
SOURCE_ID	VARCHAR(10)	Y	Identificador de la Fuente de Datos
VARIABLE_ID	VARCHAR(25)	Y	El identificador de cada variable que aplica en la transformación.
INITIAL_VALUE	VARCHAR(100)	Y	El valor inicial enviado por la fuente, el cual no se encuentra definido el diccionario de correspondencia.
CONVERTED_VALUE	VARCHAR(25)	Y	Es la traducción que se le aplicó al valor nulo o desconocido.
DATE_TRANSFORMATION	DATE	Y	Fecha en que se realizó la transformación.
DATA_BEHAVIOR	VARCHAR(25)	Y	El tipo de transformación que se le aplicó al valor inicial que puede ser transformación por no existencia o por valor nulo.

2. Tabla de Comportamiento de las Transformaciones

Nombre	ETL_DATA_BEHAVIOR
Llaves Primarias	BEHAVIOR_ID
Definición	Almacena los tipos de comportamientos o transformaciones (Inferencia, Traducción, Rechazo) que se le aplican a los datos. Solamente se implementa el comportamiento de traducción.

Atributos

Atributo	**Tipo de Dato**	**Null**	**Definición**
BEHAVIOR_ID	VARCHAR(25)	N	Identificador de la transformación.
BEHAVIOR_NAME	VARCHAR(25)	Y	Nombre de la transformación.

3. Tabla de Registro todo el Proceso de Integración

Nombre	ETL_LOG
Llaves Primarias	
Definición	Almacena las trazas del proceso de integración relativo a errores, alertas y casos de éxito.

Atributos

Atributo	Tipo de Dato	Null	Definición
CODIGO	NUMERIC(0, 0)	Y	Código del mensaje.
MENSAJE	NVARCHAR(2000)	Y	Texto del mensaje.
USUARIO	NVARCHAR(50)	Y	Usuario que realiza la operación.
INF_DETALLADA	NVARCHAR(2000)	Y	Información detallada del comportamiento
ETAPA	NVARCHAR(1000)	Y	Etapa en que se generó el mensaje, hace referencia a la etapa del proceso de ETL.
FECHA_HORA	DATE	Y	Fecha y hora del proceso.
LOG_TYPE	VARCHAR(15)	Y	Tipo de Mensaje (Alerta, Error, Éxito).
FUENTE	VARCHAR(10)	Y	Fuente involucrada en el proceso.

4. Tabla de Registro de Comportamientos por Fuente de Información

Nombre	ETL_SOURCE_CONTENT
Llaves Primarias	VARIABLE_ID,SOURCE_ID
Definición	Contiene las transformaciones que se le aplican a los valores que no existan o que sean nulos (NULL_BEHAVIOR, NULL_TRANSLATION, NOMATCHING_BEHAVIOR, NOMATCHING_TRANSLATION).

Atributos

Atributo	Tipo de Dato	Null	Definición
VARIABLE_ID	VARCHAR(25)	N	Identificador de la Variable.
SOURCE_ID	VARCHAR(10)	N	Identificador de la Fuente.
CONTENT_NULL_BEHAVIOR_ID	VARCHAR(15)	N	Identificador del comportamiento del contenido por valor nulo.
CONTENT_NOMATCHING_BEH_ID	VARCHAR(15)	N	Identificador del comportamiento del contenido por valor de no existencia.
CONTENT_NULL_TRANSLATION	VARCHAR(50)	Y	Traducción del contenido por valor nulo.
CONT_NOMATCHING_TRANSLATION	VARCHAR(50)	Y	Traducción del contenido por no existencia.

5. Tabla dell Diccionario de Correspondencia

Nombre	ETL_TRANSLATION_DICTIONARY
Llaves Primarias	VARIABLE_ID,SOURCE_ID,SOURCE_VALUE
Definición	Contiene todas las transformaciones que se le aplican a los datos de las diferentes fuentes. Este es el llamado Diccionario de Correspondencia.

Atributos

Atributo	Tipo de Dato	Null	Definición
VARIABLE_ID	VARCHAR(25)	N	Identificador de la Variable.
SOURCE_ID	VARCHAR(10)	N	Identificador de la Fuente.
SOURCE_VALUE	VARCHAR(200)	N	Valor de entrada desde la fuente.
TARGET_VALUE	VARCHAR(200)	Y	Valor de salida hacia el Almacén de Datos. (Dialecto SINSEC)
DESCRIPCION	VARCHAR(2000)	Y	Breve descripción de la transformación.

6. Tabla de Variables

Nombre	ETL_VARIABLE
Llaves Primarias	VARIABLE_ID
Definición	Define todas las variables que existen asociadas a un concepto (Ejem. Sexo, Edad, Delito, Causa de Muerte).

Atributos

Atributo	Tipo de Dato	Null	Definición
VARIABLE_ID	VARCHAR(25)	N	Identificador de la variable.
VAR_NAME	VARCHAR(40)	N	Nombre de la variable.

TABLE_NAME	VARCHAR(40)	N	Tabla asociada a la variable en el esquema de producción.

7. Tabla de Configuración de los XML

Nombre	ETL_XML_CONFIG
Llaves Primarias	CONCEPT_ID,SOURCE_ID
Definición	Almacena la configuración cada Archivo de Intercambio de Datos.

Atributos

Atributo	Tipo de Dato	Null	Definición
CONCEPT_ID	VARCHAR(15)	N	Identificador del Concepto.
SOURCE_ID	VARCHAR(10)	N	Identificador de la Fuente.
XML_TYPE_FILENAME	VARCHAR(50)	N	Nombre del Archivo de Intercambio de Datos XML.
XML_TYPE_ALIAS	VARCHAR(50)	N	Alias asociado.
XML_TYPE_TEM_DIR	VARCHAR(200)	N	Directorio temporal de almacenamiento del Archivo de Intercambio de Datos XML.
XML_TYPE_DIR	VARCHAR(200)	N	Dirección de almacenamiento del Archivo de Intercambio de Datos XML.
XML_TYPE_ACTIVE	VARCHAR(10)	N	Este campo estará configurado para Activar/Desactivar algún concepto a cargar

			por una fuente determinada. EL proceso de integración utiliza este campo para cargar los contenidos. Si este campo no está activo para un archivo, el mismo no será integrado.

8. Tabla del Registro Histórico de Carga

Nombre	ETL_XML_HISTORY
Llaves Primarias	HISTORY_TIMESTAMP,CONCEPT_ID,SOURCE_ID
Definición	Contiene un registro histórico de las cargas de datos realizadas.

Atributos

Atributo	**Tipo de Dato**	**Null**	**Definición**
HISTORY_TIMESTAMP	DATE	N	Fecha.
CONCEPT_ID	VARCHAR(15)	N	Identificador del Concepto.
SOURCE_ID	VARCHAR(10)	N	Identificador de la Fuente.
HISTORY_XML	CHAR(10)	N	Contenido del Archivo de Intercambio de Datos XML.

9. Tabla de la Estructura del XML

Nombre	ETL_XML_STRUCTURE
Llaves Primarias	CONCEPT_ID,SOURCE_ID,PATH
Definición	Define la estructura de los Archivos de Integración de Datos XML.

Atributos

Atributo	**Tipo de Dato**	**Null**	**Definición**
CONCEPT_ID	VARCHAR(15)	N	Identificador del Concepto.
SOURCE_ID	VARCHAR(10)	N	Identificador de la Fuente.
PATH	VARCHAR(100)	N	Dirección dentro del XML de cada nodo principal.
TARGET_TABLE	VARCHAR(50)	N	Tabla destino.

10. Tabla de la Estructura de los Nodos del XML

Nombre	ETL_XML_STRUCTURE_NODE
Llaves Primarias	ID_STRUCTURE_NODE
Definición	Define la estructura de cada nodo dentro del Archivo de Integración de Datos XML.

Atributos

Atributo	Tipo de Dato	Null	Definición
ID_STRUCTURE_NODE	NUMERIC(38, 0)	N	Identificador del la estructura.
CONCEPT_ID	VARCHAR(15)	N	Identificador del Concepto.
SOURCE_FIELD	VARCHAR(100)	N	Dirección del nodo principal asociado.
SOURCE_ID	VARCHAR(10)	N	Identificador de la Fuente.
TARGET_FILE_NODE	VARCHAR(100)	N	Campo asociado a cada nodo.
SOURCE_FILE_NODE	VARCHAR(100)	N	Dirección de cada nodo.

Anexo 4 Especificaciones del XML

Nivel	Campo (Nodo)	Definición	Valor de Ejemplo
1	xml	Nodo de inicio del XML	
1	dataroot	root	
2	**incidencia**	**Nodo Incidencia**	
3	incidencia_id	Identificador de negocio de una incidencia.	19
3	delegacion_id	Identificador de la Delegación a la que pertenece la Subdelegación que brinda la información.	1
3	delegacion	Nombre de la delegación.	Delegación Estadal Barinas
3	subdelegacion_id	Identificador de la Subdelegación o despacho que brinda la información	1
3	subdelegacion	Nombre de la Subdelegación	Sub Delegación: Barinas
3	fecha_estadistica	Fecha en la que se reporta el incidente (Adición del Registro en el sistema).	2008-03-12T00:00:00
3	estado_id	Identificador del estado donde ocurre el incidente.	01
3	estado	Nombre del estado.	Distrito Capital
3	municipio_id	Identificador del Municipio donde ocurre el incidente.	00
3	municipio	Nombre del Municipio	No Declarado
3	parroquia_id	Identificador de la parroquia donde ocurre	00

		el incidente.	
3	parroquia	Nombre de la parroquia	No Declarado
3	sector_id	Identificador del Sector donde ocurre el incidente.	000
3	sector	Nombre del Sector	No Declarado
3	fecha_ocurrencia	Fecha de Ocurrencia del Incidente.	2008-03-12T00:00:00
3	hora_ocurrencia_id	Identifica la Hora de Ocurrencia agrupada en grupos según el valor de este campo, Ejemplo: Ocurrió a las 15 horas, en este caso sería: 15:00-15:59 (1559)	1559
3	hora_ocurrencia	Valor del grupo de Horas	00:00-00:59
3	expediente	Número de Expediente.	H090876
3	monto	Monto indirecto	Bs 130,000.00
3	delito_id	Identificador del Delito	7
3	delito	Nombre del Delito	Homicidio
3	motivo_id	Identificador del motivo por el que ocurre el incidente.	14
3	Motivo	Motivo del delito	Homicidio Pasional
3	arma_id	Identificador del tipo de arma utilizada en el hecho.	4
3	arma	Nombre del Arma.	Arma de Fuego
3	relevancia_id	Identificador de la Relevancia del hecho	1

		(Relevante, Normal)	
3	relevancia	Nombre de la relevancia.	Normal
3	relacion_agraviado_imputado_id	Identificador de la relación entre el agraviado y la victima.	5
3	relacion_agraviado_imputado	Nombre de la Relación entre el agraviado y el imputado	Casual
3	lugar_id	Identificador del tipo de lugar donde ocurre el incidente.	1
3	Lugar	Nombre del tipo de lugar donde ocurre el incidente.	Vía Pública
3	**agraviado**	**Agraviado o Victima del incidente.**	
4	agraviado_id	Identificador de negocio.	1
4	incidencia_id	Identificador de la incidencia para relacionar la victima con el hecho.	20
4	expediente	Número del expediente delictivo.	H742361
4	delegacion_id	Identificador de la Delegación que reporta el agraviado.	13
4	Delegacion	Nombre de la Delegación que reporta al agraviado.	Dependencias del Dtto Metropolitano
4	subdelegacion_id	Identificador de la Subdelegación que reporta el agraviado.	39
4	subdelegacion	Nombre de la Subdelegación.	DIV. NAC. C/ LA DELINCUENCIA RGANIZADA

4	fecha_estadistica	Fecha en la que se registra el agraviado en el sistema. Debe ser la misma fecha en la que registraron el agraviado.	2008-03-03T00:00:00
4	Persona_juridica	Indica si el agraviado es natural o jurídico.	1 (Si)
4	agrv_indocumentado	Identifica si el agraviado es indocumentado.	0 (No)
4	sexo_id	Identificador del Sexo del agraviado.	0
4	Sexo	Nombre del sexo.	No declarado
4	condicion_id	Identificador de las condiciones del agraviado.	3
4	condicion	Condición en la que resulto el agraviado	Lesionado
3	**imputado**	**Imputado que efectuó el incidente.**	
4	imputado_id	Identificador de Negocio que identifica al imputado.	13
4	incidencia_id	Identificador de la Incidencia.	23
4	expediente	Número del Expediente Delictivo abierto a raíz del incidente.	H859172
4	delegacion_id	Identificador de la Delegación que reporta al imputado.	13
4	delegacion	Nombre de la delegación.	Dependencias del Dtto Metropolitano
4	subdelegacion_id	Identificador de la subdelegación.	56
4	subdelegacion	Nombre de la subdelegación	Subdelegación: Oeste

4	fecha_estadistica	Fecha en que se reporta el imputado.	2008-02-26T00:00:00
4	Edad	Edad del imputado.	0
4	sexo_id	Identificador del sexo del imputado.	0
4	sexo	Nombre del sexo.	No declarado
4	detencion_id	Identificador del Tipo de Detención Efectuada (No declarada, Por Flagrancia, por Investigación)	0
4	detencion	Nombre del tipo de detención.	No declarado
3	**arma_denunciada**	**Arma denunciada en el incidente (si aplica)**	
4	arma_denunciada_id	Identificador de Negocio que indica el registro de un arma denunciada.	1
4	incidente_id	Identificador de Incidente relacionado con el arma denunciada.	19
4	expediente	Número de Expediente asociado al hecho delictivo.	H090876
4	delegacion_id	Identificador de la Delegación a la que pertenece la Subdelegación que brinda la información.	1
4	delegacion	Nombre de la delegación.	Delegación Estadal Barinas
4	subdelegacion_id	Identificador de la Subdelegación que brinda la información.	1
4	subdelegacion	Nombre de la subdelegación.	Sub Delegación: Barinas
4	fecha_estadistica	Fecha en la que se reporta el arma	2008-03-12T00:00:00

		denunciada.	
4	arma_id	Identificador del Tipo de Arma denunciada.	3
4	arma	Nombre del tipo de arma denunciada.	Arma Contundente
4	cantidad	Cantidad de armas del tipo especificado.	1
4	valor	Valor calculado que corresponde al tipo y la cantidad de armas denunciadas.	98980
3	**vehiculo_denunciado**	**Vehículos denunciados en el incidente.**	
4	vehiculo_denunciado_id	Identificador de Negocio del Vehiculo denunciado.	11
4	incidencia	Identificador de la Incidencia que denuncia el vehículo.	30
4	expediente	Número de Expediente del Hecho Delictivo.	H733405
4	delegacion_id	Identificador de la Delegación a la que pertenece la Subdelegación que brinda la información.	13
4	delegacion	Nombre de la delegación.	Dependencias del Dtto Metropolitano
4	subdelegacion_id	Identificador de la Subdelegación que brinda la información.	55
4	Subdelegación	Nombre de la subdelegación	Subdelegación: Chacao
4	Fecha_estadística	Fecha en la que se reporta el vehiculo.	2008-02-26T00:00:00
4	vehiculo_id	Identificador del tipo de Vehiculo o medio	2

		de transporte.	
4	vehiculo	Nombre del tipo de vehículo.	Automóvil
4	marca_id	Identificador de la marca del vehículo.	32
4	marca	Nombre de la marca del vehículo.	Renault
4	modelo_id	Identificador del modelo del vehículo.	231
4	modelo	Nombre del modelo del vehículo.	Clio(1998-2000, 2004-2006)
4	anno	Año de fabricación del vehículo.	2000
4	valor	Valor del vehículo.	0
3	**monto_denunciado**	**Monto denunciado por conceptos.**	
4	monto_denunciado_id	Identificador de Negocio que representa la denuncia de ese monto especifico.	5
4	incidente	Identificador del Incidente delictivo.	23
4	expediente	Número de Expediente.	H271566
4	delegacion_id	Identificador de la Delegación a la que pertenece la Subdelegación que brinda la información.	13
4	delegacion	Nombre de la delegación.	Dependencias del Dtto Metropolitano
4	subdelegacion_id	Identificador de la subdelegación que brinda la información.	59
4	subdelegacion	Nombre de la subdelegación	Subdelegación: El Valle
4	fecha_estadistica	Fecha en la que se denuncia el monto.	2008-02-26T00:00:00

4	monto_id	Identificador al tipo de elemento denunciado (Joyas, Dinero, etc.)	1
4	Monto	Nombre del monto, o concepto.	Joyas
4	Valor	Valor monetario que representa la denuncia del concepto. (Monto numérico)	10000