

Universidad de las Ciencias Informáticas

Facultad 15



Título: Análisis y Diseño del Data Warehouse para el Sistema informativo de la actividad de Estructura y Composición de las FAR.

Trabajo de diploma para optar por el título de
Ingeniero en Ciencias Informáticas

Autor: Cadete Yudisleidy Arias Pérez

Tutor: Ing. 1Tte José Gabriel Fernández Pérez

Ciudad de La Habana, 2010.

Año 52 de la Revolución

DECLARACIÓN DE AUTORÍA

Declaro ser autora de la presente tesis y reconozco a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo.

Para que así conste firmo la presente a los ____ días del mes de _____ del año _____.

Autora

Yudisleidy Arias Pérez

Tutor

1Tte José Gabriel Fernández Pérez

DATOS DE CONTACTO

Tutor:

1Tte. José Gabriel Fernández Pérez.

Ingeniero Informático.

Centro laboral: Centro Principal de Automatización de las FAR (CPA).

e-mail: cpa@unicom.co.cu

Categoría Científica: Aspirante a Investigador.

RESUMEN

Entre las principales áreas de nuestra Universidad se encuentra la Unidad de Compatibilización, Integración y Desarrollo de Software para la Defensa (UCID), la cual cuenta con varios proyectos tales como el Sistema Informativo para el aseguramiento de la actividad de estructura y composición de las FAR (SIDECE), a través del cual se garantizan los procesos de perfeccionamiento y organización de las estructuras de las FAR.

Presente la solicitud por parte del Departamento de Organización y Personal de la Dirección de Organización y Personal del Ministerio de las Fuerzas Armadas Revolucionarias (Minfar), de garantizar el almacenamiento y recuperación de la información histórica generada por el SIDECE, para su posterior análisis; se propone, diseñar un Almacén de Datos que tras concebirse su implementación se podría satisfacer esta necesidad. Es precisamente mediante novedosas técnicas de almacenamiento de datos que se podrán auxiliar los especialistas de este órgano para garantizar sus misiones con la mayor eficiencia y prontitud posibles.

En el presente trabajo se realiza un análisis y la propuesta de diseño de un Almacén de Datos para el SIDECE.

TABLA DE CONTENIDOS

RESUMEN.....	II
CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA.....	4
1.1 Introducción.....	4
1.2 Sistemas Operacionales o de Producción	4
1.3 Sistemas Informacionales	4
1.4 Procesamiento Transaccional en Línea (OLTP)	4
1.5 Procesamiento analítico en Línea (OLAP).....	5
1.6 Almacenes de Datos	5
1.6.1 Definición de almacén de datos	6
1.6.2 Objetivos generales de los Data Warehouse.....	6
1.6.3 Características de los almacenes de datos	7
1.6.4 Estructura básica de la arquitectura DW	7
1.6.5 Ventajas.....	8
1.7 Aplicación de los almacenes de datos en el mundo.....	9
1.7.1 Soluciones cubanas.....	10
1.8 Herramientas.....	11
1.8.1 Plataforma BI	11
1.8.2 ETL.....	11
1.8.3 PostgreSQL 8.3 (Gestor de Base de Datos)	11
1.8.4 Apache (Servidor de Aplicaciones Web)	12
1.8.5 Mozilla (Navegador Web).....	12
1.8.6 Herramienta CASE. Visual Paradigm	13
1.9 Metodología	13
1.9.1 HEFESTO.....	13
1.10 Conclusiones del capítulo	14
CAPÍTULO 2 CARACTERÍSTICAS DEL SIDEC	15
2.1 Introducción.....	15
2.2 Análisis de los requerimientos.....	15
2.2.1 Identificar preguntas.....	15
2.2.2 Identificar indicadores y perspectivas de análisis	17
2.2.3 Modelo Conceptual	21
2.3 Dimensión Tiempo	23
2.4 Conclusiones del capítulo	23

CAPÍTULO 3 ANÁLISIS Y DISEÑO DEL ALMACÉN DE DATOS	24
3.1 Introducción.....	24
3.2 Establecer correspondencias con los requerimientos	24
3.3 Selección de los campos que integrarán cada perspectiva. Nivel de granularidad....	54
3.4 Elaboración del modelo lógico de la estructura del almacén de datos.....	60
3.4.1 Esquema en Constelación	61
3.4.2 Diseñar tablas de dimensiones.....	61
3.5 Diseñar tablas de hechos	65
3.5.1 Casos para tener en cuenta en la creación de la Tabla de hechos.....	66
3.6 Flujo de datos	69
3.7 Conclusiones del capítulo	69
CONCLUSIONES.....	70
RECOMENDACIONES.....	71
BIBLIOGRAFÍA.....	72
REFERENCIAS BIBLIOGRÁFICAS.....	73
GLOSARIO DE TÉRMINOS	74

INTRODUCCIÓN

Con la informatización de la sociedad, ha crecido a nivel mundial la capacidad de generación y almacenamiento de la información, que no puede ser analizada por los métodos tradicionales existentes, mientras mayor es la capacidad para almacenar datos, mayor es la incapacidad para extraer información realmente útil en las entidades. Mucha información importante, queda sepultada y disgregada, y los sistemas existentes no están preparados para el nuevo reto.

Mucho se ha hablado de la Era de la Información y sus ventajas; con las nuevas posibilidades se acortan las distancias y crecen los beneficios para quienes tienen acceso al gran caudal de datos. Sin embargo, lo que constituye un valioso recurso para todos, se ha tornado en el gran problema de principios de siglo, manejar de forma óptima grandes volúmenes de información.

Las entidades necesitan depositar toda su confianza en la Toma de Decisiones, decisiones que tienen que ser rápidas y deben de ser tomadas sobre una cantidad de hechos y cifras, constituyendo las unidades básicas de la gestión. Las buenas decisiones son las bases para conseguir un rendimiento excepcional. Ahora las entidades no dependen tan solo de factores como ubicación, productos, etc., sino que también del conocimiento. Tal conocimiento, basado en información comprensible, detallada y relevante es crucial para lograr y sostener ventaja en eficiencia y factibilidad. El poseer conocimientos correctos significa tener respuestas correctas y realizar decisiones estratégicas para la ejecución de la entidad. La tarea de recolectar, procesar, limpiar y transformar la información necesaria para la toma de decisiones no es una tarea sencilla, sobre todo si consideramos que una entidad tiene distintas áreas, que a veces, se encuentran alejadas de los jefes principales. Por otro lado, se dispone de fuentes de datos cada vez más numerosas, desconectadas entre sí y a menudo incompatibles. Fuentes de datos que tienen que cambiar a lo largo de la evolución de las estrategias de las entidades. Necesitamos herramientas que nos ayuden a minimizar el tiempo para analizar toda esa información con mayor velocidad y precisión; logrando de esta manera mantenernos actualizados y reaccionar ante los cambios que se produzcan.

El componente de **Inteligencia de Negocios** – Business Intelligence – que resuelve este caos de los datos para una rápida toma de decisiones es el **Almacén de datos** –Data Warehouse (DW).

El Data Warehouse (DW), básicamente se encarga de consolidar, integrar y centralizar los datos que la entidad genera en todos los ámbitos de una actividad o proceso (Compras, Ventas, Producción, entre otros), para luego ser almacenados mediante una estructura que permite el acceso y exploración de la

información requerida con buena actualización, facilitando posteriormente, una amplia gama de posibilidad de análisis multivariados, que permitirá la toma de decisiones estratégicas y tácticas.

En Cuba y específicamente en las FAR el uso y explotación de técnicas de DW es muy joven, pero se está trabajando para que se incremente este desarrollo, producto de la importancia y necesidad que tienen las entidades de obtener resultados a partir de esta tecnología.

El Minfar se encuentra inmerso en un constante proceso de desarrollo y perfeccionamiento de sus procesos de trabajo, para lo cual se desarrollan varios sistemas informáticos en la UCID, entre los que se encuentra el Sistema de Planificación de Recursos Empresariales para las FAR (ERP-FAR), del cual su columna vertebral lo constituye el SIDEC, siendo este, el marco normativo para la gestión del personal y el material de guerra y a través del cual se garantizan los procesos de perfeccionamiento y organización de las estructuras de las FAR, para satisfacer algunas de las exigencias de la defensa del país. Sin embargo el SIDEC no cuenta con ninguna funcionalidad que permita el almacenamiento y recuperación de la información histórica generada por él, para su posterior análisis.

Después de un análisis de la situación problemática, queda conformado el **problema a resolver** mediante el cuestionamiento de ¿Cómo garantizar los procesos de almacenamiento y recuperación de la información histórica generada por el SIDEC para su posterior análisis?

Este problema presenta como **objeto de estudio** los Sistemas informacionales. Teniendo como **campo de acción** el Almacén de Datos para el SIDEC.

Por lo que el **objetivo general** consiste en Realizar el análisis y el diseño de un almacén de datos que posterior a su implementación garantice los procesos de almacenamiento y recuperación de la información histórica generada por el SIDEC para su posterior análisis.

Para dar cumplimiento al objetivo general fueron identificados los siguientes **objetivos específicos**:

1. Realizar el estado del arte.
2. Realizar análisis de requerimientos.
3. Realizar análisis de los OLTP.
4. Realizar el modelo lógico de la estructura del DW.

Por cada objetivo fue identificada una o más **tareas investigativas**, a continuación serán listadas según su relación con el objetivo en específico.

1. Realizar el estudio del arte.
 - Realizar el análisis de la bibliografía del tema.
 - Definir las herramientas y metodologías a utilizar.
2. Realizar análisis de requerimientos.
 - Identificar preguntas.
 - Identificar indicadores y perspectivas de análisis.
 - Elaborar modelos conceptuales.
3. Realizar análisis de los OLTP.
 - Establecer correspondencia con los requerimientos.
 - Seleccionar los campos que integrarán cada perspectiva.
4. Realizar el modelo lógico de la estructura del DW.
 - Diseñar tablas de dimensiones.
 - Diseñar tablas de hechos.

Tomando como **idea a defender**: Si se realiza el análisis y el diseño de un almacén de datos que permita su implementación entonces se podrán garantizar los procesos de almacenamiento y recuperación de la información histórica generada por el SIDEC para su posterior análisis

Estructura del trabajo:

Capítulo 1: Fundamentación teórica: se realiza un estudio detallado del estado del arte de los DW. Se hace referencia a las herramientas, técnicas y metodologías que se utilizan en la UCID para el desarrollo de los DW, así como sus características.

Capítulo 2: Características del sistema: se realiza un análisis que ayudará a conocer el funcionamiento y accionar del sistema, lo que permitirá examinar e interpretar de forma óptima las necesidades de información de la misma.

Capítulo 3: Análisis y diseño del sistema: Se realiza el análisis del sistema y se modela el diseño.

CAPÍTULO 1: FUNDAMENTACIÓN TEÓRICA.

1.1 Introducción

En este capítulo se realiza una descripción sobre los diferentes aspectos relacionados con los sistemas informacionales que existen hoy en día, especificando algunos de ellos, así como explicando brevemente a lo que se dedican. Posterior a la revisión de estos se enfatizará en las peculiaridades de los sistemas DW. Se realiza un estudio sobre su empleo actual, tanto en el mundo como en Cuba. Además, se mencionan las diferentes herramientas, tecnologías y metodologías a utilizar para el desarrollo de este trabajo, las cuales ya se encuentran definidas por el Centro de Apoyo a la Toma de Decisiones (CATD) que pertenece a la UCID.

1.2 Sistemas Operacionales o de Producción

Los Sistemas Operacionales o Sistemas de Producción, son los sistemas que se utilizan para garantizar el desempeño del negocio en tiempo real; es decir, se encargan de registrar y controlar los procesos que constituyen el núcleo del comportamiento diario de la entidad. Entre los objetivos de estos sistemas operacionales, además de apoyar las funciones diarias de la empresa, se encuentran brindar servicios de oficina y entregar la información de manera automatizada, asegurando la calidad y la protección de la misma. (1)

1.3 Sistemas Informacionales

Los Sistemas Informacionales son los sistemas que se utilizan para administrar y controlar la entidad, es decir, se basan en puntos estables en el tiempo o datos históricos que se diseñan principalmente, para ejecutar consultas complejas, no planas o dimensionales y de sólo lectura. El objetivo fundamental de los sistemas informacionales es posibilitar el hecho de mantener disponible un compendio de la información histórica garantizando una fuente única, contribuyendo a realizar análisis y toma de decisiones estratégicas a largo plazo. (1)

1.4 Procesamiento Transaccional en Línea (OLTP)

Uno de los aspectos esenciales que distingue a los sistemas operacionales de los informacionales, es el tipo de procesamiento de la información que realizan, a saber, transaccional o analítico.

Históricamente, los datos de la corporación suelen residir en bases de datos (BD) que se diseñaron principalmente para introducir y almacenar datos, mediante el llamado Proceso de Transacciones On Line

(OLTP). Este método es idóneo para insertar, modificar o borrar registros, pero no lo es tanto para responder a complejas consultas. La relación entre los datos responde a las técnicas llamadas de Entidad-Relación.

Un sistema operacional o de procesamiento en línea es un sistema como el de administración de recursos humanos, de asignación de bancarios de recuperación y control de cartera o de control de seguros, y su función principal es dar el soporte a las necesidades del día; son sistemas normalmente optimizados para el manejo de un conjunto predefinido de transacciones.

Los sistemas operacionales de los cuales se transferirá la información seleccionada, pueden haber sido contruidos utilizando manejadores de datos relacionados, manejadores de archivos jerárquicos, de archivos planos u otro tipo de manejadores. Por ello, es necesario analizar y definir cuidadosamente de los sistemas operacionales aquellos datos que representen la esencia o filosofía del negocio que se pretenda manejar, para que al transferir los datos a la bodega, ese conocimiento primordial se capture en lo que se conoce como metadatos, que son precisamente, los que describen a los datos provenientes de los sistemas operacionales. **¡Error! No se encuentra el origen de la referencia. (1) ¡Error! No se encuentra el origen de la referencia.¡Error! No se encuentra el origen de la referencia.¡Error! No se encuentra el origen de la referencia.**

1.5 Procesamiento analítico en Línea (OLAP)

El Procesamiento Analítico en Línea, OLAP (On Line Analytical Processing), conocido también como Procesamiento para la Toma de Decisiones, soporta las actividades de investigación y navegación del usuario terminal, sustenta el estudio del comportamiento del negocio y su proyección, se caracteriza por un análisis dimensional y dinámico, desde diferentes puntos de vista de los datos consolidados de la empresa, ayudando al usuario a sintetizar la información de la empresa a través de vistas personalizadas, análisis históricos y pronósticos. (1) **¡Error! No se encuentra el origen de la referencia.¡Error! No se encuentra el origen de la referencia.**

1.6 Almacenes de Datos

Un almacén de datos se define como un ambiente para la toma de decisiones que refuerza los datos almacenados en diferentes fuentes, organizándolos y poniéndolos a disposición de los encargados de esta responsabilidad en la empresa, independientemente de su nivel de habilidad técnica o de las plataformas de trabajo; en pocas palabras, es el manejo de los datos y la tecnología del análisis. (1)

1.6.1 Definición de almacén de datos

El almacén de datos posibilita la extracción de datos de sistemas operacionales y fuentes externas, permite la integración y homogenización de los datos de toda la empresa, provee información que ha sido transformada y sumariada, para que ayude en el proceso de toma de decisiones estratégicas y tácticas.

El DW, convertirá entonces los datos operacionales de la empresa en una herramienta competitiva, debido a que pondrá a disposición de los usuarios indicados la información pertinente, correcta e integrada, en el momento que se necesita.

Una de las definiciones más famosas sobre DW, es la de William. H. Inmon, quien define: "Un data warehouse es una colección de datos orientada al negocio, integrada, variante en el tiempo y no volátil para el soporte del proceso de toma de decisiones de la gerencia".

Debido a que W. H. Inmon, es reconocido mundialmente como el padre del DW, la explicación de las características más sobresalientes de esta herramienta se basó en su definición. (2)

Ralph Kimball es otro conocido autor en el tema de los data warehouse, 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 data warehouse 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. (3)

Las definiciones anteriores se centran en los datos en sí mismos. 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 la inteligencia empresarial, herramientas para extraer, transformar y cargar datos en el almacén de datos, y herramientas para gestionar y recuperar los metadatos.

1.6.2 Objetivos generales de los Data Warehouse

- **Hacer accesible la información de la organización:** La información contenida en el Data Warehouse debe ser navegable, fácilmente comprendida por los usuarios, y sobre todo de acceso rápido.
- **Hacer que la información de la organización sea consistente:** La información de un departamento de la organización puede ser contrastada con la información de otro departamento. Si dos mediciones tienen el mismo nombre, entonces significan lo mismo, por el contrario, si dos

mediciones representan conceptos diferentes, deben llamarse de distinto modo de manera que toda la información sea la correcta y esté al día.

- **Ser una fuente adaptable de información:** El Data Warehouse está diseñado para afrontar con éxito continuos cambios. Cuando surgen nuevas necesidades de información, nuevas preguntas o nuevos datos añadidos, las tecnologías y los datos existentes no deben verse afectadas. El diseño de núcleos de información separados (Data Marts) debe ser distribuido e incremental.
- **Ser la base para la toma de decisiones:** Los datos contenidos en el Data Warehouse son adecuados para justificar decisiones estratégicas de la organización. Las decisiones se toman una vez que el Data Warehouse ha aportado los datos que las justifican. (2)

1.6.3 Características de los almacenes de datos

Del estudio realizado, distintas son las características que se perciben en varias bibliografías acerca del tema, pero sería muy preciso señalar que no pocos de estos materiales coinciden en algunas de estas. Las características a continuación son las descritas por W. H. Inmon. (4)

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

Orientada a tema: 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 queden 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 solo lectura, y se mantiene para futuras consultas.

1.6.4 Estructura básica de la arquitectura DW

Datos operacionales: un origen de datos para el componente de almacenamiento físico DW.

Extracción de datos: selección sistemática de datos operacionales usados para poblar el componente de almacenamiento físico DW.

Transformación de datos: Procesos para realizar sumario u otros cambios en los datos operacionales para reunir los objetivos de orientación a temas e integración principalmente.

Carga de datos: inserción sistemática de datos en el componente de almacenamiento físico DW.

Data Warehouse: almacenamiento físico de datos de la arquitectura DW.

Herramientas de Acceso al componente de almacenamiento físico DW: herramientas que proveen acceso a los datos.

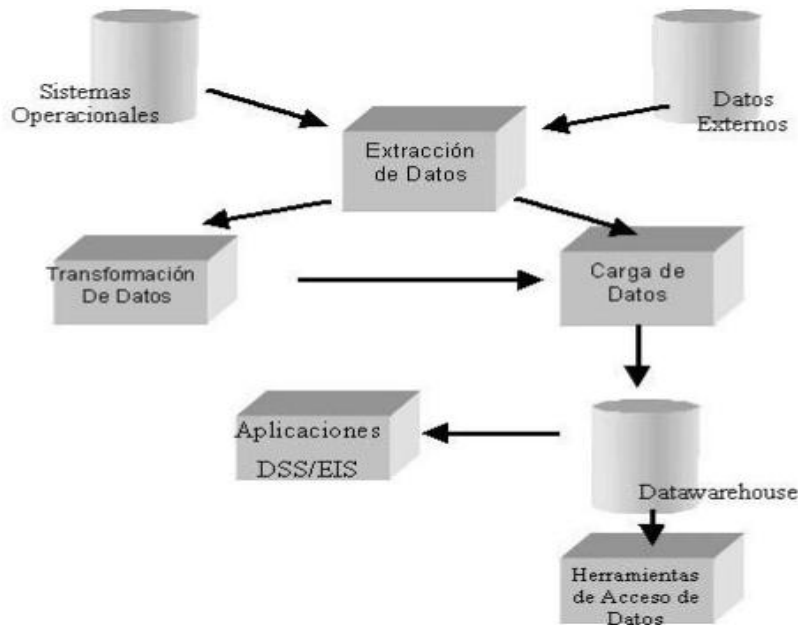


Figura 1: Estructura Básica DW.

1.6.5 Ventajas

A continuación se enumerarán algunas de las ventajas más sobresalientes que trae aparejada la implementación de un DW y que ejemplifican de mejor modo sus características y cualidades: (2)

- ✓ Transforma datos orientados a las aplicaciones en información orientada a la toma de decisiones.
- ✓ Integra y consolida diferentes fuentes de datos y departamentos empresariales, que anteriormente formaban islas, en una única plataforma sólida y centralizada.
- ✓ Provee la capacidad de analizar y explotar las diferentes áreas de trabajo y de realizar un análisis inmediato de las mismas.
- ✓ Elimina la producción y el procesamiento de datos que no son utilizados ni necesarios, producto de aplicaciones mal diseñadas o ya no utilizadas.

- ✓ Mejora la entrega de información, es decir, información completa, correcta, consistente, oportuna y accesible. Información que los usuarios necesitan, en el momento adecuado y en el formato apropiado.
- ✓ Logra un impacto positivo sobre los procesos empresariales. Cuando los usuarios tienen acceso a una mejor calidad de información, la empresa puede lograr por sí misma: aprovechar el enorme valor potencial de sus recursos de información y transformarlo en valor verdadero; eliminar los retardos de los procesos empresariales que resultan de información incorrecta, inconsistente y/o inexistente; integrar y optimizar procesos a través del uso compartido e integrado de las fuentes de información; permitir al usuario adquirir mayor confianza acerca de sus propias decisiones y de las del resto, y lograr así, un mayor entendimiento de los impactos ocasionados.
- ✓ Los usuarios pueden acceder directamente a la información en línea, lo que contribuye a su capacidad para operar con mayor efectividad en las tareas rutinarias o no. Además, pueden tener a su disposición una gran cantidad de valiosa información multidimensional, presentada coherentemente como fuente única, confiable y disponible en sus estaciones de trabajo. Así mismo, los usuarios tienen la facilidad de contar con herramientas que les son familiares para manipular y evaluar la información obtenida en el DW, tales como: hojas de cálculo, procesadores de texto, software de análisis de datos, software de análisis estadístico, reportes, etc. **¡Error! No se encuentra el origen de la referencia.¡Error! Marcador no definido.**
- ✓ Permite la toma de decisiones estratégicas, tácticas y operativas.

1.7 Aplicación de los almacenes de datos en el mundo.

Hoy en día en el mundo son utilizados los almacenes de datos de forma significativa por miles de empresas y entidades, por ejemplo la compañía Sun Microsystems establece record mundial con el data warehouse más grande del mundo. Sun ofrece tecnologías líderes en la industria, enfocadas a toda la infraestructura de tecnologías informáticas, en la que hardware, software y servicios se combinan en una sola. Con una sólida reputación de compartir experiencias, construir comunidades, e innovación, Sun crea oportunidades, tanto social como económicamente, alrededor del mundo. (5)

Otro ejemplo lo podemos ver con la compañía VISA UE, la cual ha aprovechado el potencial de la extensa cantidad de datos financieros de los que dispone, para ayudar a sus bancos socios a entender el

CAPÍTULO I. FUNDAMENTACIÓN TEÓRICA

comportamiento de los consumidores mucho mejor que antes. VISA UE, junto con Sun Microsystems y Accenture, han creado uno de los almacenes de datos más grandes de Europa para almacenar detalles de los 3.000 millones de transacciones con tarjetas de crédito y débito que se realizan al año en 26 países europeos. Los bancos podrán acceder a los datos de diferentes formas para descubrir los cambios en las tendencias de la economía y conocer con más detalle los patrones de gasto. Todo ello, dentro del estricto marco legal de la Directiva de Protección de Datos de la Unión Europea. Los datos individuales de los clientes nunca son develados. Los 6.000 bancos miembros de VISA UE podrán también evaluarse a sí mismos y compararse con los competidores de su misma categoría, a través de una serie de indicadores clave de negocio y el comportamiento de los consumidores. VISA UE también es consciente del potencial interés que tiene su almacén de datos para las consultoras que dependen fuertemente del análisis de datos en bruto y para los vendedores que necesitan conocer los cambios en los patrones de consumo, aunque no sean empresas con proyectos en esta área. (6)

También tenemos a la compañía de DATATUR, que contienen un almacén de datos para el análisis y difusión de la información estadística del turismo en España. Se da una visión global del sistema tanto desde su perspectiva metodológica como técnica. El objetivo de esta compañía es poner al alcance de los analistas de datos estadísticos del turismo, una herramienta que permita analizar esta información y trabajar con ella de un modo más sencillo y poder además publicar esta información a través de internet para ponerla al alcance de todo analista que la necesite. En general el sistema DATATUR se ha desarrollado para facilitar y agilizar el almacenamiento, recuperación, análisis y difusión de la información. (7)

1.7.1 Soluciones cubanas.

Como mencionamos anteriormente, en nuestro país, el desarrollo de los almacenes de datos es muy joven, sin embargo ya existen algunos resultados de los cuales el más sobresaliente resulta ser el Sistema Data Warehouse Comercial de la Corporación CIMEX.

La Corporación CIMEX se dedica fundamentalmente a la Exportación e Importación de mercancías. El Sistema Data Warehouse comercial de la Corporación CIMEX centra su atención en la actividad del comercio, principalmente en la gestión de inventario, permitiendo una gestión de compra-venta eficiente, con una finalidad fundamental: "Disminuir los costos, sin afectar al cliente, permitiendo prestaciones eficientes y con la calidad requerida, aumentando las ganancias o utilidades de las empresas". (8)

1.8 Herramientas

Siguiendo las políticas definidas para el desarrollo de software en las FAR, las herramientas a emplear para la ejecución del presente trabajo de tesis, son de uso libre, aprovechando además, entre otras ventajas, la posibilidad de intercambio de experiencia.

Se debe señalar que las herramientas a utilizar en el desarrollo de este proyecto, tanto en su diseño como en su posterior implementación están definidas por el CATD de la UCID.

A continuación se mencionarán las herramientas que se utilizarán para el desarrollo de este trabajo:

1.8.1 Plataforma BI

Se optó por utilizar la plataforma suite Pentaho Open Source Business Intelligence como plataforma BI. Esto es debido a que se han creado muchos módulos que se adaptan a esta plataforma. Además es una alternativa de licencia libre. Pentaho funciona sobre cualquier navegador, soporta la ejecución de Dashboards y reportes en tiempo real, permite gestionar usuarios, colgar documentación, monitorear la ejecución de “jobs” y finalmente, posee una versión de demostración que permite aprender rápido el manejo de la plataforma, reduciendo la curva de aprendizaje, contando con ejemplos de los cuales se puede partir. La documentación para consultar es basta. Tiene una interfaz amigable para los usuarios, la que es posible editar para ajustarse más a las necesidades de estos.

1.8.2 ETL

Para el proceso de ETL, se utilizará Kettle Data Integration. Kettle es una herramienta libre que se encarga de este proceso y que ahora forma parte del proyecto Pentaho. Kettle posee los módulos Pan y Spoon, para diseñar el proceso de ETL. Kitchen y Chef, para la programación de los Jobs de la ETL. En general, Kettle es una herramienta gráfica, lo que facilita su aprendizaje y su uso; puede trabajar con diversas fuentes de datos y conectarse a muchos motores de base de datos, tanto como fuente, como para destino. Otra de las ventajas de Kettle, es que posee rutinas que facilitan el proceso de limpieza de datos.

1.8.3 PostgreSQL 8.3 (Gestor de Base de Datos)

PostgreSQL 8.3 es un sistema de base de datos relacional perteneciente al ámbito del software libre que destaca por su robustez, escalabilidad y cumplimiento de los estándares SQL.

Incluye otras funcionalidades como:

- Posibilidad de almacenar y manipular cubos multidimensionales de información, unidad fundamental de almacenamiento de los Data Warehouse.
- Alta capacidad de almacenar un inmenso volumen de información, con una alta velocidad de respuesta en las consultas por muy complejas y/o extensas que sean.
- Eficiente gestión de la seguridad: Posee entre sus características el manejo eficiente de usuarios, permisos y roles, que tendrán acceso a distintas partes de los cubos multidimensionales, en cuyo contenido también estará almacenado información sensible por lo que se requiere de una factible gestión de la seguridad.
- Alta compatibilidad con las herramientas de Extracción, Transformación y Carga (ETL por sus siglas en inglés).

1.8.4 Apache (Servidor de Aplicaciones Web)

Es un servidor web flexible, rápido y eficiente, continuamente actualizado y adaptado a los nuevos protocolos (HTTP 1.1). Entre sus características se destacan:

- Multiplataforma
- Es un servidor web conforme al protocolo HTTP/1.1.
- Modular: Puede ser adaptado a diferentes entornos y necesidades, con los diferentes módulos de apoyo que proporciona, y con la Interfaz de Programación de Aplicaciones (API) de programación de Módulos, para el desarrollo de módulos específicos.

1.8.5 Mozilla (Navegador Web)

Mozilla Firefox es un navegador web libre descendiente de Mozilla Application Suite, desarrollado por la Corporación Mozilla, la Fundación Mozilla y un gran número de voluntarios externos. Es el segundo navegador más usado detrás de todas las versiones sumadas de Internet Explorer.

Firefox es un navegador multiplataforma y está disponible en varias versiones de Microsoft Windows, Mac OS X, GNU/Linux y algunos sistemas basados en Unix. Su código fuente es software libre, publicado bajo una triple licencia GPL/LGPL/MPL. Para visualizar páginas web, Firefox usa el motor de renderizado

Gecko, que implementa algunos estándares web actuales además de otras funciones, algunas de las cuales están destinadas a anticipar probables adiciones a los estándares web.

Incluye navegación por pestañas, corrector ortográfico, búsqueda progresiva, marcadores dinámicos, un administrador de descargas y un sistema de búsqueda integrado que utiliza el motor de búsqueda que desee el usuario.

1.8.6 Herramienta CASE. Visual Paradigm

Como herramienta para modelar visualmente se utilizó Visual Paradigm. Visual Paradigm, genera documentos, está disponible para varios sistemas operativos como son: Windows, Linux y Unix.

Es una herramienta amigable para el usuario, puede ser usada en varios idiomas y cada componente utilizado en el diagrama que se esté creando. Sugiere nuevos posibles componentes a utilizar, por lo que ya no es necesario localizarlos en la barra y así se crea fácilmente cualquier tipo de diagrama. Posee un amplio número de estereotipos que proporciona la creación de diagramas de fácil entendimiento, además de que estos diagramas los organiza automáticamente.

1.9 Metodología

Conjunto de procedimientos, técnicas y herramientas que ayuda a los desarrolladores a realizar un nuevo producto.

1.9.1 HEFESTO

HEFESTO es una metodología propia para la construcción de los almacenes de datos, cuya propuesta está fundamentada en una amplia investigación, comparación de metodologías existentes y experiencias propias en procesos de confección de almacenes de datos. (2)

Con esta metodología, la construcción e implementación de un almacén de datos puede adaptarse muy bien a cualquier ciclo de vida de desarrollo de software, con la salvedad de que para algunas fases en particular, las acciones que se han de realizar serán diferentes. No se requieren fases extensas de reunión de requerimientos y análisis, fases de desarrollo monolítico que conlleven demasiado tiempo y fases de despliegue muy largas. Garantiza la entrega de una primera implementación que satisfaga una parte de las necesidades, para demostrar las ventajas del DW y motivar a los usuarios.

La metodología HEFESTO, puede ser embebida en cualquier ciclo de vida que cumpla con la condición antes declarada.

Esta metodología cuenta con las siguientes características:

- ❖ Los objetivos y resultados esperados en cada fase se distinguen fácilmente y son sencillos de comprender.
- ❖ Se basa en los requerimientos del usuario, por lo cual su estructura es capaz de adaptarse con facilidad y rapidez ante los cambios en el negocio.
- ❖ Reduce la resistencia al cambio, ya que involucra al usuario final en cada etapa para que tome decisiones respecto al comportamiento y funciones del DW.
- ❖ Utiliza modelos conceptuales y lógicos, los cuales son sencillos de interpretar y analizar.
- ❖ Es independiente del tipo de ciclo de vida que se emplee para contener la metodología.
- ❖ Es independiente de las herramientas que se utilicen para su implementación.
- ❖ Es independiente de las estructuras físicas que contengan el DW y de su respectiva distribución.
- ❖ Cuando se culmina con una fase, los resultados obtenidos se convierten en el punto de partida para llevar a cabo el paso siguiente.
- ❖ Se aplica tanto para Data Warehouse como para Data Mart.

1.10 Conclusiones del capítulo

En este capítulo se ha abordado descriptivamente las principales características, definiciones y aspectos relacionados con los DW. Se realizó un estudio de los sistemas informacionales que actualmente existen. Se expusieron las distintas herramientas y la metodología a emplear para el desarrollo de este trabajo de diploma. Fueron abordados varios conceptos, definiciones y características sobre muchos de los términos relacionados con el sistema que se desarrolla.

CAPÍTULO 2 CARACTERÍSTICAS DEL SIDEC

2.1 Introducción

El objetivo principal de este capítulo, es el de obtener e identificar las necesidades de información clave de alto nivel, que son esenciales para llevar a cabo las metas y estrategias de la entidad, y que facilitará la toma de decisiones de manera más eficiente y eficaz. El análisis de los requerimientos de los diferentes usuarios, es el punto de partida de este trabajo, ya que ellos son los que deben, en cierto modo, guiar la investigación hacia un desarrollo que refleje claramente lo que se espera del depósito de datos, en relación con sus funciones y cualidades. En este capítulo solo se trabajará con las siguientes actividades de la metodología HEFESTO:

- Identificar preguntas.
- Identificar Indicadores y perspectivas de análisis.
- Modelo Conceptual.

2.2 Análisis de los requerimientos

2.2.1 Identificar preguntas

Una forma de asegurarse de que se ha realizado un buen análisis, es que el resultado del mismo debe hacer explícitos los objetivos estratégicos planteados por la entidad que se está estudiando. En el caso que se ha desarrollado, se ha encaminado al relevamiento, en pos de enfocar las necesidades de información en los procesos principales que desarrolla el SIDEC. La idea central es, que se formulen preguntas complejas sobre el negocio, que incluyan variables de análisis que se consideren relevantes, ya que son estas las que permitirán estudiar la información desde diferentes perspectivas.

Se indagó con los usuarios en busca de sus necesidades de información, pero las mismas abarcaban casi todas las actividades de la entidad, por lo cual se les pidió que escogieran las que considerasen más importantes dentro del proceso de perfeccionamiento estructural y que a su vez estuviesen soportadas por algún OLTP. A continuación, se procedió a identificar que era lo que les interesaba conocer acerca de este proceso y cuáles eran las variables o perspectivas que debían tenerse en cuenta para poder tomar decisiones basadas en ello.

De conjunto con los especialistas, se identificaron las siguientes necesidades informativas:

CAPÍTULO II. CARACTERÍSTICAS DEL SIDEC

- Cantidad de Órganos en TP de cada estructura en un tiempo determinado.
- Cantidad de Órganos en TG de cada estructura en un tiempo determinado.
- Cantidad de Plantillas en TP de cada estructura en un tiempo determinado.
- Cantidad de Plantillas en TG de cada estructura en un tiempo determinado.
- Cantidad de Órganos en TP de cada estructuraop de cada estructura en un tiempo determinado.
- Cantidad de Órganos en TG de cada estructuraop de cada estructura en un tiempo determinado.
- Cantidad de Completos en TP de cada estructuraop de cada estructura en un tiempo determinado.
- Cantidad de Completos en TG de cada estructuraop de cada estructura en un tiempo determinado.
- Cantidad de Cargos Civiles en TP de cada cargo civil de cada estructuraop de cada estructura en un tiempo determinado.
- Cantidad de Cargos Militares en TP de cada cargo militar de cada estructuraop de cada estructura en un tiempo determinado.
- Cantidad de Cargos Militares en TG de cada cargo militar de cada estructuraop de cada estructura en un tiempo determinado.
- Cantidad de Módulos en TP de cada cargo militar de cada estructuraop de cada estructura en un tiempo determinado.
- Cantidad de Módulos en TG de cada cargo militar de cada estructuraop de cada estructura en un tiempo determinado.
- Cantidad de Técnicas en TP de cada técnica de cada estructuraop de cada estructura en un tiempo determinado.
- Cantidad de Técnicas en TG de cada técnica de cada estructuraop de cada estructura en un tiempo determinado.
- Cantidad de Directivas de cada estructura en un tiempo determinado.
- Cantidad de Ajustes de cada estructuraop de cada estructura de cada directiva en un tiempo determinado.

2.2.2 Identificar indicadores y perspectivas de análisis

Una vez que se han establecido los requerimientos, se debe proceder a su descomposición para descubrir los indicadores que se utilizarán y las perspectivas de análisis que intervendrán.

Indicadores: Constituyen valores numéricos y representan lo que se desea analizar concretamente, por ejemplo: por cientos, promedios, cantidades, sumatorias, fórmulas.

Perspectivas: se refieren a los objetos mediante los cuales se quiere examinar los indicadores, con el fin de responder a las preguntas planteadas, por ejemplo: clientes, proveedores, sucursales, países, productos, estructuras. Cabe destacar, que el Tiempo es comúnmente una perspectiva.

A continuación se procederá a identificar las perspectivas para el análisis de los indicadores, serán mencionadas y analizadas tanto estas perspectivas como el objetivo que se obtendrá.

1. Cantidad de Órganos en TP de cada estructura en un tiempo determinado.

INDICADOR

PERSPECTIVAS

2. Cantidad de Órganos en TG de cada estructura en un tiempo determinado.

INDICADOR

PERSPECTIVAS

3. Cantidad de Plantillas en TP de cada estructura en un tiempo determinado.

INDICADOR

PERSPECTIVAS

4. Cantidad de Plantillas en TG de cada estructura en un tiempo determinado.

INDICADOR

PERSPECTIVAS

5. Cantidad de Órganos en TP de cada estructuraop de cada estructura en un tiempo determinado.

CAPÍTULO II. CARACTERÍSTICAS DEL SIDEC

INDICADOR

PERSPECTIVAS

6. Cantidad de Órganos en TG de cada estructuraop de cada estructura en un tiempo determinado.

INDICADOR

PERSPECTIVAS

7. Cantidad de Completos en TP de cada estructuraop de cada estructura en un tiempo determinado.

INDICADOR

PERSPECTIVAS

8. Cantidad de Completos en TG de cada estructuraop de cada estructura en un tiempo determinado.

INDICADOR

PERSPECTIVAS

9. Cantidad de Cargos Civiles en TP de cada cargo civil de cada estructuraop de cada estructura en un

INDICADOR

PERSPECTIVAS

tiempo determinado.

PERSPECTIVAS

10. Cantidad de Cargos Militares en TP de cada cargo militar de cada estructuraop de cada estructura en

INDICADOR

PERSPECTIVAS

un tiempo determinado.

PERSPECTIVAS

11. Cantidad de Cargos Militares en TG de cada cargo militar de cada estructuraop de cada estructura

INDICADOR

en un tiempo determinado.

PERSPECTIVAS

PERSPECTIVAS

12. Cantidad de Módulos en TP de cada cargo militar de cada estructuraop de cada estructura en un

INDICADOR

tiempo determinado.

PERSPECTIVAS

PERSPECTIVAS

13. Cantidad de Módulos en TG de cada cargo militar de cada estructuraop de cada estructura en un

INDICADOR

tiempo determinado.

PERSPECTIVAS

PERSPECTIVAS

14. Cantidad de Técnicas en TP cada técnica de cada estructuraop de cada estructura en un tiempo

CAPÍTULO II. CARACTERÍSTICAS DEL SIDEC

INDICADOR

determinado.

PERSPECTIVAS

15. Cantidad de Técnicas en TG cada técnica de cada estructuraop de cada estructura en un tiempo

INDICADOR

determinado.

PERSPECTIVAS

16. Cantidad de Directivas de cada estructura en un tiempo determinado.

INDICADOR

PERSPECTIVAS

17. Cantidad de Ajustes de cada estructuraop de cada estructura de cada directiva en un tiempo

INDICADOR

determinado

PERSPECTIVAS

En síntesis los indicadores y las perspectivas de análisis son:

INDICADORES	PERSPECTIVAS
<ul style="list-style-type: none"> ▪ Cantidad de Plantillas en TP ▪ Cantidad de Plantillas en TG ▪ Cantidad de Órganos en TP 	<ul style="list-style-type: none"> ▪ Estructura ▪ Estructuraop ▪ Cargo Civil

<ul style="list-style-type: none"> ▪ Cantidad de Órganos en TG ▪ Cantidad de Cargos Civiles en TP ▪ Cantidad de Cargos Militares en TP ▪ Cantidad de Cargos Militares en TG ▪ Cantidad de Técnicas en TP ▪ Cantidad de Técnicas en TG ▪ Cantidad de Módulos en TP ▪ Cantidad de Módulos en TG ▪ Cantidad de Completos en TP ▪ Cantidad de Completos en TG ▪ Cantidad de Directivas ▪ Cantidad de Ajustes. 	<ul style="list-style-type: none"> ▪ Cargo Militar ▪ Técnica ▪ Directiva ▪ Tiempo
---	---

Tabla 1: Indicadores y Perspectivas de análisis.

2.2.3 Modelo Conceptual

A continuación se construirá el **modelo conceptual** a partir de la identificación de los **indicadores** y **perspectivas** obtenidas. A través de este modelo, se podrá observar con claridad cuáles son los alcances del proyecto, para luego poder trabajar sobre ellos, y así poseer un alto nivel de definición de los datos, permitiendo que puedan ser presentados ante los usuarios y explicados con facilidad.

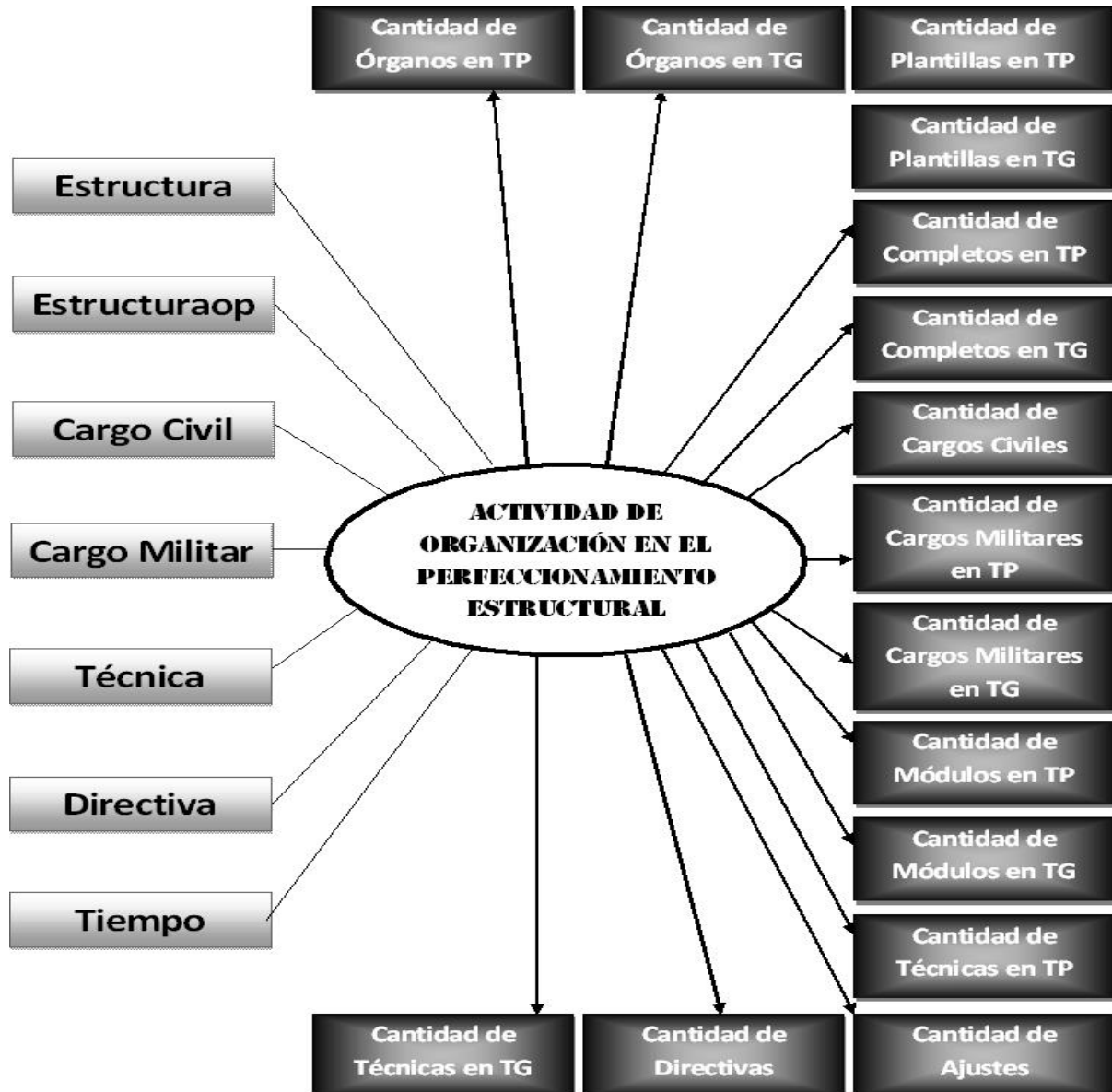


Figura 2: Modelo Conceptual: Actividad de Organización en el Perfeccionamiento Estructural.

La relación mediante la cual se unen las diferentes perspectivas, (se encuentran en el extremo izquierdo del gráfico), para obtener como resultado los indicadores requeridos por los usuarios, (se encuentran en el extremo derecho del gráfico), es precisamente la “Actividad de Organización en el Perfeccionamiento Estructural”.

2.3 Dimensión Tiempo

En un DW, la dimensión Tiempo es obligatoria, y la definición de granularidad y jerarquía de la misma depende de la dinámica del negocio que se esté analizando, toda la información dentro de la bodega, como ya se ha explicado, posee su propio sello de tiempo que determina la ocurrencia y ubicación con elementos en iguales condiciones, representando de esta manera diferentes versiones de una misma situación. Es importante tener en cuenta que el tiempo no es solo una secuencia cronológica representada de forma numérica, sino que posee fechas especiales que inciden notablemente en las actividades de la organización. Esto se debe a que los usuarios podrán por ejemplo analizar los ajustes realizados a una estructura teniendo en cuenta el día de la semana en que se produjeron, quincena, mes, trimestre, semestre, año, entre otros. Existen muchas maneras de diseñar esta tabla, y en adición a ello no es una tarea sencilla de llevar a cabo. Por estas razones se considera una buena práctica evaluar con cuidado la temporalidad de los datos, la forma en que trabaja la organización, los resultados que se esperan obtener del almacén de datos relacionados con una unidad de tiempo y la flexibilidad que se desea obtener de dicha tabla.

2.4 Conclusiones del capítulo

El principal objetivo de este capítulo lo constituyen las metas específicas que los especialistas deberían cumplir. A partir de un análisis de las mismas fueron obtenidas las funcionalidades que debe realizar el DW. Fue realizado el artefacto Modelo Conceptual para que a través de los indicadores y las perspectivas se tuviese un mejor entendimiento de estas funcionalidades del sistema.

CAPÍTULO 3 ANÁLISIS Y DISEÑO DEL ALMACÉN DE DATOS

3.1 Introducción

El objetivo de este capítulo, es el de examinar los OLTP disponibles que contengan la información requerida, como así también sus características, para poder identificar las correspondencias entre el modelo conceptual y las fuentes de datos. En el caso de los indicadores, deben explicitarse como se calcularán, más aún si son fórmulas u operaciones complejas. Se establecerá las relaciones con los OLTP. Se examinarán y seleccionarán los campos que contendrá cada perspectiva, ya que será a través de estos por los que se manipularán y filtrarán los indicadores. Se confeccionará el modelo lógico de la estructura del DW, teniendo como base el modelo conceptual que ya ha sido creado. Se seleccionará cuál será el tipo de esquema que se utilizará para contener la estructura del depósito de datos, que se adapte mejor a los requerimientos y necesidades del usuario. Se diseñarán las tablas de dimensiones, para ello se tomará cada perspectiva con sus atributos relacionados y se les realizará el siguiente proceso:

- Se elegirá un nombre que identifique la dimensión.
- Se añadirá un campo que represente su clave principal.
- Se redefinirán los nombres de los atributos si es que no son lo bastante explicativos.

Por último, en este capítulo se diseñarán las tablas de hechos, se le realizarán las uniones con las tablas de dimensiones y se determinará la jerarquía.

3.2 Establecer correspondencias con los requerimientos

Para realizar este análisis fueron extraídos fragmentos del modelo entidad relación de la base de datos del SIDEC. A continuación, será expuesta la correspondencia entre estos modelos.

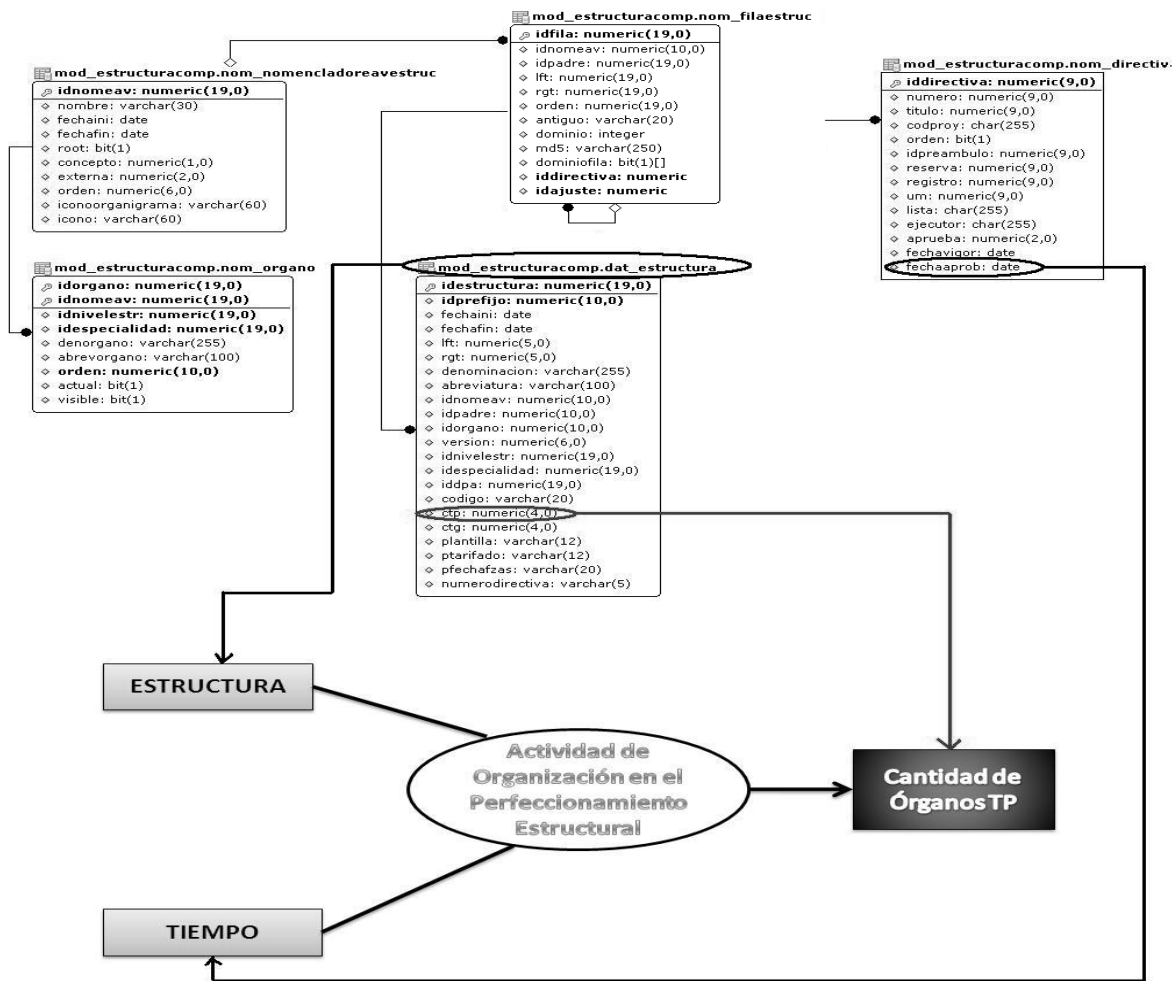


Figura 3: Correspondencia de Requerimientos, Cantidad de Órganos en TP.

Relaciones identificadas:

- La tabla “mod_estructuracomp.dat_estructura” se relaciona con la perspectiva “Estructura”.
- El campo “fechaaprob” de la tabla “mod_estructuracomp.nom_directiva” con la perspectiva “Tiempo” (debido a que es la fecha principal del proceso de la actividad de organización en el perfeccionamiento estructural).
- El campo “ctp” de la tabla “mod_estructuracomp.dat_estructura” con el indicador “Cantidad de Órganos en TP”.

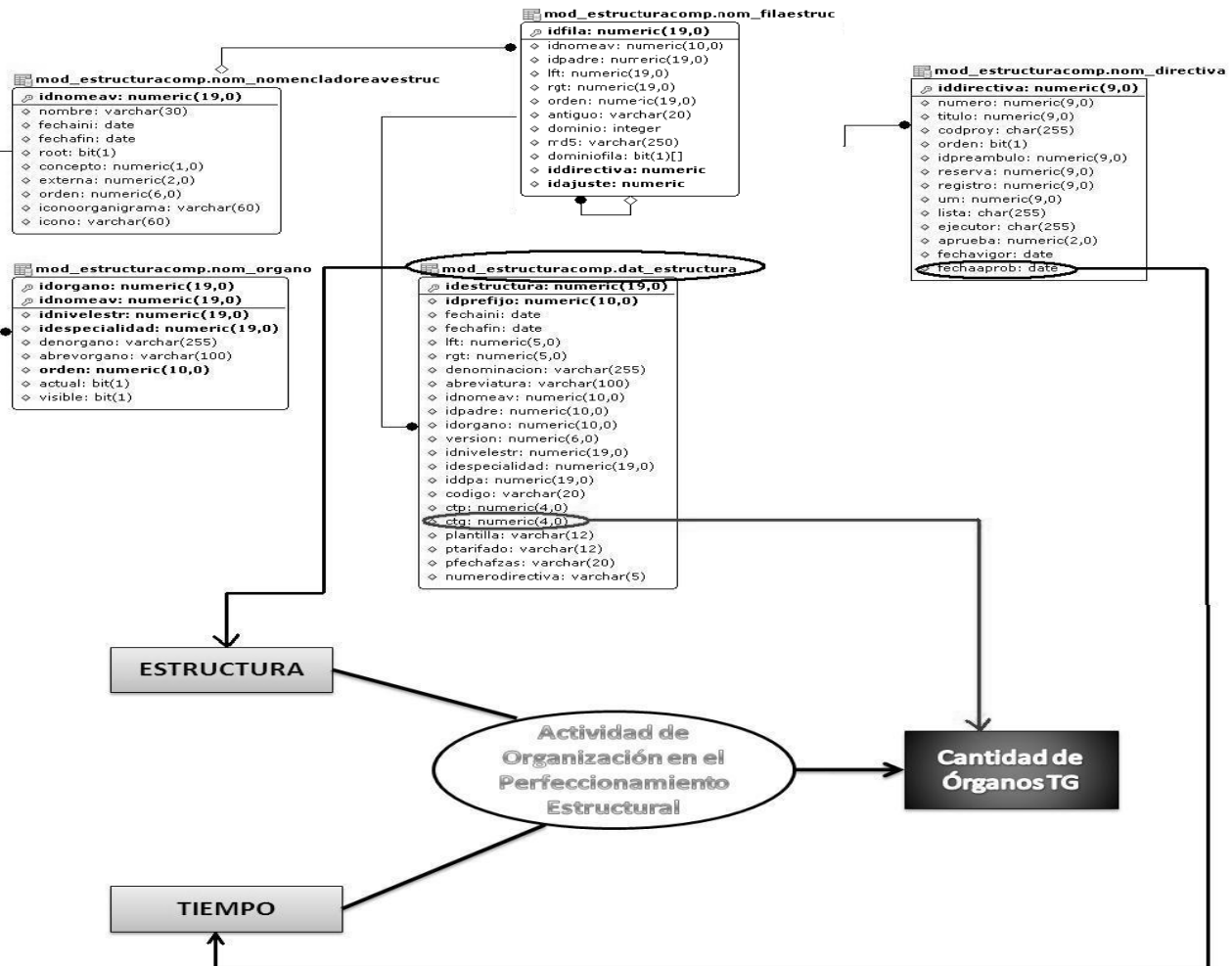


Figura 4: Correspondencia de Requerimientos, Cantidad de Órganos en TG.

Relaciones identificadas:

- La tabla “mod_estructuracomp.dat_estructura” se relaciona con la perspectiva “Estructura”.
- El campo “fechaaprob” de la tabla “mod_estructuracomp.nom_directiva” con la perspectiva “Tiempo” (debido a que es la fecha principal del proceso de la actividad de organización en el perfeccionamiento estructural).
- El campo “ctg” de la tabla “mod_estructuracomp.dat_estructura” con el indicador “Cantidad de Órganos en TG”.

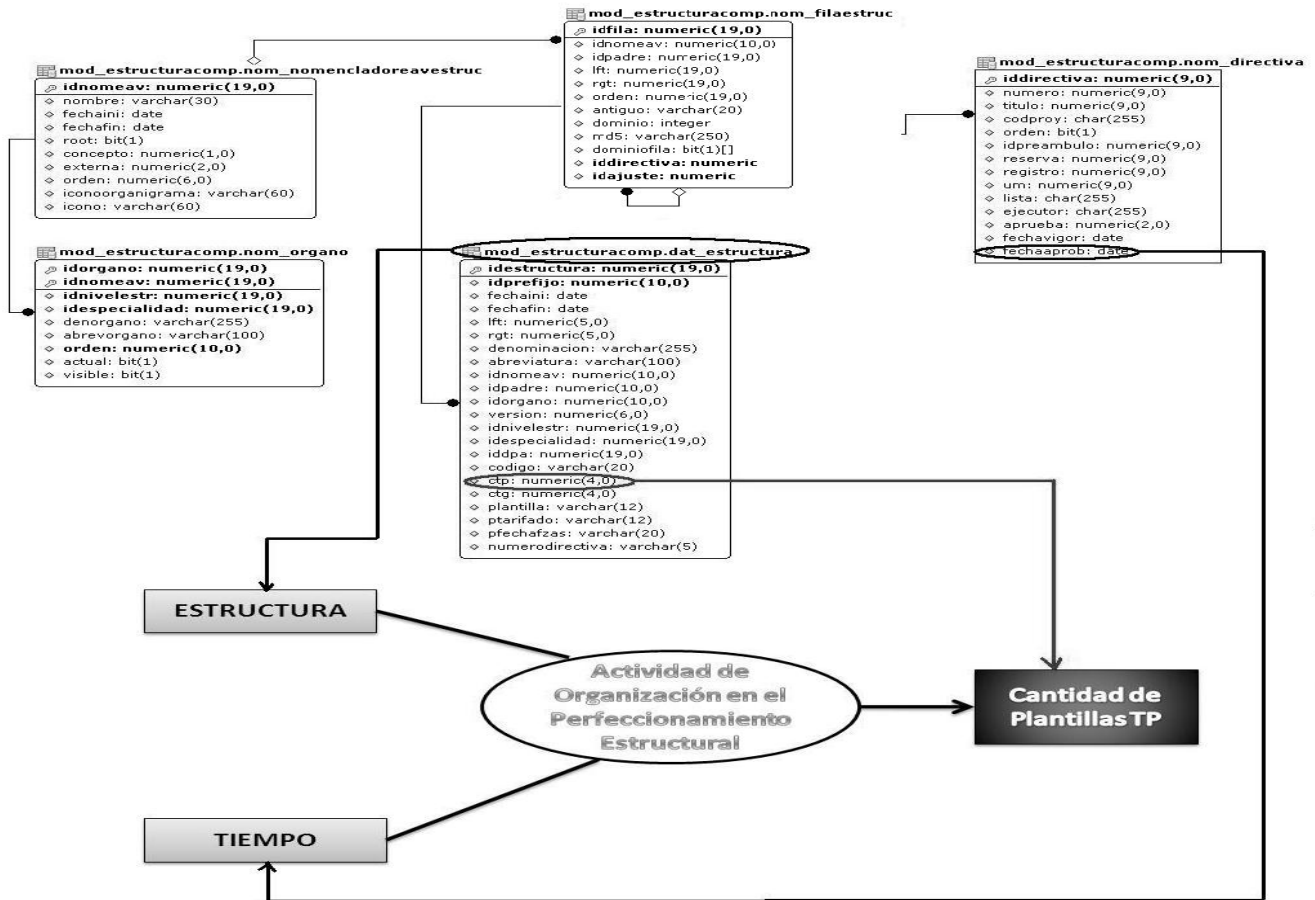


Figura 5: Correspondencia de Requerimientos, Cantidad de Plantillas en TP.

Relaciones identificadas:

- La tabla “mod_estructuracomp.dat_estructura” se relaciona con la perspectiva “Estructura”.
- El campo “fechaaprob” de la tabla “mod_estructuracomp.nom_directiva” con la perspectiva “Tiempo” (debido a que es la fecha principal del proceso de la actividad de organización en el perfeccionamiento estructural).
- El campo “ctp” de la tabla “mod_estructuracomp.dat_estructura” con el indicador “Cantidad de Plantillas en TP”.

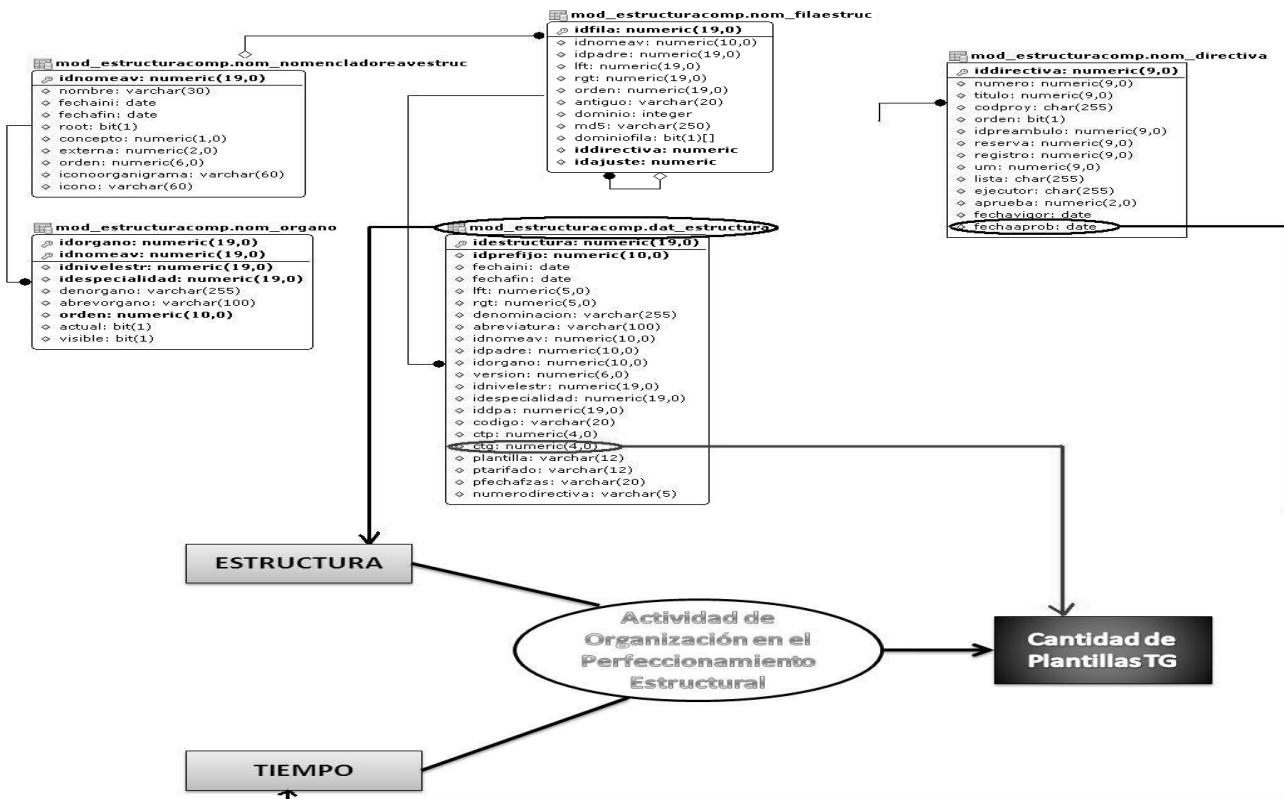


Figura 6: Correspondencia de Requerimientos, Cantidad de Plantillas en TG.

Relaciones identificadas:

- La tabla “mod_estructuracomp.dat_estructura” se relaciona con la perspectiva “Estructura”.
- El campo “fechaaprob” de la tabla “mod_estructuracomp.nom_directiva” con la perspectiva “Tiempo” (debido a que es la fecha principal del proceso de la actividad de organización en el perfeccionamiento estructural).
- El campo “ctg” de la tabla “mod_estructuracomp.dat_estructura” con el indicador “Cantidad de Plantillas en TG”.

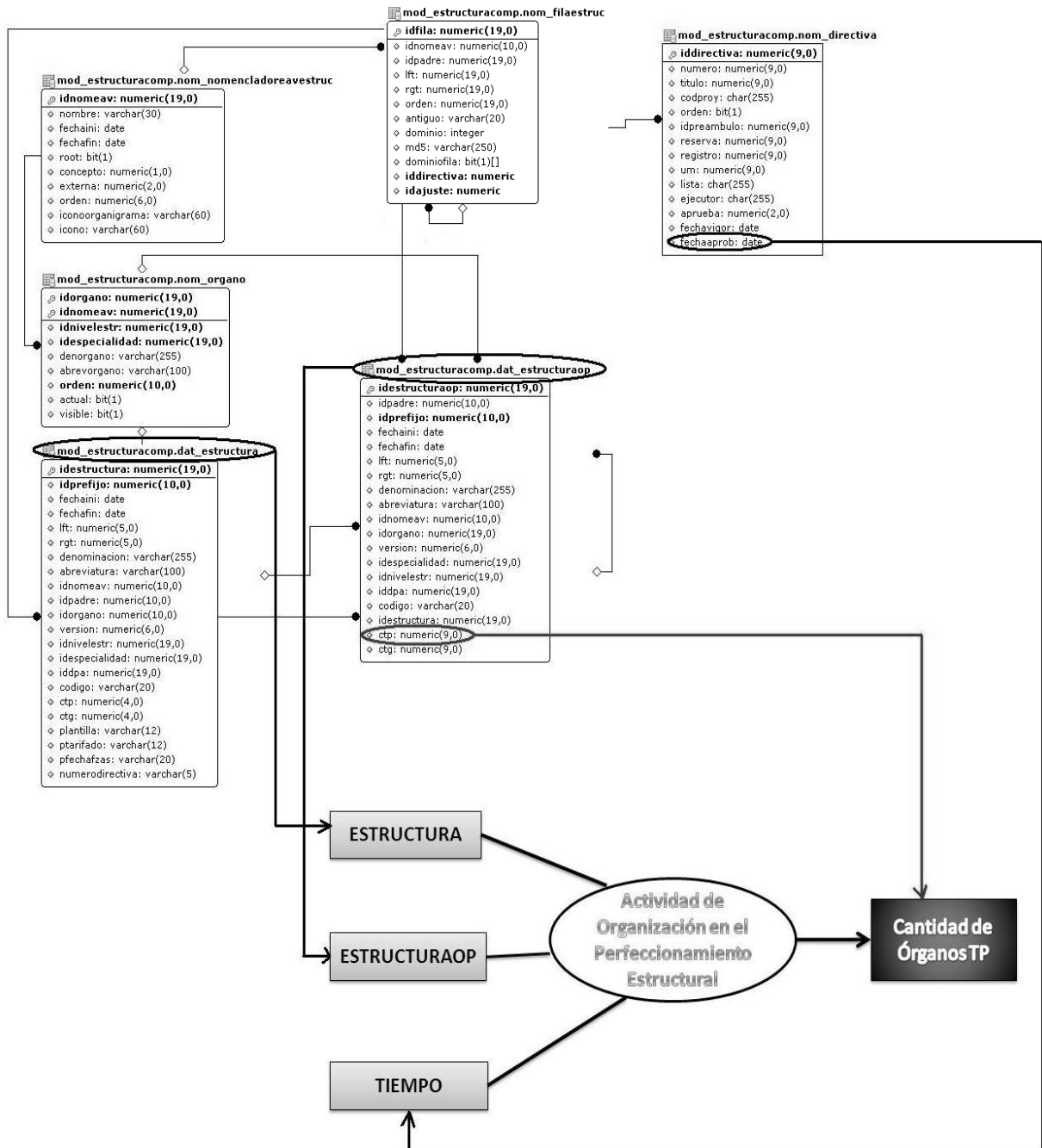


Figura 7: Correspondencia de Requerimientos, Cantidad de Órganos en TP.

Relaciones identificadas:

- La tabla “mod_estructuracomp.dat_estructura” se relaciona con la perspectiva “Estructura”.

CAPÍTULO III. ANÁLISIS Y DISEÑO DEL ALMACÉN DE DATOS

- La tabla “mod_estructuracomp.dat_estructuraop” se relaciona con la perspectiva Estructuraop”.
- El campo “fechaaprob” de la tabla “mod_estructuracomp.nom_directiva” con la perspectiva “Tiempo” (debido a que es la fecha principal del proceso de la actividad de organización en el perfeccionamiento estructural).
- El campo “ctp” de la tabla “mod_estructuracomp.dat_estructuraop” con el indicador “Cantidad de Órganos en TP”.

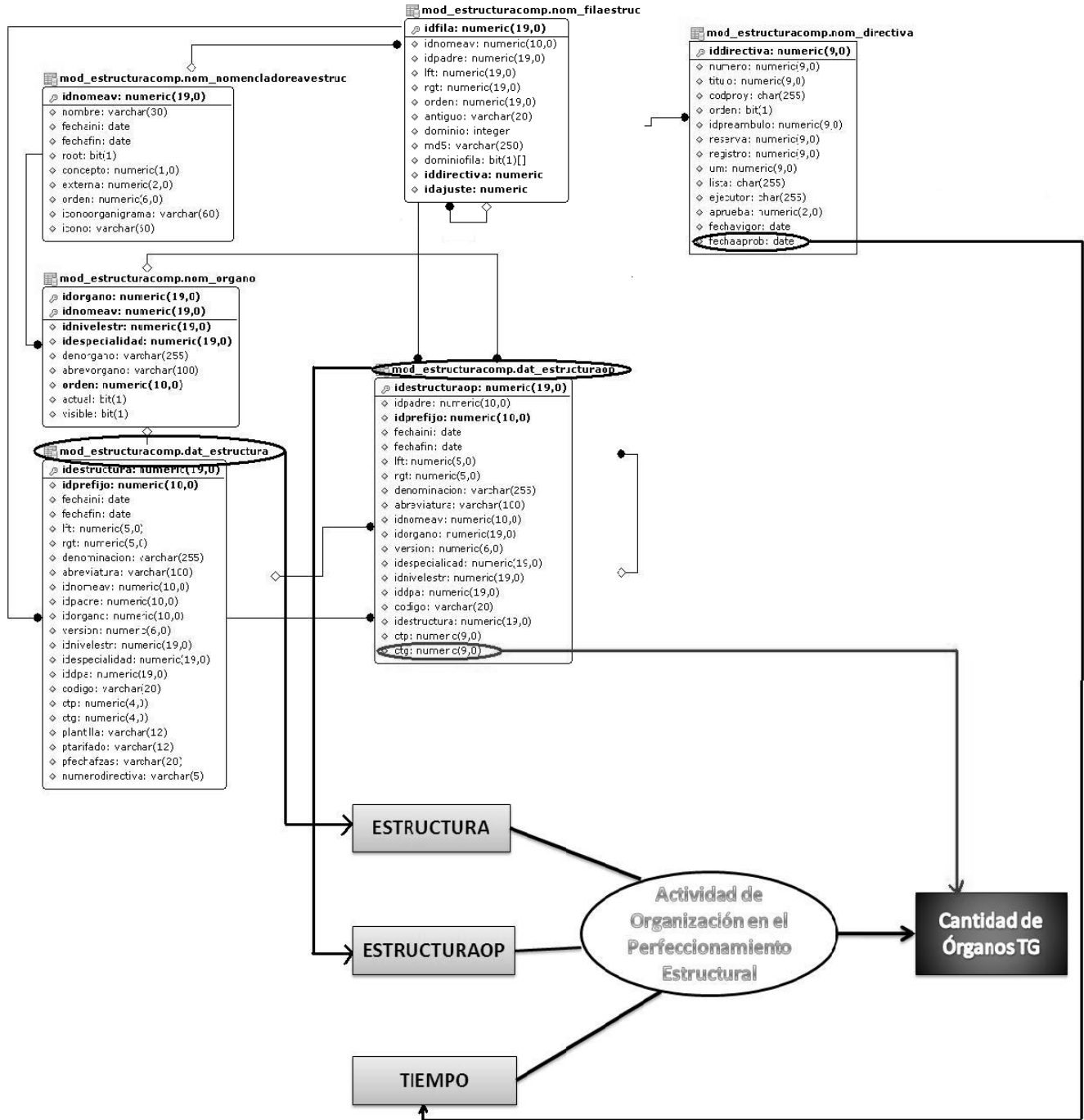


Figura 2: Correspondencia de Requerimientos, Cantidad de Órganos en TG.

Relaciones identificadas:

CAPÍTULO III. ANÁLISIS Y DISEÑO DEL ALMACÉN DE DATOS

- La tabla “mod_estructuracomp.dat_estructura” se relaciona con la perspectiva “Estructura”.
- La tabla “mod_estructuracomp.dat_estructuraop” se relaciona con la perspectiva “Estructuraop”.
- El campo “fechaaprob” de la tabla “mod_estructuracomp.nom_directiva” con la perspectiva “Tiempo” (debido a que es la fecha principal del proceso de la actividad de organización en el perfeccionamiento estructural).
- El campo “ctg” de la tabla “mod_estructuracomp.dat_estructuraop” con el indicador “Cantidad de Órganos en TG”.

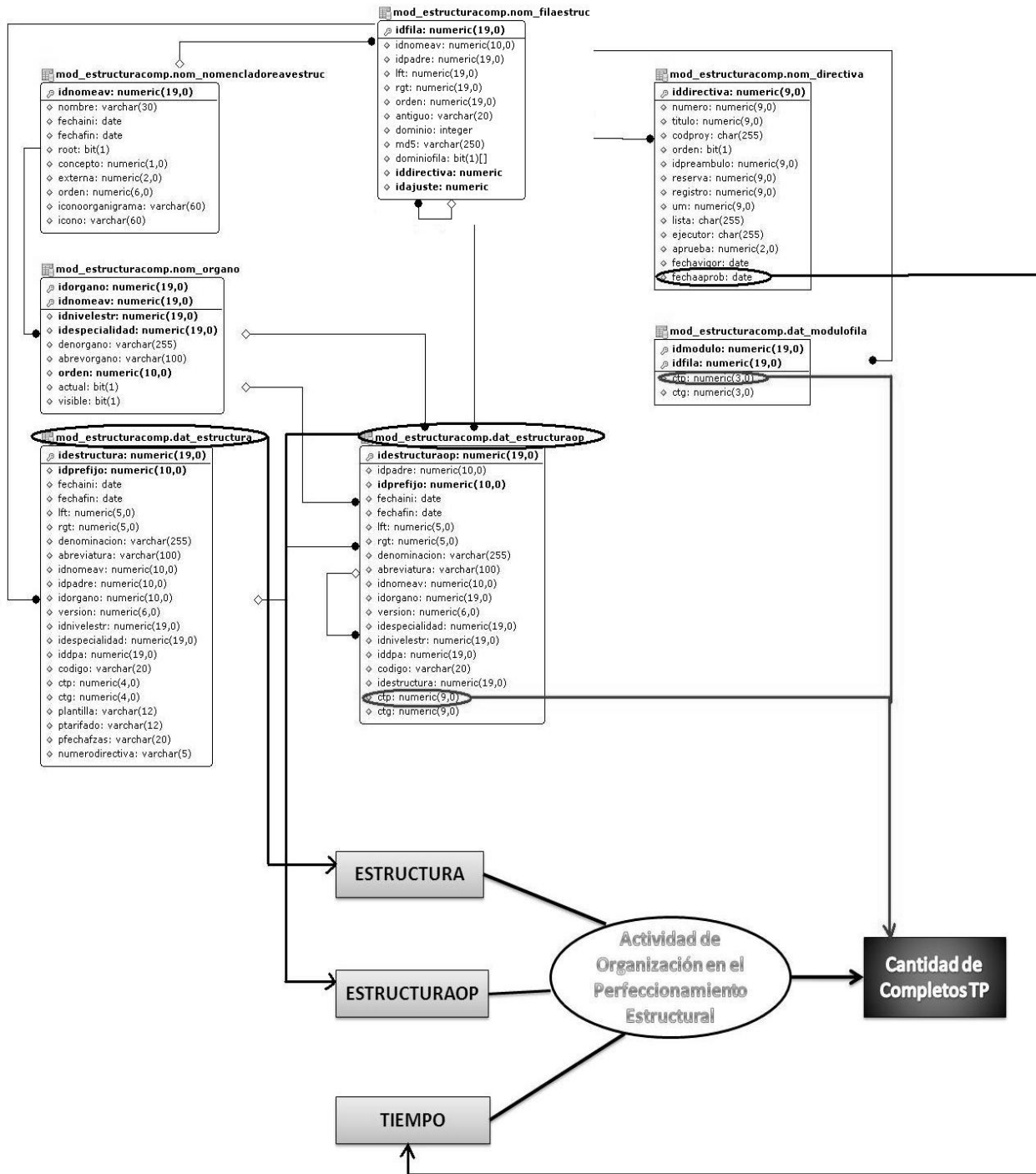


Figura 9: Correspondencia de Requerimientos, Cantidad de Completos en TP.

Relaciones identificadas:

- La tabla “mod_estructuracomp.dat_estructura” se relaciona con la perspectiva “Estructura”.

CAPÍTULO III. ANÁLISIS Y DISEÑO DEL ALMACÉN DE DATOS

- La tabla “mod_estructuracomp.dat_estructuraop” se relaciona con la perspectiva “Estructuraop”.
- El campo “fechaaprob” de la tabla “mod_estructuracomp.nom_directiva” con la perspectiva “Tiempo” (debido a que es la fecha principal del proceso de la actividad de organización en el perfeccionamiento estructural).
- El campo “ctp” de la tabla “mod_estructuracomp.dat_estructuraop” multiplicado por el campo “ctp” de la tabla “mod_estructuracomp.dat_modulofila” con el indicador “Cantidad de Completos en TP”.

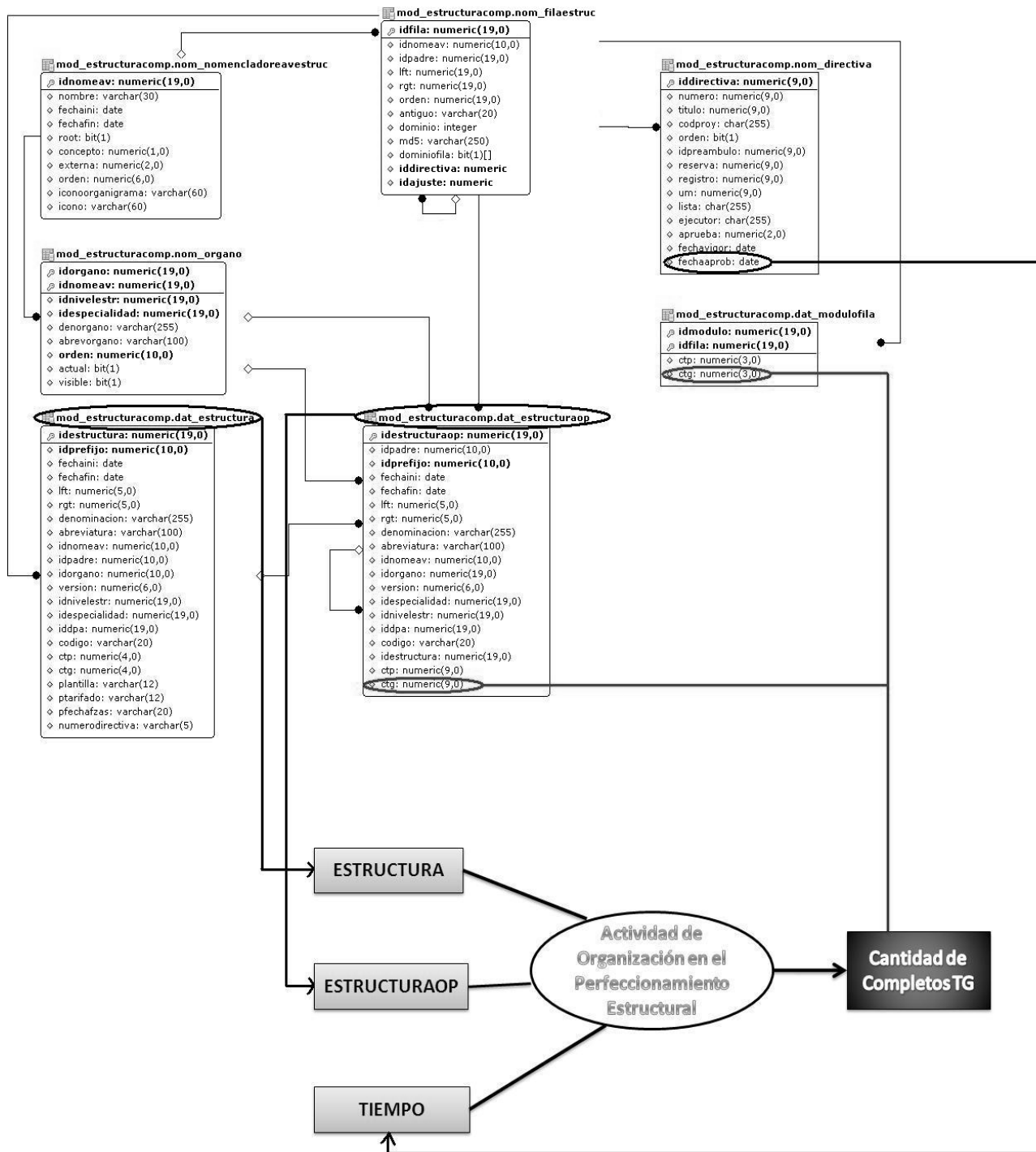


Figura 10: Correspondencia de Requerimientos, Cantidad de Completos en TG.

Relaciones identificadas:

- La tabla “mod_estructuracomp.dat_estructura” se relaciona con la perspectiva “Estructura”.

CAPÍTULO III. ANÁLISIS Y DISEÑO DEL ALMACÉN DE DATOS

- La tabla “mod_estructuracomp.dat_estructuraop” se relaciona con la perspectiva “Estructuraop”.
- El campo “fechaaprob” de la tabla “mod_estructuracomp.nom_directiva” con la perspectiva “Tiempo” (debido a que es la fecha principal del proceso de la actividad de organización en el perfeccionamiento estructural).
- El campo “ctg” de la tabla “mod_estructuracomp.dat_estructuraop” multiplicado por el campo “ctg” de la tabla “mod_estructuracomp.dat_modulofila” con el indicador “Cantidad de Completos en TG”.

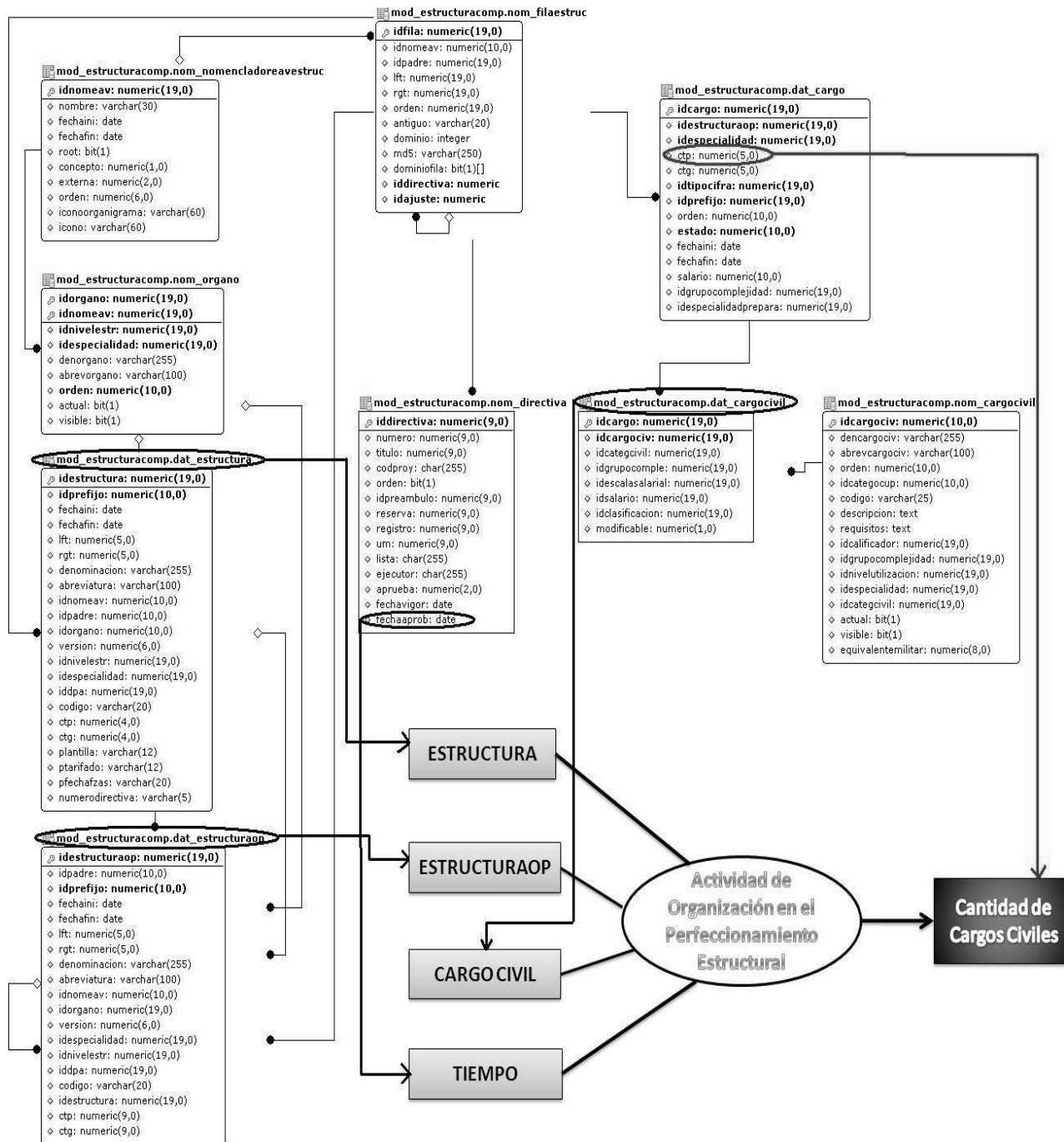


Figura 11: Correspondencia de Requerimientos, Cantidad de Cargos Civiles.

Relaciones identificadas:

- La tabla "mod_estructuracomp.dat_estructura" se relaciona con la perspectiva "Estructura".

CAPÍTULO III. ANÁLISIS Y DISEÑO DEL ALMACÉN DE DATOS

- La tabla “mod_estructuracomp.dat_estructuraop” se relaciona con la perspectiva “Estructuraop”.
- La tabla “mod_estructuracomp.dat_cargocivil” se relacionada con la perspectiva “Cargo Civil”.
- El campo “fechaaprob” de la tabla “mod_estructuracomp.nom_directiva” con la perspectiva “Tiempo” (debido a que es la fecha principal del proceso de la actividad de organización en el perfeccionamiento estructural).
- El campo “ctp” de la tabla “mod_estructuracomp.dat_cargo” con el indicador “Cantidad de Cargos Civiles”.

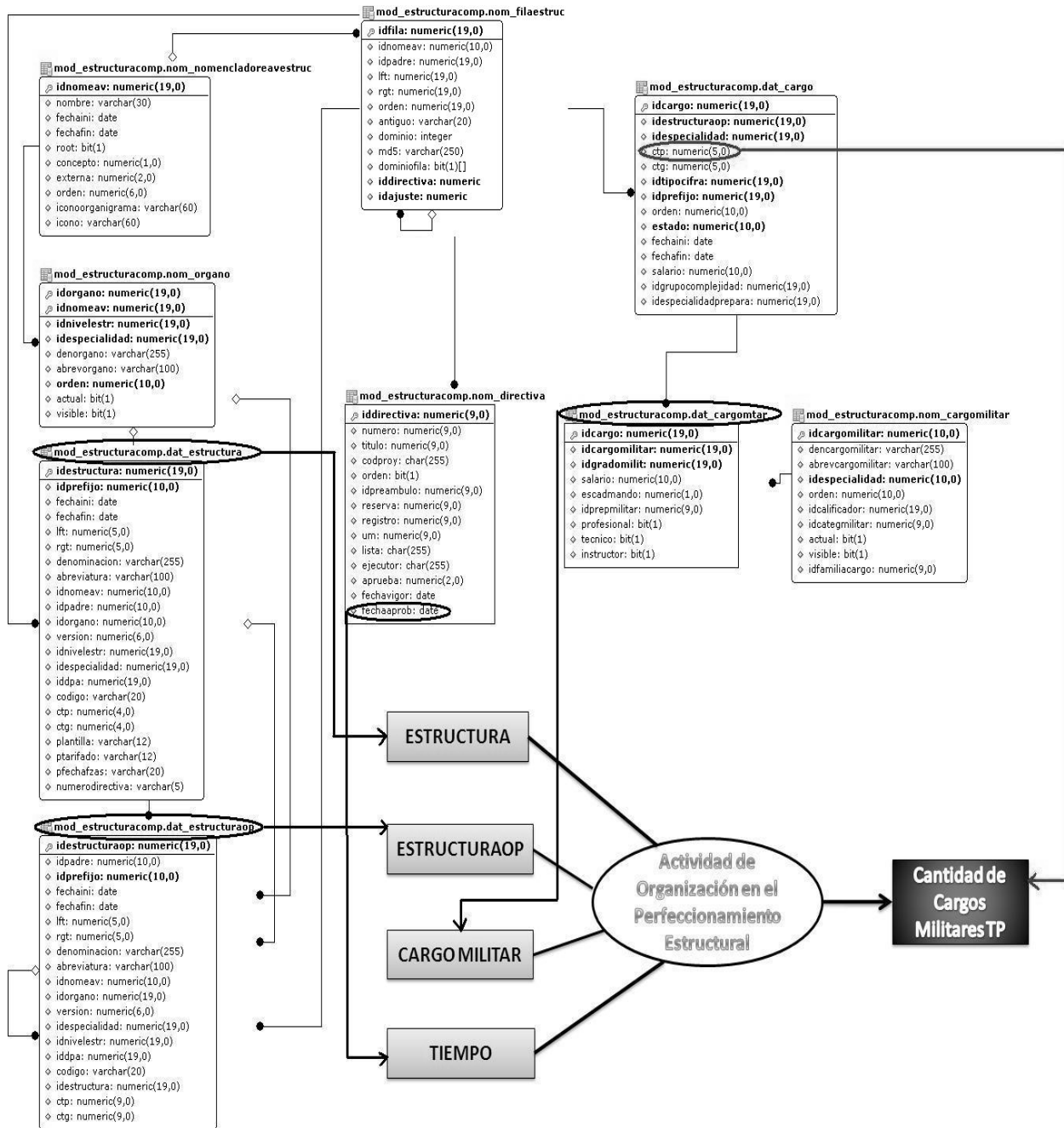


Figura 12: Correspondencia de Requerimientos, Cantidad de Cargos Militares en TP.

Relaciones identificadas:

CAPÍTULO III. ANÁLISIS Y DISEÑO DEL ALMACÉN DE DATOS

- La tabla “mod_estructuracomp.dat_estructura” se relaciona con la perspectiva “Estructura”.
- La tabla “mod_estructuracomp.dat_estructuraop” se relaciona con la perspectiva “Estructuraop”.
- La tabla “mod_estructuracomp.dat_cargomilitar” se relaciona con la perspectiva “Cargo Militar”.
- El campo “fechaaprob” de la tabla “mod_estructuracomp.nom_directiva” con la perspectiva “Tiempo” (debido a que es la fecha principal del proceso de la actividad de organización en el perfeccionamiento estructural).
- El campo “ctp” de la tabla “mod_estructuracomp.dat_cargo” con el indicador “Cantidad de Cargos Militares en TP”.

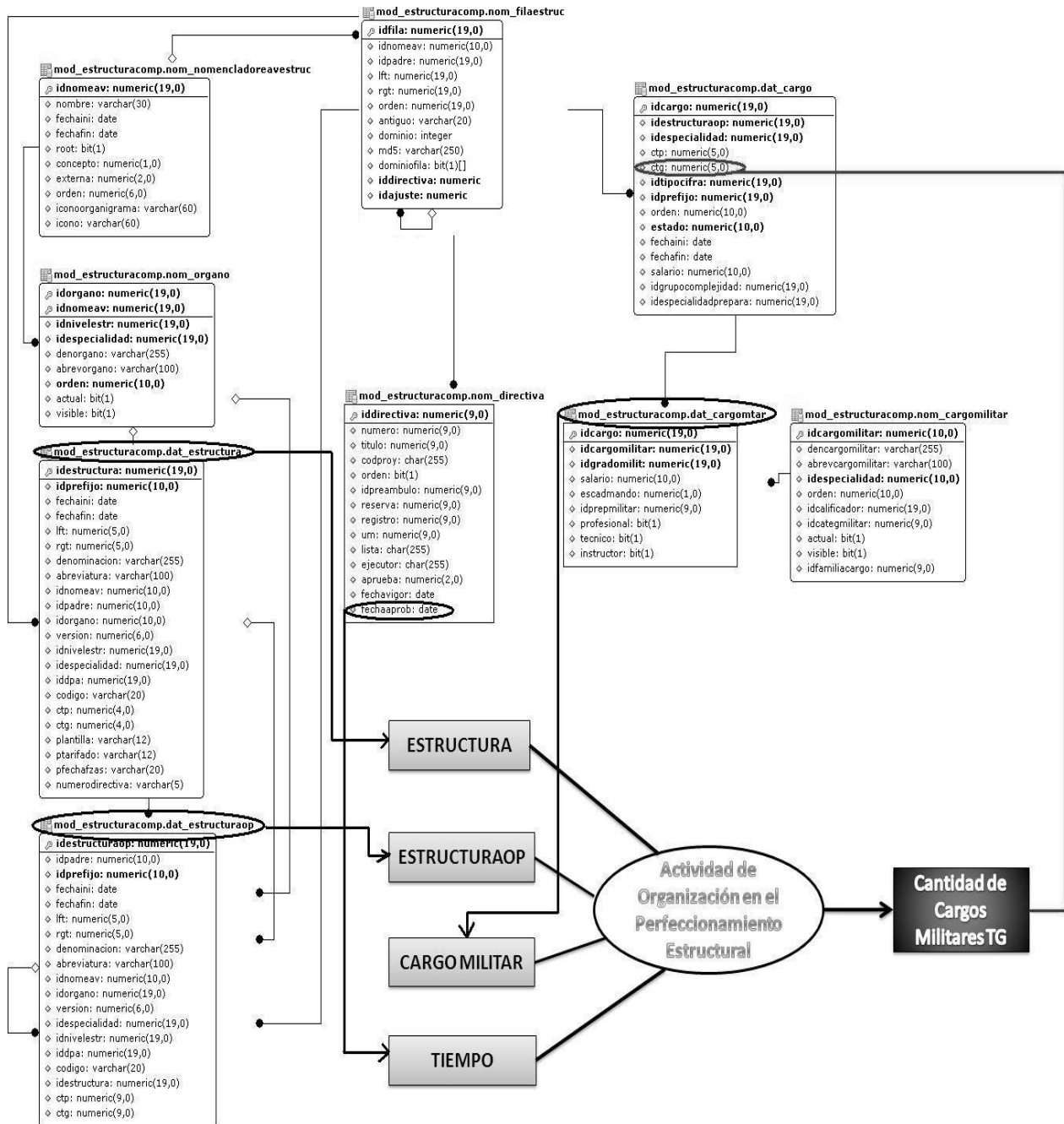


Figura 13: Correspondencia de Requerimientos, Cantidad de Cargos Militares en TG.

Relaciones identificadas:

- La tabla “mod_estructuracomp.dat_estructura” se relaciona con la perspectiva “Estructura”.

CAPÍTULO III. ANÁLISIS Y DISEÑO DEL ALMACÉN DE DATOS

- La tabla “mod_estructuracomp.dat_estructuraop” se relaciona con la perspectiva “Estructuraop”.
- La tabla “mod_estructuracomp.dat_cargomilitar” se relaciona con la perspectiva “Cargo Militar”.
- El campo “fechaaprob” de la tabla “mod_estructuracomp.nom_directiva” con la perspectiva “Tiempo” (debido a que es la fecha principal del proceso de la actividad de organización en el perfeccionamiento estructural).
- El campo “ctg” de la tabla “mod_estructuracomp.dat_cargo” con el indicador “Cantidad de Cargos Militares en TG”.

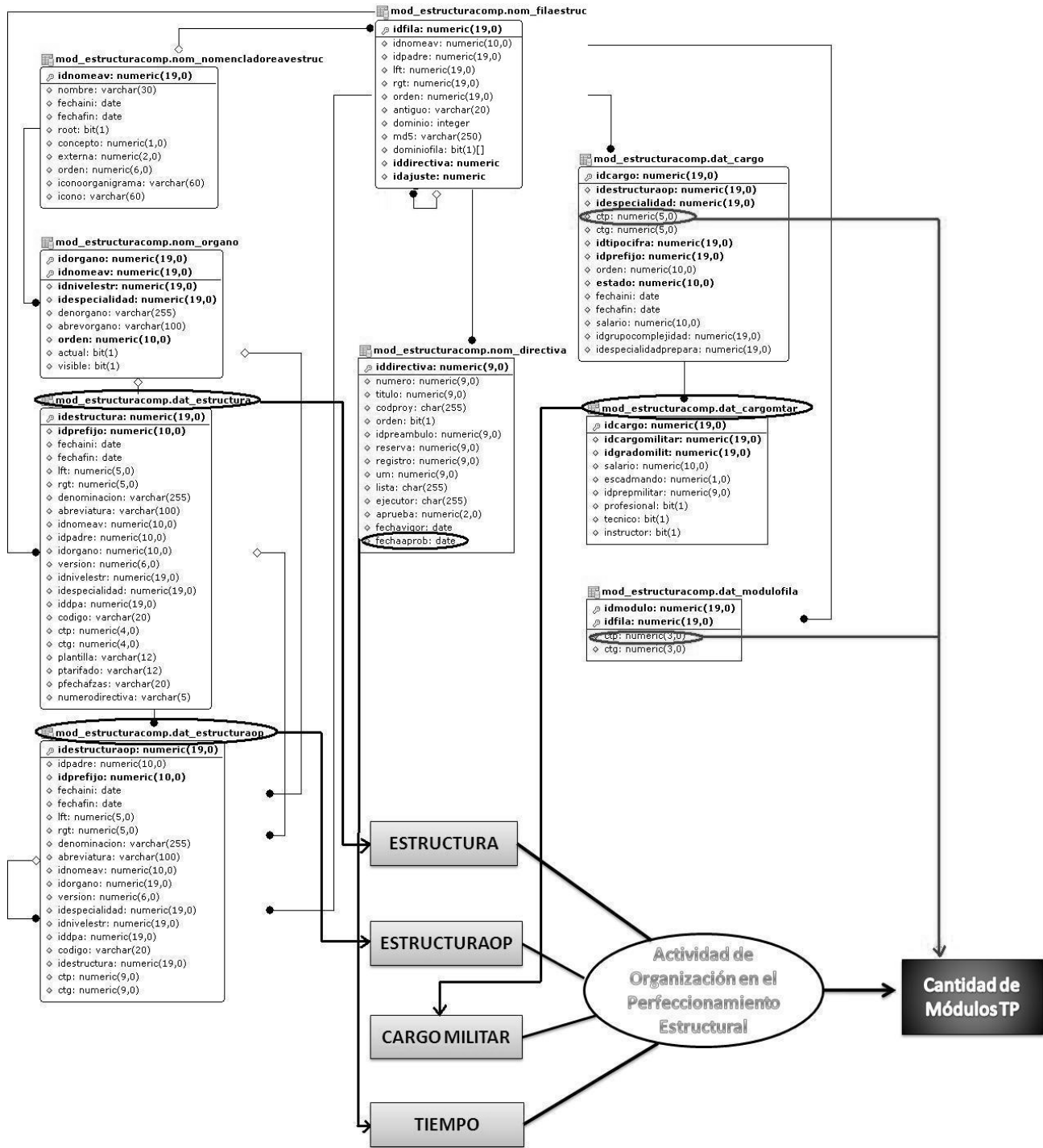


Figura 14: Correspondencia de Requerimientos, Cantidad de Módulos en TP.

Relaciones identificadas:

- La tabla “mod_estructuracomp.dat_estructura” se relaciona con la perspectiva “Estructura”.

CAPÍTULO III. ANÁLISIS Y DISEÑO DEL ALMACÉN DE DATOS

- La tabla “mod_estructuracomp.dat_estructuraop” se relaciona con la perspectiva “Estructuraop”.
- La tabla “mod_estructuracomp.dat_cargomilitar” se relaciona con la perspectiva “Cargo Militar”.
- El campo “fechaaprob” de la tabla “mod_estructuracomp.nom_directiva” con la perspectiva “Tiempo” (debido a que es la fecha principal del proceso de la actividad de organización en el perfeccionamiento estructural).
- El campo “ctp” de la tabla “mod_estructuracomp.dat_cargo” multiplicado con el campo “ctp” de la tabla “mod_estructuracomp.dat_modulofila” con el indicador “Cantidad de Módulos en TP”.

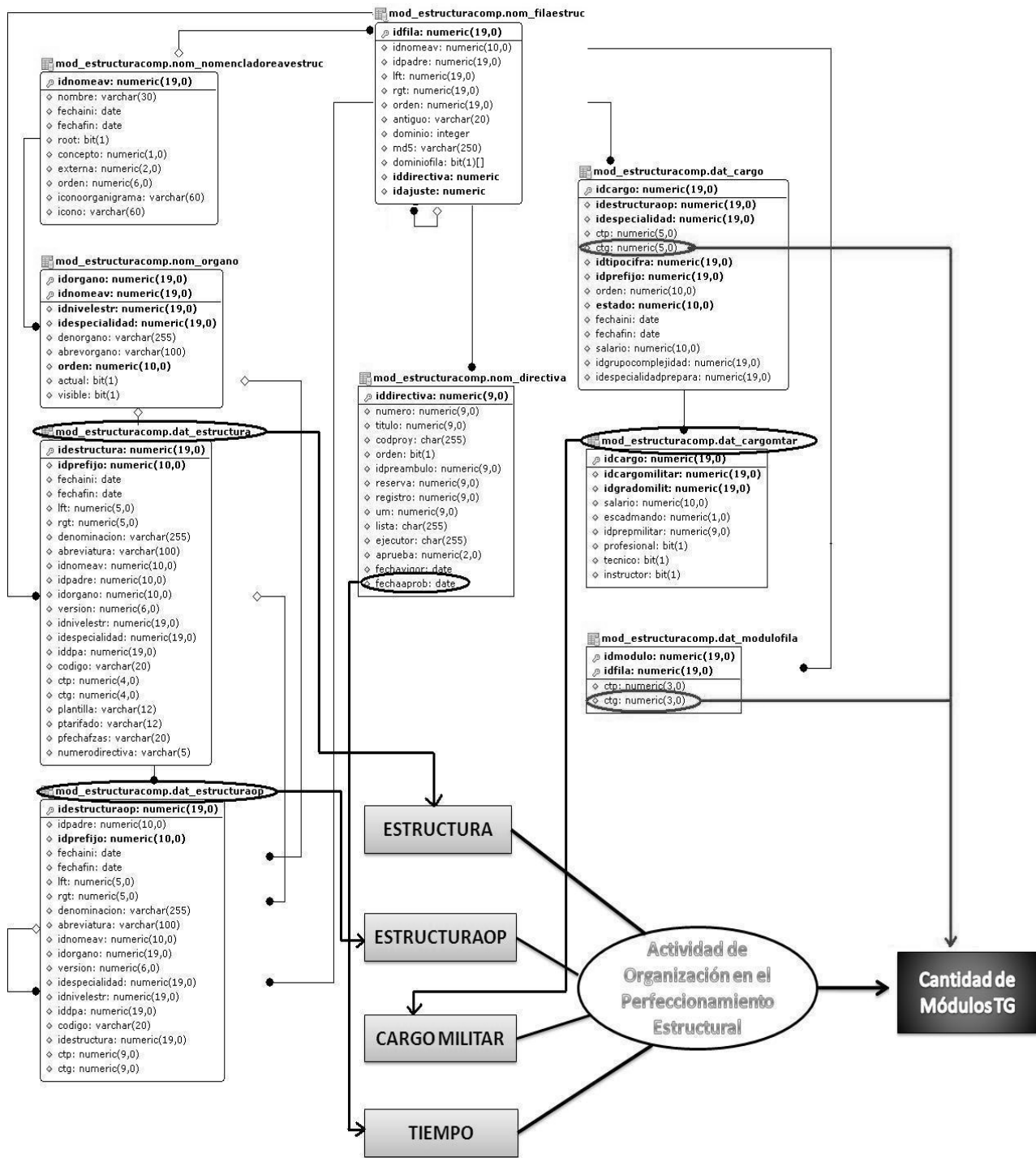


Figura 15: Correspondencia de Requerimientos, Cantidad de Módulos en TG.

Relaciones identificadas:

- La tabla “mod_estructuracomp.dat_estructura” se relaciona con la perspectiva “Estructura”.

CAPÍTULO III. ANÁLISIS Y DISEÑO DEL ALMACÉN DE DATOS

- La tabla “mod_estructuracomp.dat_estructuraop” se relaciona con la perspectiva “Estructuraop”.
- La tabla “mod_estructuracomp.dat_cargomilitar” se relaciona con la perspectiva “Cargo Militar”.
- El campo “fechaaprob” de la tabla “mod_estructuracomp.nom_directiva” con la perspectiva “Tiempo” (debido a que es la fecha principal del proceso de la actividad de organización en el perfeccionamiento estructural).
- El campo “ctg” de la tabla “mod_estructuracomp.dat_cargo” multiplicado con el campo “ctp” de la tabla “mod_estructuracomp.dat_modulofila” con el indicador “Cantidad de Módulos en TG”.

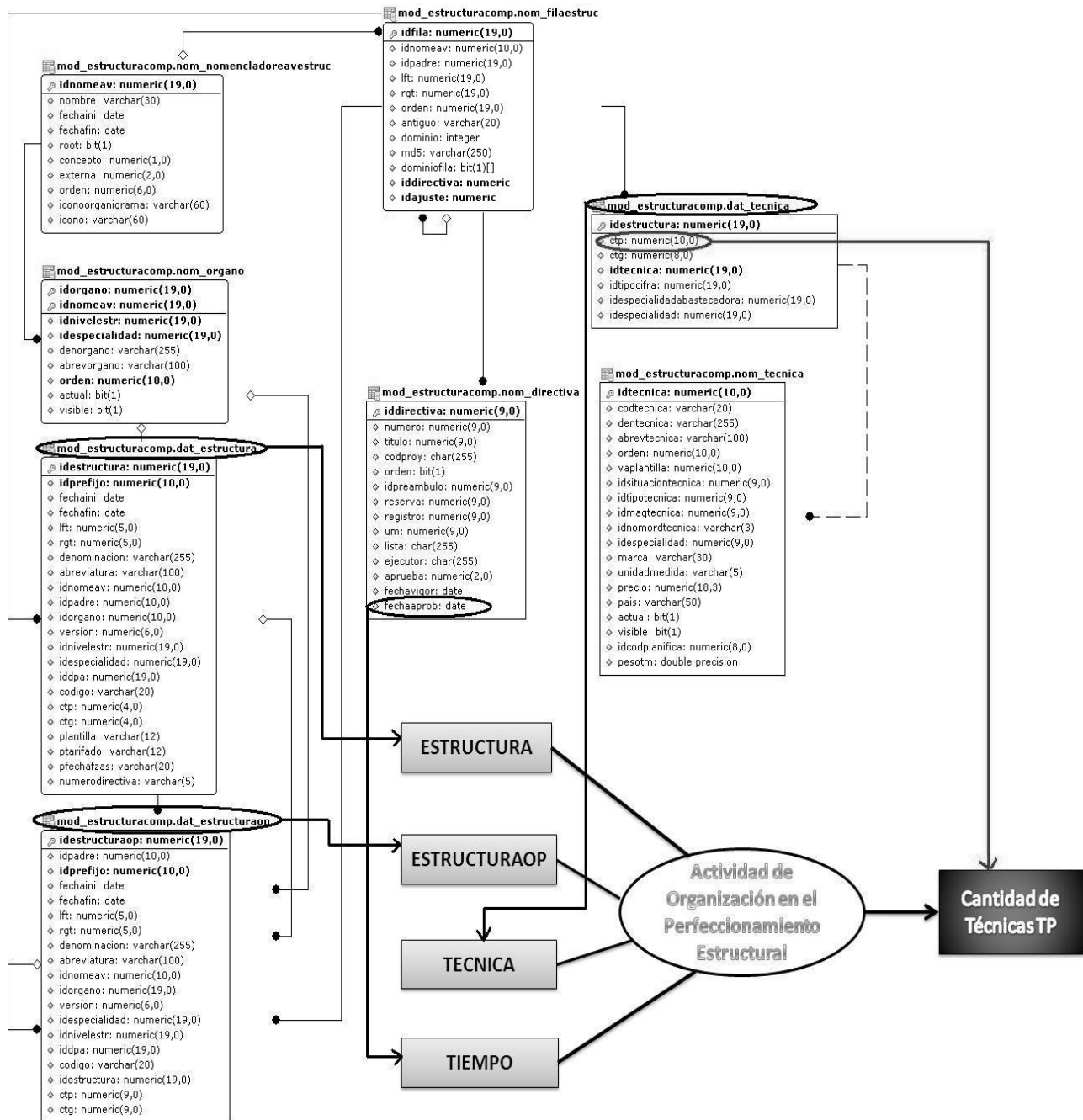


Figura 16: Correspondencia de Requerimientos, Cantidad de Técnicas en TP.

Relaciones identificadas:

- La tabla “mod_estructuracomp.dat_estructura” se relaciona con la perspectiva “Estructura”.

CAPÍTULO III. ANÁLISIS Y DISEÑO DEL ALMACÉN DE DATOS

- La tabla “mod_estructuracomp.dat_estructuraop” se relaciona con la perspectiva “Estructuraop”.
- La tabla “mod_estructuracomp.dat_tecnica” se relaciona con la perspectiva “Técnica”.
- El campo “fechaaprob” de la tabla “mod_estructuracomp.nom_directiva” con la perspectiva “Tiempo” (debido a que es la fecha principal del proceso de la actividad de organización en el perfeccionamiento estructural).
- El campo “ctp” de la tabla “mod_estructuracomp.dat_tecnica” con el indicador “Cantidad de Técnicas en TP”.

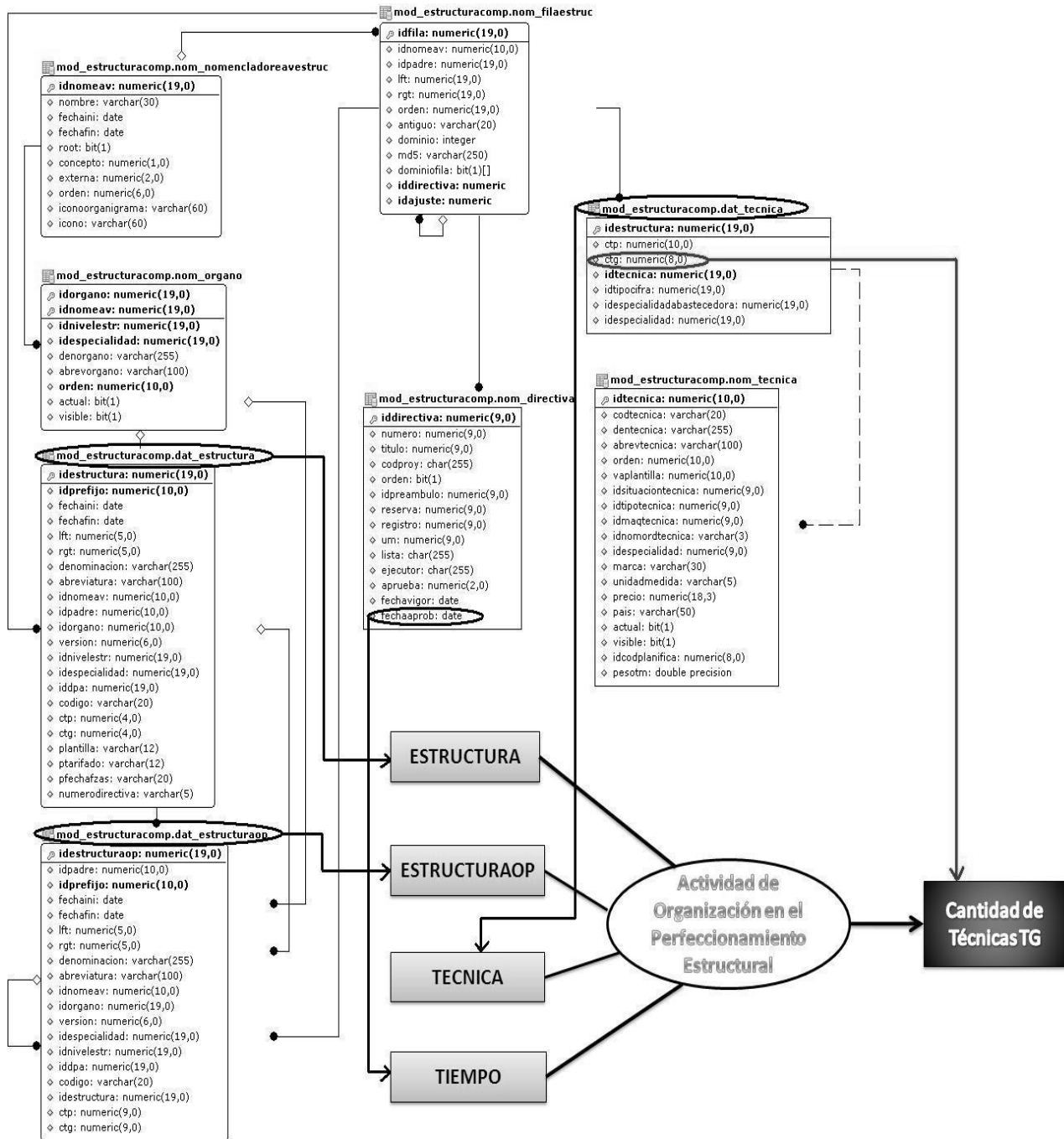


Figura 17: Correspondencia de Requerimientos, Cantidad de Técnicas en TG.

Relaciones identificadas:

CAPÍTULO III. ANÁLISIS Y DISEÑO DEL ALMACÉN DE DATOS

- La tabla “mod_estructuracomp.dat_estructura” se relaciona con la perspectiva “Estructura”.
- La tabla “mod_estructuracomp.dat_estructuraop” se relaciona con la perspectiva “Estructuraop”.
- La tabla “mod_estructuracomp.dat_tecnica” se relaciona con la perspectiva “Técnica”.
- El campo “fechaaprob” de la tabla “mod_estructuracomp.nom_directiva” con la perspectiva “Tiempo” (debido a que es la fecha principal del proceso de la actividad de organización en el perfeccionamiento estructural).
- El campo “ctg” de la tabla “mod_estructuracomp.dat_tecnica” con el indicador “Cantidad de Técnicas en TG”.

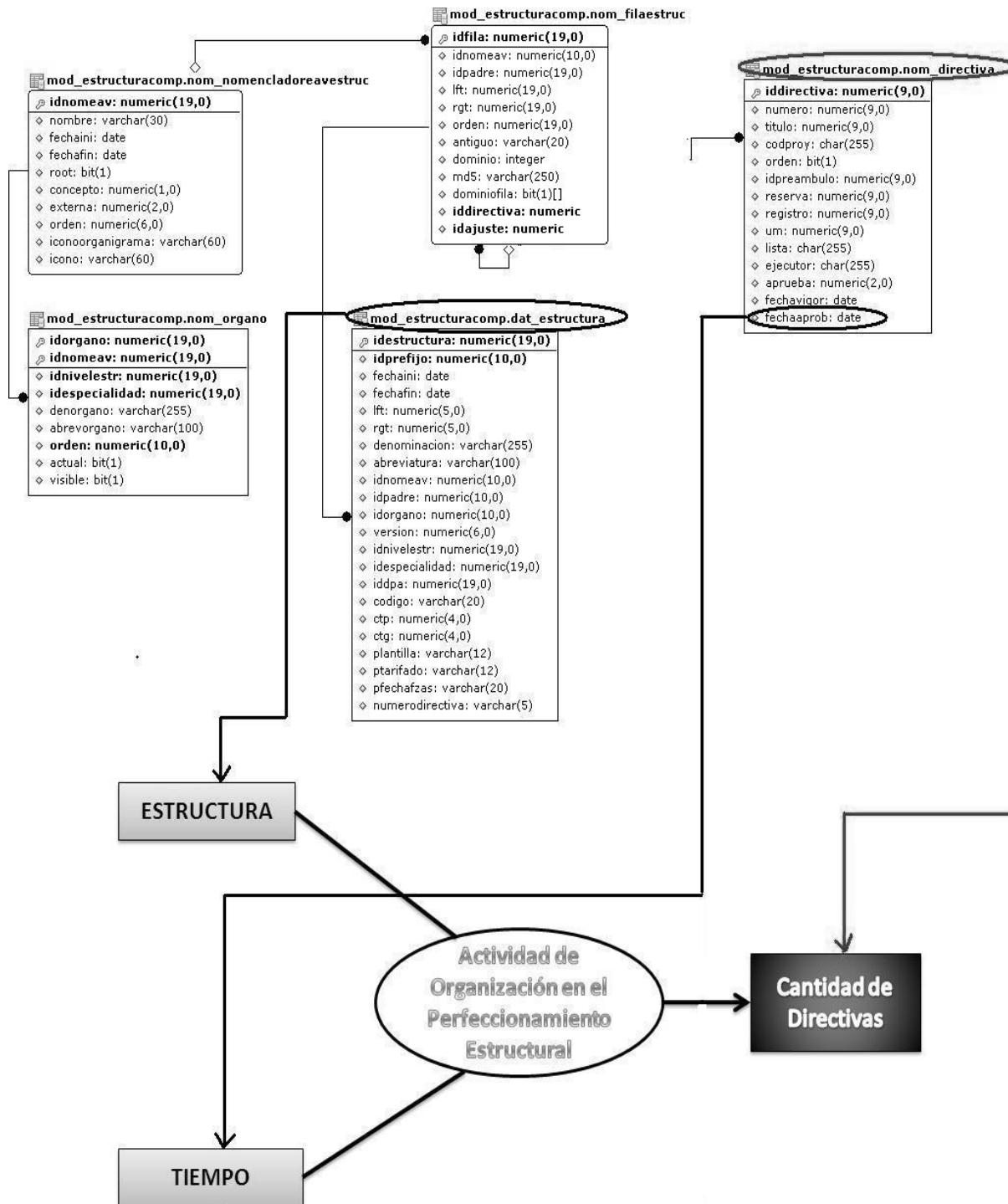


Figura 13: Correspondencia de Requerimientos, Cantidad de Directivas.

Relaciones identificadas:

- La tabla “mod_estructuracomp.dat_estructura” se relaciona con la perspectiva “Estructura”.

CAPÍTULO III. ANÁLISIS Y DISEÑO DEL ALMACÉN DE DATOS

- El campo “fechaaprob” de la tabla “mod_estructuracomp.nom_directiva” con la perspectiva “Tiempo” (debido a que es la fecha principal del proceso de la actividad de organización en el perfeccionamiento estructural).
- La tabla “mod_estructuracomp.nom_directiva” con el indicador “Cantidad de Directivas”.

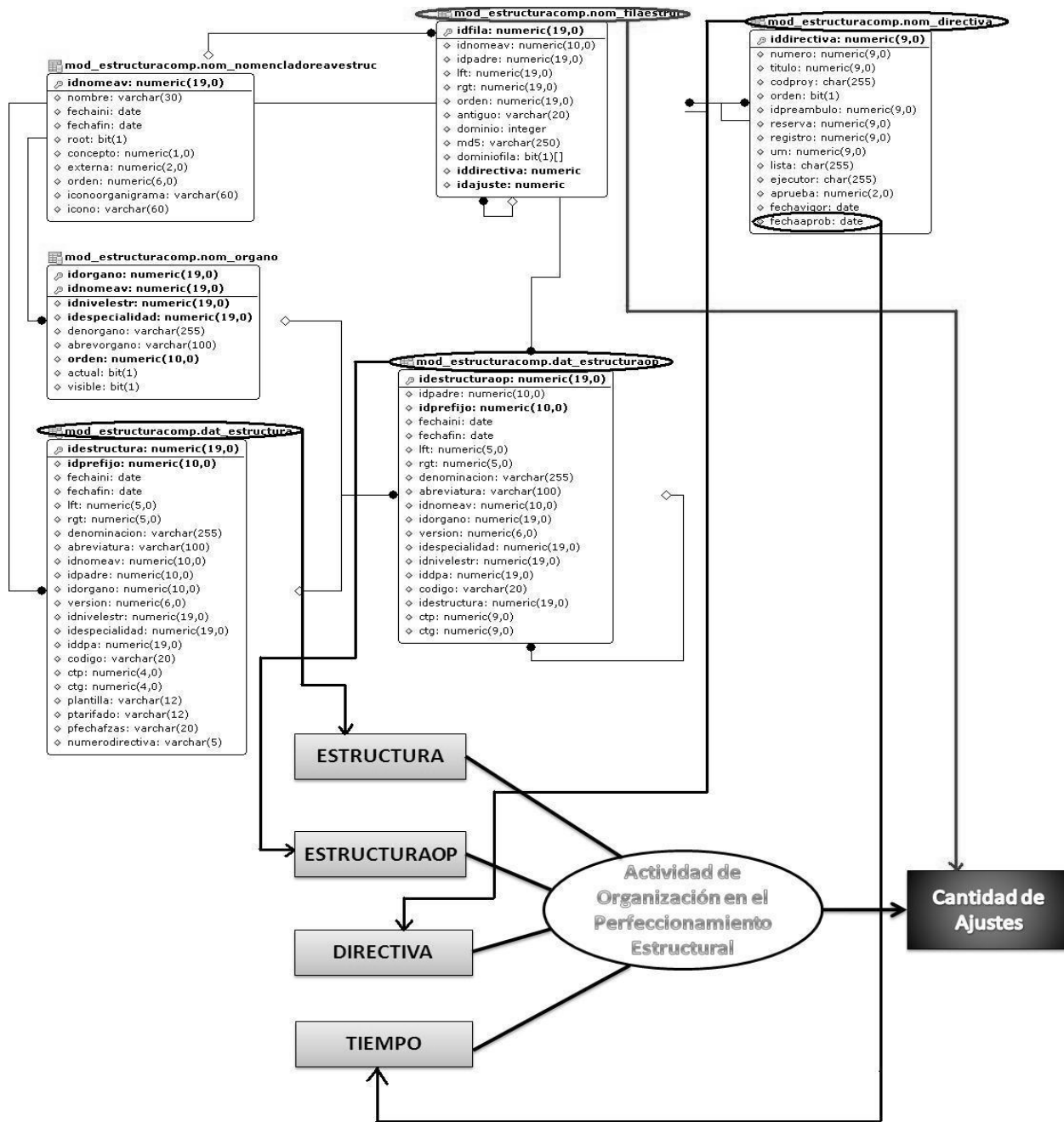


Figura 19: Correspondencia de Requerimientos, Cantidad de Ajustes.

Relaciones identificadas:

- La tabla “mod_estructuracomp.dat_estructura” se relaciona con la perspectiva “Estructura”.
- La tabla “mod_estructuracomp.nom_directiva” se relaciona con la perspectiva “Directiva”.

- El campo “fechaaprob” de la tabla “mod_estructuracomp.nom_directiva” con la perspectiva “Tiempo” (debido a que es la fecha principal del proceso de la actividad de organización en el perfeccionamiento estructural).
- La tabla “mod_estructuracomp.nom_filaestruc” con el indicador “Cantidad de Ajustes”.

3.3 Selección de los campos que integrarán cada perspectiva. Nivel de granularidad.

Luego de exponer frente al usuario los datos existentes, explicando su significado, valores posibles y características, este debe decidir cuáles son los que considera relevantes para consultar los indicadores y cuáles no. Con respecto a la perspectiva “Tiempo”, es muy importante definir el ámbito mediante el cual se agruparán o sumarán los datos. Este punto es fundamental precisarlo con claridad, debido a que, determinará la granularidad de la información encontrada en el DW. Sus campos posibles pueden ser: día de la semana, quincena, mes, trimestre, semestre, año, etc.

De acuerdo con las correspondencias establecidas, se analizaron los campos residentes en cada tabla a la que se hacía referencia, a través de dos métodos diferentes. Primero se examinó la base de datos para intuir los significados de cada campo, y luego se consultó con el encargado del sistema sobre algunos aspectos de los cuales no se comprendía su sentido. Como puede apreciarse en el diagrama de entidad relación antes expuesto, los nombres de los campos son bastante explícitos y se deducen con facilidad, sin embargo, fue necesario investigarlos para evitar cualquier tipo de inconvenientes.

- Con respecto a la perspectiva “**Estructura**”, los datos disponibles son los siguientes:
 - idestructura: Llave primaria de la tabla ,guarda el identificador de la estructura
 - idprefijo: Representa a través de una llave foránea el mando donde se aprobó la estructura.
 - fechaini: Fecha de inicio del uso de la estructura
 - fechafin: Fecha de fin del uso de la estructura
 - rgt: Hijo izquierdo de la estructura
 - denominación: Alias de la estructura
 - abreviatura: Abreviatura de la estructura
 - idnomeav: Representa a través de una llave foránea el tipo de estructura. Ejemplo: Órgano, Plantilla, Agrupación.

- idpadre: Guarda el id del padre de la estructura.
 - idorgano: Representa a través de una llave foránea la denominación de la estructura. Ejemplo: Región Naval, Campamento Oriental.
 - idnivelestr: Representa a través de una llave foránea el nivel estructural de la estructura.
 - idespecialidad: Representa a través de una llave foránea la especialidad.
 - iddpa: Representa a través de una llave foránea la división político administrativa a la que pertenece la estructura.
 - código: Guarda el código de la estructura
 - ctp: Cantidad en tiempo de Paz
 - ctg: Cantidad en tiempo de Guerra
 - fechavigor: Fecha en que se pone en vigor
 - iddirectiva: Identificador de nom_directiva
 - plantilla: Numero de la plantilla
 - idclasifunidades: Llave foránea de nom_clasifunidades Representa a través de una llave foránea la clasificación de unidades.
- Con respecto a la perspectiva “**Estructuraop**”, los datos disponibles son los siguientes:
- idestructuraop: Llave primaria de la tabla ,guarda el identificador de la estructura
 - idpadre: Guarda el id del padre de la estructuraop
 - idprefijo: Llave foránea de la tabla nom_prefijo , guarda los id.
 - fechaini: Fecha de inicio del uso de la estructuraop
 - fechafin: Fecha de fin del uso de la estructuraop
 - lift: Hijo izquierdo de la estructuraop
 - rgt: Hijo derecho de la estructuraop
 - denominación: Denominación de la estructuraop
 - abreviatura: Abreviatura de la estructura
 - idnomeav: Atributo para evitar la redundancia. La tabla al que pertenece no está relacionada, la tabla nomencladorEAV
 - idorgano: Llave foránea de la tabla nom_organo , guarda los id

CAPÍTULO III. ANÁLISIS Y DISEÑO DEL ALMACÉN DE DATOS

- idnivelestr: Guarda el nivel estructural de la estructura, llave foránea de la tabla nom_nivelestruct , guarda los id
 - idespecialidad: Llave foránea de la tabla especialidad , guarda los id
 - iddpa: Llave foránea de la tabla nom_dpa , guarda los id
 - código: Guarda el código de la estructuraop
 - ctp: Cantidad en tiempo de Paz
 - ctg: Cantidad en tiempo de Guerra
- Con respecto a la perspectiva “**Cargo Civil**”, los datos disponibles son los siguientes:
- Idcargocivil: Llave foránea de la tabla dat_cargo , guarda los id.
 - Idcargocivil: Guarda el identificador de la tabla nom_cargocivil
 - Idcategcivil: Guarda el identificador de la tabla nom_categcivil
 - idgrupocomple: Guarda el identificador de la tabla nom_grupocomple
 - idclasificacion: Llave foránea de la tabla nom_clasificacion, guarda los id
- Con respecto a la perspectiva “**Cargo Militar**”, los datos disponibles son los siguientes:
- Idcargomilitar: Guarda el identificador del cargo militar
 - Idgradomilit: Guarda el identificador del grado militar
 - escadmando: Guarda el mando escalar del cargo militar
 - idcargo: Guarda el identificador del cargo
 - idprepmilitar: Guarda el identificador de prepmilitar
 - profesional: Si es profesional o no
 - técnico: Si es técnico o no
 - instructor: Si es instructor o no
- Con respecto a la perspectiva “**Técnica**”, los datos disponibles son los siguientes:
- idestructura: Llave foránea de la tabla dat_estructuraop, guarda los id
 - ctp: Cantidad en tiempo de Paz

- ctg: Cantidad en tiempo de Guerra
 - idtecnica: Llave primaria del nomenclador, se guarda el identificador de la tabla
 - idtipocifra: Llave foránea de la tabla nom_ tipocifra, guarda los id
 - idespecialidadabastecedora: Identificador de la especialidad abastecedora
 - idespecialidad: Llave foránea de la tabla nom_especialidad, guarda los id
 - codproy: Identificador de nom_proyecto
- Con respecto a la perspectiva “**Directiva**”, los datos disponibles son los siguientes:
- iddirectiva: Identificación de la directiva
 - numero: numero de la directiva
 - titulo: Titulo de la directiva
 - codproy: Identificador de proyecto
 - orden: Orden de la directiva
 - idpreambulo: Llave foránea de nom_preambulo
 - reserva: Clasificación de la directiva
 - registro: Registro de la directiva (machón)
 - um: Unidad militar de la directiva (machón)
 - lista: Lista de la directiva (machón)
 - ejecutor: Registro de la directiva (machón)
 - aprueba: Llave foránea de nom_firmas. Guarda el identificador del jefe que aprueba
 - fechavigor: Fecha en que se pone en vigor la directiva
 - fechaaprob: Fecha en que se aprueba la directiva
- Con respecto a la perspectiva “**Tiempo**”, que es la que determinará la granularidad del depósito de datos, los atributos más típicos que pueden emplearse son los siguientes:
- Año.
 - Semestre.
 - Cuatrimestre.
 - Trimestre.

- Número de mes.
- Nombre del mes.
- Quincena.
- Semana.
- Número de día.

Una vez que se recolectó toda la información pertinente y se consultó con los usuarios cuales eran los datos que consideraban de interés para analizar los indicadores ya expuestos, los resultados obtenidos fueron los siguientes:

En la perspectiva “Estructura” se utilizarán los campos que hacen referencia al código (“codigo” de la tabla “mod_estructuracomp.dat_estructura”), a la denominación (“denominacion” de la tabla “mod_estructuracomp.dat_estructura”), a la especialidad (“denespecialidad” de la tabla “mod_nomencladores.nom_especialidad”), al nivel estructural (“dennivelestr” de la tabla “mod_estructuracomp.nom_nivelestr”), al número de la plantilla (“plantilla” de la tabla “mod_estructuracomp.dat_estructura”), al completo (“esmodulo” de la tabla “mod_estructuracomp.nom_modulo”), y al tipo de ajuste (“abreviatura” de la tabla “mod_estructuracomp.nom_tipoajuste”).

En la perspectiva “Estructuraop” se utilizarán los campos que hacen referencia al código (“codigo” de la tabla “mod_estructuracomp.dat_estructuraop”), la denominación (“denominacion” de la tabla “mod_estructuracomp.dat_estructuraop”), la especialidad (“denespecialidad” de la tabla “mod_nomencladores.nom_especialidad”), al nivel estructural (“dennivelestr” de la tabla “mod_estructuracomp.nom_nivelestr”), al completo (“esmodulo” de la tabla “mod_estructuracomp.nom_modulo”), y al tipo de ajuste (“abreviatura” de la tabla “mod_estructuracomp.nom_tipoajuste”).

En la perspectiva “CargoCivil” se utilizará los campos que hacen referencia al código (“codigo” de la tabla “mod_estructuracomp.nom_cargocivil”), la denominación (“dencargociv” de la tabla “mod_estructuracomp.nom_cargocivil”), la categoría funcional (“dencategcivil” de la tabla “mod_estructuracomp.nom_categcivil”), el tipo de cargo (“dentipocifra” de la tabla “mod_estructuracomp.nom_tipocifra”), la función del cargo (“denominacion” de la tabla “mod_estructuracomp.nom_funciones”), la especialidad del cargo (“denespecialidad”, obtenido a través de

CAPÍTULO III. ANÁLISIS Y DISEÑO DEL ALMACÉN DE DATOS

la unión con la tabla “mod_nomencladores.nom_especialidad”) y al tipo de ajuste (“abreviatura” de la tabla “mod_estructuracomp.nom_tipoajuste”).

En la perspectiva “Cargo Militar” se utilizarán los campos que hacen referencia al código (“codigo” de la tabla “mod_estructuracomp.nom_cargomilitar”), a la especialidad de preparación “denespecialidad” de la tabla “mod_nomencladores.nom_especialidad”), el grado (“dengradomilit” de la tabla “mod_estructuracomp.nom_gradomilit”), la denominación (“dencargomilitar” de la tabla “mod_estructuracomp.dat_cargomilitar”), el tipo de cargo (“dentipocifra” de la tabla “mod_estructuracomp.nom_tipocifra”), la función del cargo (“denominacion” de la tabla “mod_estructuracomp.nom_funciones”), la especialidad del cargo (“denespecialidad” de la tabla “mod_nomencladores.nom_especialidad”), al nivel de preparación (“denprepmilitar” de la tabla “mod_estructuracom.nom_prepmilitar”), al módulo (“esmodulo” de la tabla “mod_estructuracomp.nom_modulo”) y al tipo de ajuste (“abreviatura” de la tabla “mod_estructuracomp.nom_tipoajuste”).

En la perspectiva “Técnica” se utilizarán los campos que hacen referencias a la especialidad del medio (“denespecialidad” de la tabla “mod_nomencladores.nom_especialidad”), a la especialidad abastecedora (“denespecialidad” de la tabla “mod_nomencladores.nom_especialidad”), al código (“codtecnica” de la tabla “mod_estructuracomp.nom_tecnica”), a la denominación (“dentecnica” de la tabla “mod_estructuracomp.nom_tecnica”) y al tipo de ajuste (“abreviatura” de la tabla “mod_estructuracomp.nom_tipoajuste”).

En la perspectiva “Directiva” se utilizarán los campos que hacen referencias al número de la directiva (“numero” de la tabla “mod_estructuracomp.nom_directiva”), al jefe que la aprueba (“aprueba” de la tabla “mod_estructuracomp.nom_directiva”), a la fecha de aprobación (“fechaaprob” de la tabla “mod_estructuracomp.nom_directiva”), a la fecha de puesta en vigor (“fechavigor” de la tabla “mod_estructuracomp.nom_directiva”), al título de la directiva (“titulo” de la tabla “mod_estructuracomp.nom_directiva”), a la clasificación del documento (“reserva” de la tabla “mod_estructuracomp.nom_directiva”), y al registro del documento (“registro” de la tabla “mod_estructuracomp.nom_directiva”).

En la perspectiva “Tiempo”, se utilizará el campo que hace referencia a la fecha en que se aprobó la directiva (“fechaaprob” de la tabla “mod_estructuracomp.nom_directiva”).

Teniendo esto en cuenta, se completará el diseño del diagrama conceptual:

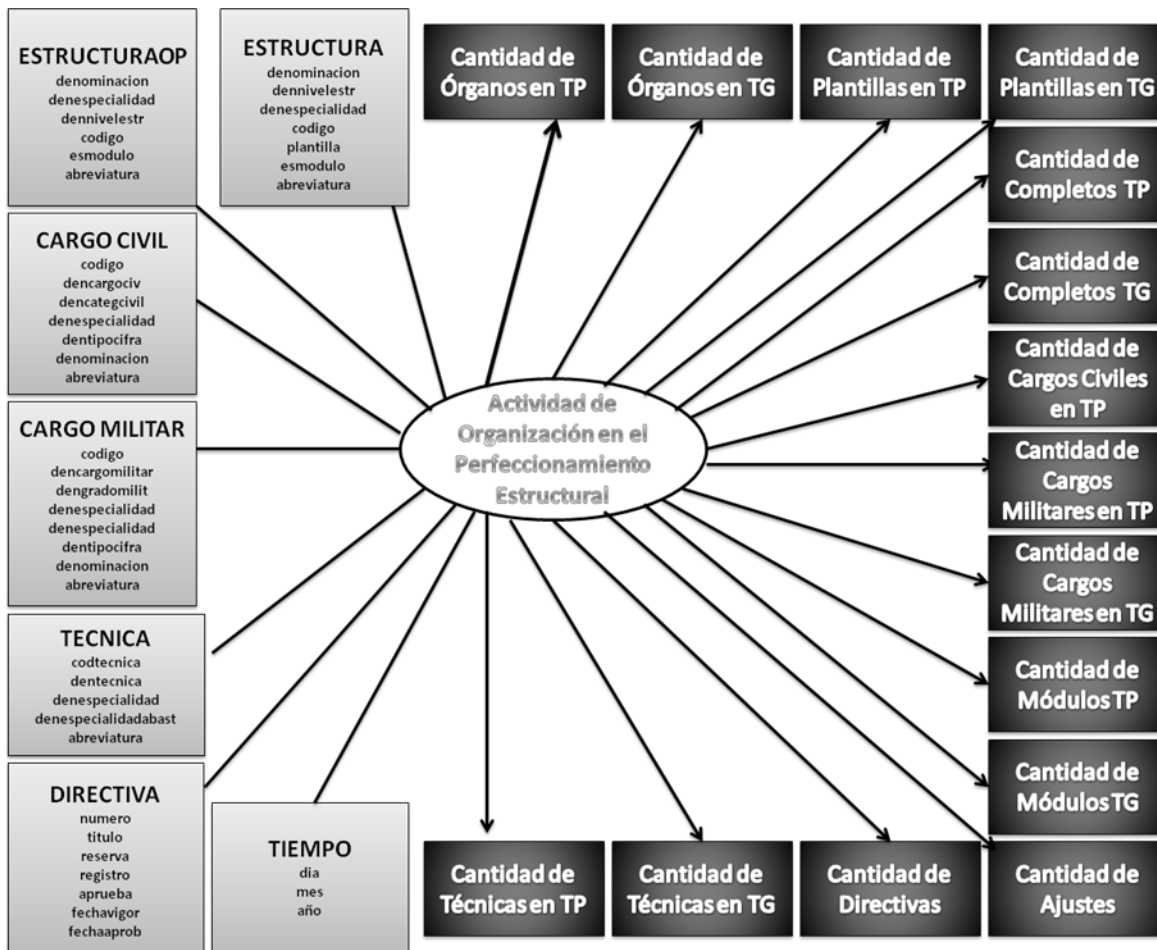


Figura 20: Modelo Conceptual Ampliado.

3.4 Elaboración del modelo lógico de la estructura del almacén de datos.

A continuación, se confeccionará el modelo lógico de la estructura del almacén de datos, teniendo como base el modelo conceptual que ya ha sido creado. Se debe seleccionar cuál será el tipo de esquema (estrella, constelación o copo de nieve) que se utilizará para contener la estructura del depósito de datos, que se adapte mejor a los requerimientos y necesidades del usuario.

En este caso se utilizará el Esquema en Constelación ya que este se adapta mejor a las necesidades propias de las funcionalidades del usuario, además de las características y ventajas que presenta sobre otros esquemas.

3.4.1 Esquema en Constelación

Este modelo está compuesto por una serie de esquemas en estrella, y tal como se puede apreciar en la siguiente figura, está formado por una tabla de hechos principal (“HECHOS_A”) y por una o más tablas de hechos auxiliares (“HECHOS_B”), las cuales pueden ser sumarios de la principal. Dichas tablas yacen en el centro del modelo y están relacionadas con sus respectivas tablas de dimensiones. No es necesario que las diferentes tablas de hechos compartan las mismas tablas de dimensiones, ya que, las tablas de hechos auxiliares pueden vincularse con solo algunas de las tablas de dimensiones asignadas a la tabla de hechos principal, y también pueden hacerlo con nuevas tablas de dimensiones.

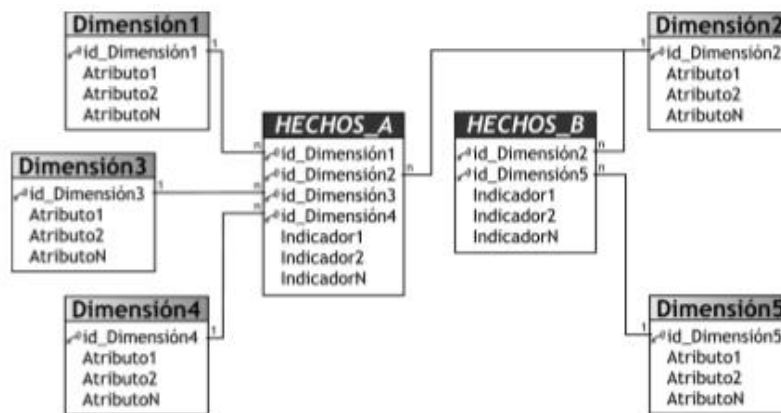


Figura 25: Esquema en Constelación

3.4.2 Diseñar tablas de dimensiones.

Lo primero que se hará será crear las dimensiones que tendrá el DW, para ello se tomará cada perspectiva con sus atributos relacionados y se les realizará el siguiente proceso:

- Se elegirá un nombre que identifique la dimensión.
 - Se añadirá un campo que represente su clave principal.
 - Se redefinirán los nombres de los atributos si es que no son lo bastante explicativos.
- Perspectiva “Estructura”.
 - La nueva dimensión tendrá el nombre “dat_estructura”.
 - Se le agregará una clave principal con el nombre “id_estructura”.
 - Se modificará el nombre del atributo “abreviatura” por “tipoajuste”.

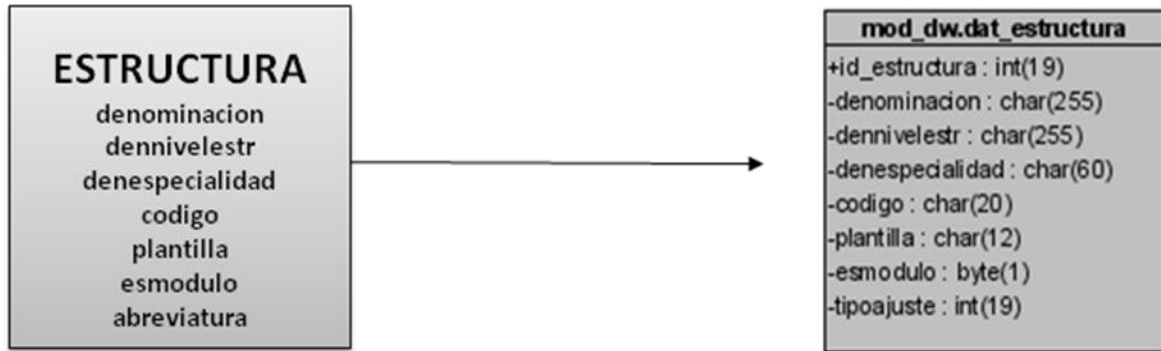


Figura 26: Caso práctico, dimensión “ESTRUCTURA”.

- Perspectiva “Estructuraop”.
 - La nueva dimensión tendrá el nombre “dat_estructuraop”.
 - Se le agregará una clave principal con el nombre “id_estructuraop”.
 - Se modificará el nombre del atributo “abreviatura” por “tipoajuste”.



Figura 27: Caso práctico, dimensión “ESTRUCTURAOP”.

- Perspectiva “Cargo Civil”.
 - La nueva dimensión tendrá el nombre “dat_cargocivil”.
 - Se le agregará una clave principal con el nombre “id_cargocivil”.
 - Se modificarán los nombres de los atributos “abreviatura” por “tipoajuste” y “denominacion” por “función”.



Figura 28: Caso práctico, dimensión “CARGO CIVIL”.

- Perspectiva “Cargo Militar”.
 - La nueva dimensión tendrá el nombre “dat_cargomilitar”.
 - Se le agregará una clave principal con el nombre “id_cargomilitar”.
 - Se modificarán los nombres de los atributos “abreviatura” por “tipoajuste” y “denominacion” por “funcion”.

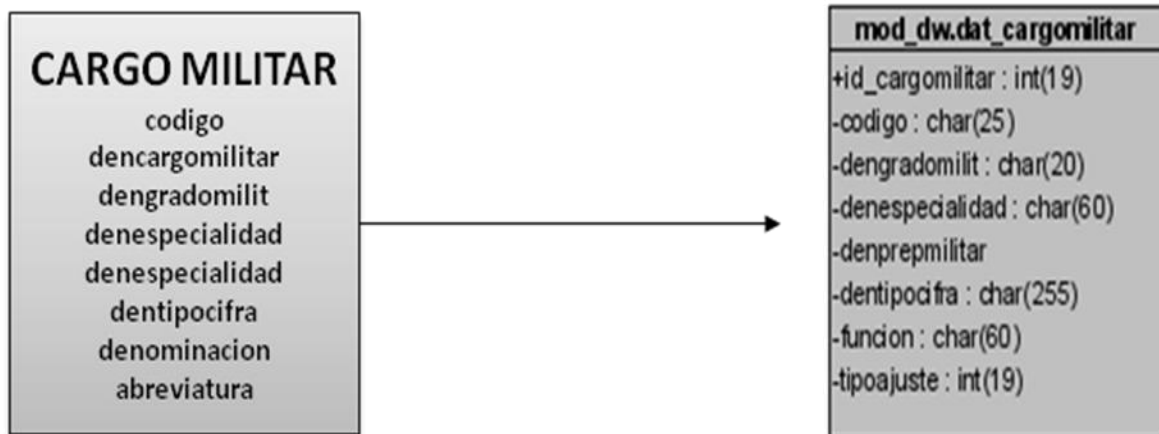


Figura 29: Caso práctico, dimensión “CARGO MILITAR”.

- Perspectiva “Técnica”.
 - La nueva dimensión tendrá el nombre “dat_tecnica”.
 - Se le agregará una clave principal con el nombre “id_tecnica”.

- Se modificará el nombre del atributo “denespecialidad” por “especialidad”.
- Se modificará el nombre del atributo “abreviatura” por “tipoajuste”.

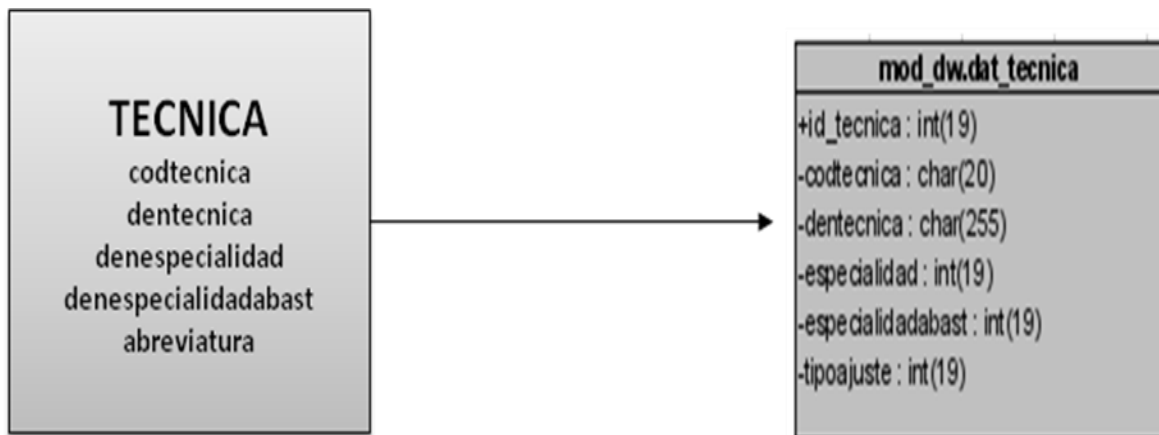


Figura 30: Caso práctico, dimensión “TECNICA”.

- Perspectiva “Directiva”.
 - La nueva dimensión tendrá el nombre “dat_directiva”.
 - Se le agregará una clave principal con el nombre “id_directiva”.
 - No fue cambiado ningún atributo.

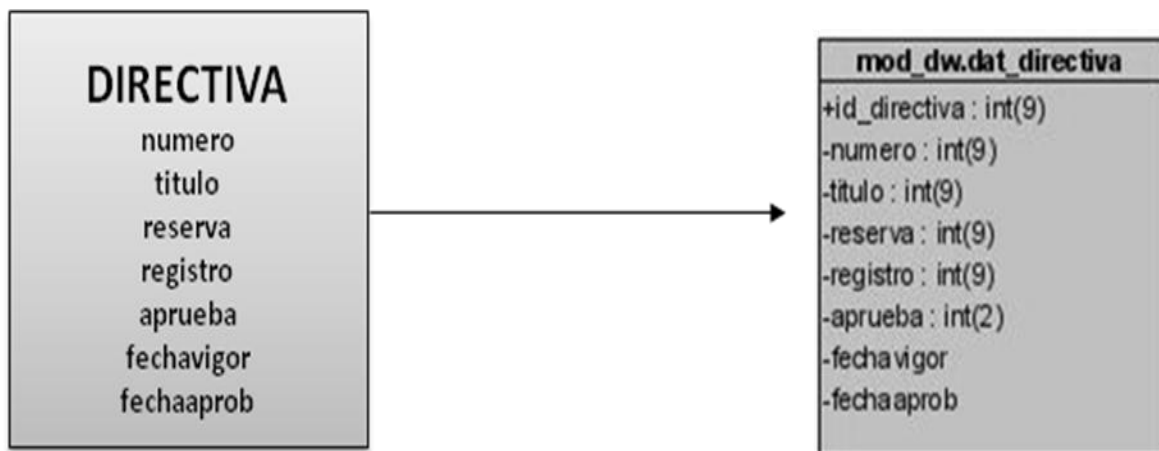


Figura 31: Caso práctico, dimensión “DIRECTIVA”.

- Perspectiva “Tiempo”.
 - La nueva dimensión tendrá el nombre “tiempo”.
 - Se le agregará una clave principal con el nombre “id_fecha”.
 - No fue cambiado ningún atributo.

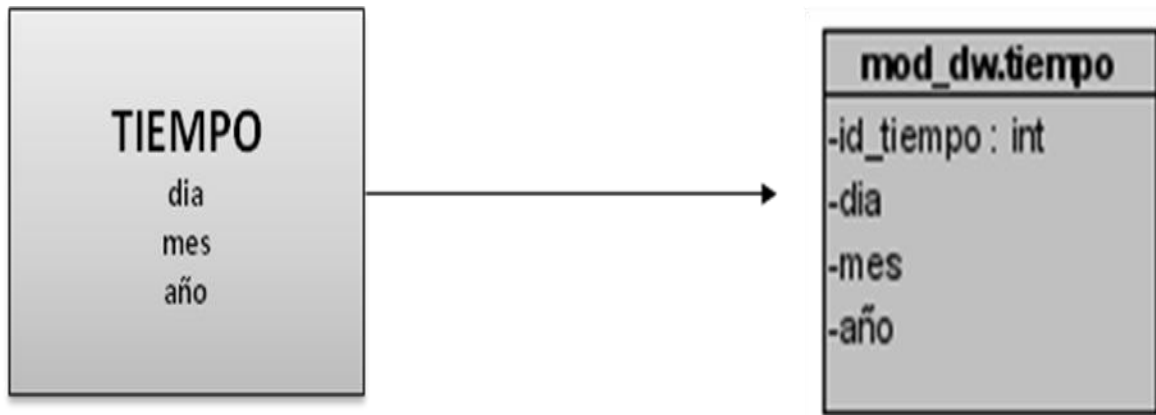


Figura 32: Caso práctico, dimensión “TIEMPO”.

3.5 Diseñar tablas de hechos

Las tablas de hechos contienen los hechos, medidas o indicadores que serán utilizados por los analistas de negocio para apoyar el proceso de toma de decisiones. Los hechos son datos instantáneos en el tiempo, que son filtrados, agrupados y explorados a través de condiciones definidas en las tablas de dimensiones. Los datos presentes en las tablas de hechos constituyen el volumen del almacén de datos, y pueden estar compuestos por millones de registros dependiendo de su granularidad y de los intervalos de tiempo de los mismos. Los más importantes son los de tipo numérico. Una tabla de hecho posee una clave primaria que está compuesta por las claves primarias de las tablas de dimensiones relacionadas con esta.

Los hechos son todas aquellas sumalizaciones o acumulaciones preestablecidas que residen en una tabla de hechos para agilizar las consultas y permitir que los datos puedan ser accedidos por las diferentes dimensiones, y desde luego, explorados por ellas. Las sumalizaciones no están referidas solo a sumas, sino que también a promedios, mínimos, máximos, totales por sector, porcentajes, fórmulas predefinidas, dependiendo de los requerimientos de información del negocio.

3.5.1 Casos para tener en cuenta en la creación de la Tabla de hechos.

Caso 1: Si en dos o más preguntas figuran los mismos indicadores pero con diferentes dimensiones de análisis, existirán tantas tablas de hechos como preguntas cumplan esta condición. Por ejemplo:

"Analizar el Indicador1 por Dimensión1 y por Dimensión2".
"Analizar el Indicador1 por Dimensión2 y por Dimensión3".

Figura 33: Caso 1, preguntas.

Entonces se obtendrá:

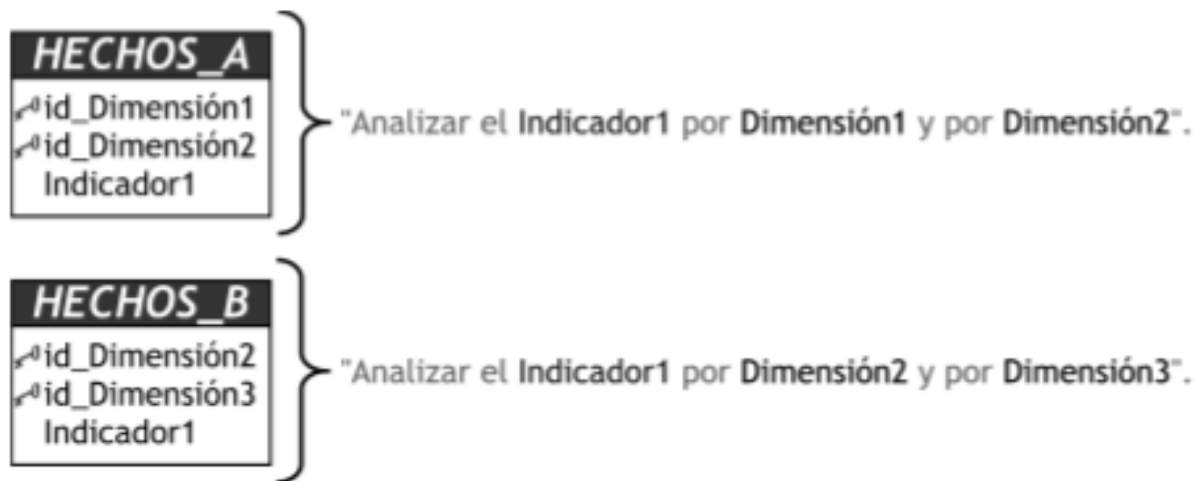


Figura 34: Caso 1, diseño de tablas de hechos

Caso 2: Si en dos o más preguntas figuran diferentes indicadores con diferentes dimensiones de análisis, existirán tantas tablas de hechos como preguntas cumplan esta condición. Por ejemplo:

"Analizar el Indicador1 por Dimensión1 y por Dimensión2".
"Analizar el Indicador2 por Dimensión2 y por Dimensión3".

Figura 35: Caso 2, preguntas

Entonces se obtendrá:

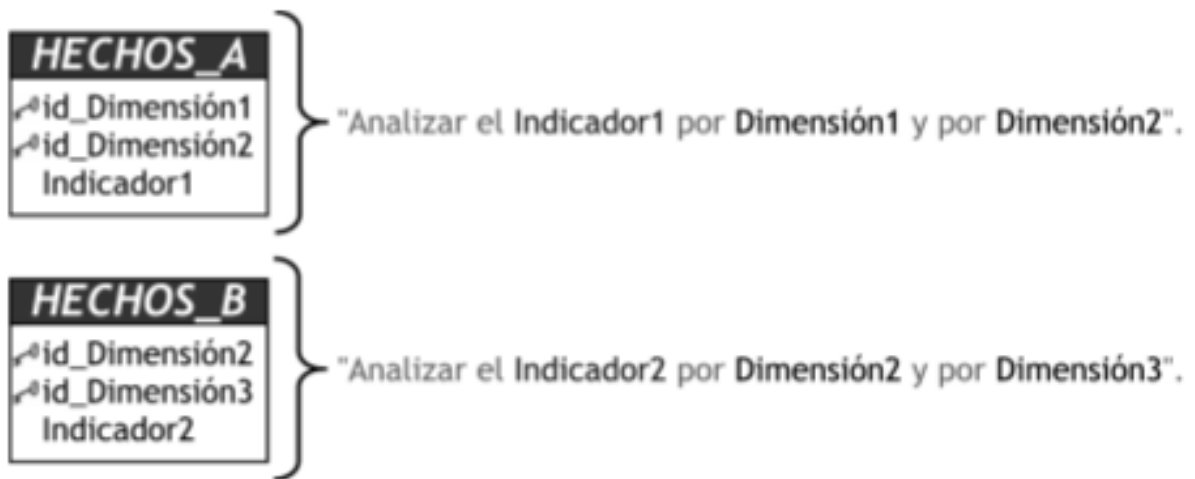


Figura 36: Caso 2, diseño de tablas de hechos

Caso 3: Si el conjunto de preguntas cumple con las condiciones de los dos puntos anteriores se deberán unificar aquellas interrogantes que posean diferentes indicadores pero iguales dimensiones de análisis, para luego reanudar el estudio de las preguntas. Por ejemplo:

"Analizar el Indicador1 por Dimensión1 y por Dimensión2".
"Analizar el Indicador2 por Dimensión1 y por Dimensión2".

Figura 37: Caso 3, preguntas

Se unificarán en:

"Analizar el Indicador1 y el Indicador2 por Dimensión1 y por Dimensión2".

Figura 38: Caso 3, unificación

A continuación se expondrá el diseño de la tabla de hechos del DW así como sus respectivas uniones con las tablas de dimensiones, propuesto para el SIDEC.

CAPÍTULO III. ANÁLISIS Y DISEÑO DEL ALMACÉN DE DATOS

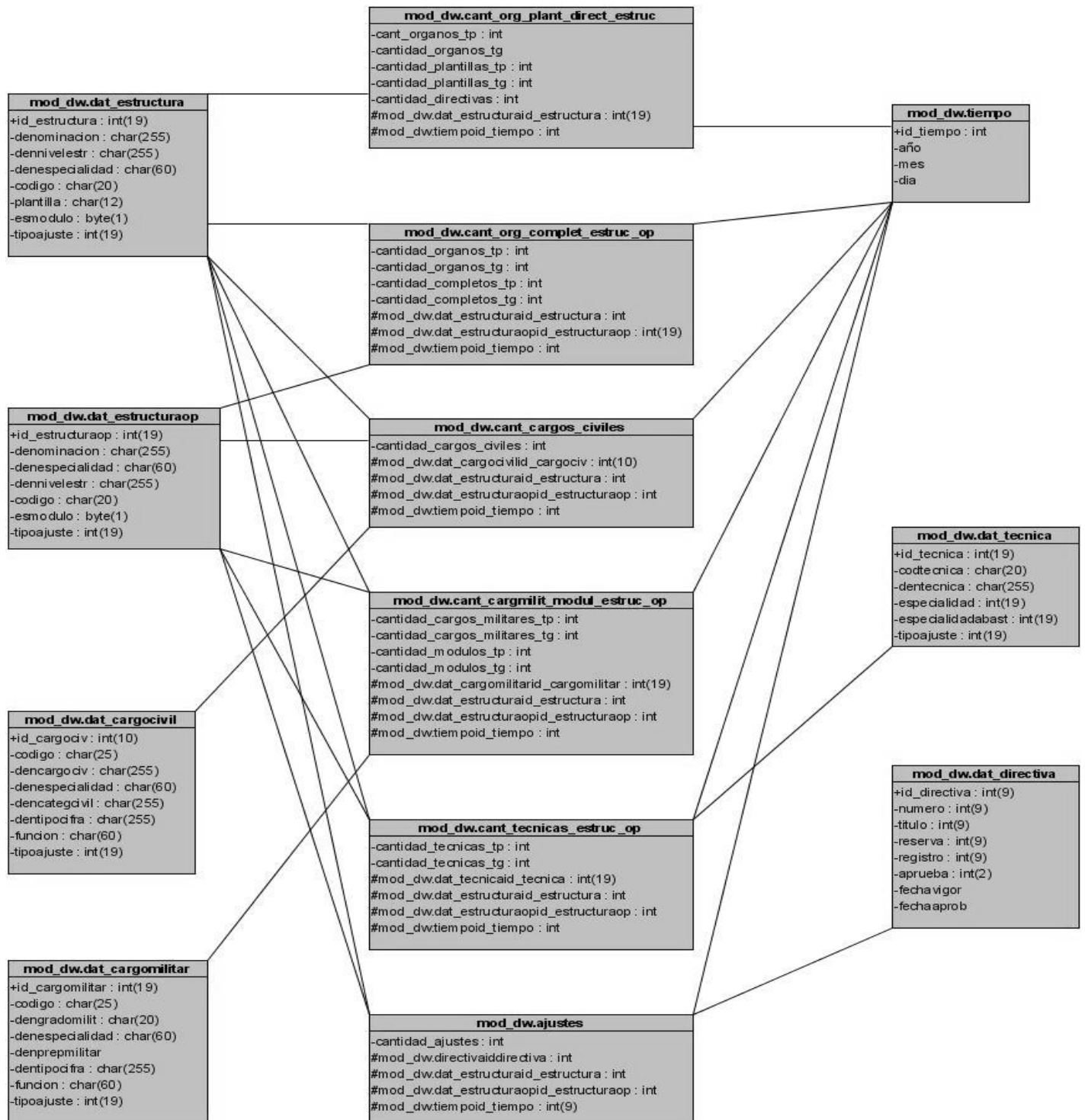


Figura 39: Tabla de hechos

3.6 Flujo de datos

El flujo de los datos desde sus orígenes hasta su presentación al cliente en forma de reportes tabulados, se realiza de manera unidireccional, es decir, en una sola dirección, sufriendo los cambios y las transformaciones necesarias. Se le realizará procesos de extracción y transformación hacia otra Base de datos que contendrá un diseño multidimensional, ubicado en el servidor central de la institución, montada y preparada en el gestor PostgreSQL, dicha información ya almacenada en cubos de datos preparados en el servidor destinado para el DW, podrá ser consultada desde distintas aplicaciones.

La Base de Datos operacional actualizará periódicamente sus datos en la base de datos Multidimensional utilizando paquetes PostgreSQL.

3.7 Conclusiones del capítulo

Como resultado de este capítulo se obtuvo una descripción de los artefactos que fueron creados, fueron descritos todos los pasos necesarios para concebir la creación del Almacén de Datos, se crearon las tablas de Dimensión y Hechos para culminar así con la arquitectura del Almacén de Datos.

CONCLUSIONES

1. Se realizó un estudio del comportamiento de los almacenes de datos así como de la situación existente en la actualidad acerca del tema y se obtuvo como resultado un conjunto de información imprescindible para trazar las pautas hacia donde se debía encaminar este trabajo.
2. Se mencionaron y explicaron de forma breve las características de las herramientas y la metodología que se encuentran definidas por el CATD para el desarrollo de este trabajo.
3. Se realizó un análisis del SIDEC, así como de sus características para identificar las necesidades informativas de los usuarios a partir del almacenamiento de la información histórica generada por sistema. .
4. Partiendo de las necesidades identificadas, se diseñó el Modelo Conceptual del sistema y se realizó la correspondencia de los requerimientos con la base de datos del SIDEC; mediante lo cual se construyó el Modelo Conceptual Ampliado.
5. Se crearon las tablas de Dimensión y Hechos para culminar de esta forma con la arquitectura básica del Almacén de Datos.

RECOMENDACIONES

1. Desarrollar de forma inmediata el Almacén de Datos diseñado en el presente Trabajo de Diploma, para que pueda ser explotado lo antes posible por los usuarios del SIDEC.
2. Luego de concluir el desarrollo del Almacén de Datos para el SIDEC se debe comenzar a explotar inmediatamente para en un lapso no menor de tres meses comenzar a obtener los primeros resultados a través de los indicadores.
3. Realizar un estudio en dependencia de las necesidades de la Jefatura de las FAR para aplicar técnicas de minería de Datos con el objetivo de extraer conocimiento útil y comprensible, de identificar patrones válidos y novedosos para asegurar el proceso de Toma de Decisión.

BIBLIOGRAFÍA

1. **Tamargo, MSc. Lic. Lourdes Cerezal.** La Revista del Empresario Cubano. La Revista del Empresario Cubano. [En línea] Disaic. [Citado el: 10 de enero de 2010.] http://www.betsime.disaic.cu/secciones/tec_ja_04.html.
2. HEFESTO: Metodología propia para la Construcción de un Data Warehouse. **Bernabeu, Ricardo Dario.** Córdoba: Argentina, 2007.
3. **Kimball, R. y Margy, R. 2002. . 2002.** The Data Warehouse Toolkit. Arizona: John Wiley & Sons Ltd, 2002. ISBN-10: 0471200247.
4. **Inmon, W. H. 2002.** Building the Data Warehouse. New York: New York: John Wiley & Sons, 2002. ISBN-10: 0471141615.
5. Sun MicroSystem Hispanoamérica. Sun MicroSystem Hispanoamérica. [En línea] 2008. <http://www.mx.sun.com/sunnews/press/2008/081003.jsp>.
6. Visa aprovecha pontecial de los datos de sus tarjetas. Visa aprovecha pontecial de los datos de sus tarjetas. [En línea] <http://es.sun.com/sunnews/success/pdf/visa.pdf>.
7. **Jorge Rubio Navarro, José Manuel Salinas.** DataTur Almacén de datos para el análisis y difusión de la información estadística del turismo en España. DataTur Almacén de datos para el análisis y difusión de la información estadística del turismo en España. [En línea] 2002. <http://www.turismo.uma.es/turitec/turitex2002/actas/Microsoft Word - 3.Rubio.pdf>.
8. Sistema Data Warehouse comercial de la corporación CIMEX. Sistema Data Warehouse comercial de la corporación CIMEX. [En línea] <http://www.redciencia.cu/empres/Intempres2002/Seminarios/CIMEXSAponencia.pdf>.

REFERENCIAS BIBLIOGRÁFICAS

1. **Tamargo, MSc. Lic. Lourdes Cerezal.** La Revista del Empresario Cubano. La Revista del Empresario Cubano. [En línea] Disaic. [Citado el: 10 de enero de 2010.] http://www.betsime.disaic.cu/secciones/tec_ja_04.html.
2. HEFESTO: Metodología propia para la Construcción de un Data Warehouse. **Bernabeu, Ricardo Dario.** Córdoba : Argentina, 2007.
3. **Kimball, R. y Margy, R. 2002. . 2002.** The Data Warehouse Toolkit. Arizona : John Wiley & Sons Ltd, 2002. ISBN-10: 0471200247.
4. **Inmon, W. H. 2002.** Building the Data Warehouse. New York : New York: John Wiley & Sons, 2002. ISBN-10: 0471141615.
5. Sun MicroSystem Hispanoamérica. Sun MicroSystem Hispanoamérica. [En línea] 2008. <http://www.mx.sun.com/sunnews/press/2008/081003.jsp> .
6. Visa aprovecha pontencial de los datos de sus tarjetas. Visa aprovecha pontencial de los datos de sus tarjetas. [En línea] <http://es.sun.com/sunnews/success/pdf/visa.pdf>.
7. **Jorge Rubio Navarro, José Manuel Salinas.** DataTur Almacén de datos para el análisis y difusión de la información estadística del turismo en España. DataTur Almacén de datos para el análisis y difusión de la información estadística del turismo en España. [En línea] 2002. <http://www.turismo.uma.es/turitec/turitex2002/actas/Microsoft Word - 3.Rubio.pdf>.
8. Sistema Data Warehouse comercial de la corporación CIMEX. Sistema Data Warehouse comercial de la corporación CIMEX. [En línea] <http://www.redciencia.cu/empres/Intempres2002/Seminarios/CIMEXSAponencia.pdf>.

GLOSARIO DE TÉRMINOS

OLTP: Del ingles (On Line Transaction Processing). Conocido también como Procesamiento Operacional en línea.

OLAP: Del ingles (On Line Analytical Processing). Conocido también como Procesamiento analítico en línea, usado para la Toma de Decisiones.

Data Warehouse: Este término se refiere a un Almacén de Datos o bodega de datos, también ha sido utilizada la abreviatura DW.

Indicadores: Constituyen valores numéricos y representan lo que se desea analizar concretamente, por ejemplo: saldos, promedios, cantidades, sumatorias, fórmulas.

Perspectivas: Se refieren a los objetos mediante los cuales se quiere examinar los indicadores, con el fin de responder a las preguntas planteadas, por ejemplo: clientes, proveedores, sucursales, países, productos, rubros. Cabe destacar, que el Tiempo es muy comúnmente una perspectiva.

Modelo Conceptual: Especie de planilla compuesta por Indicadores, Perspectivas y Relaciones.

ETL: Del ingles (Extraction transformation and load). Procesos de transformación extracción y carga de Datos en un Almacén de Datos.