

Universidad de las Ciencias Informáticas

Facultad 3



Almacén de datos para el Sistema de Planificación de Actividades SIPAC.

Trabajo de Diploma para optar por el título de Ingeniero en Ciencias Informáticas.

Autores: Susel Milagro Rosabal Colas

Juan Fuentes Bauta

Tutores: Ing. Maria Teresa Rosales González

Ing. Dionny Cardoso Carmona

La Habana, septiembre de 2013

“Año 55 de la Revolución”

DECLARACIÓN DE AUTORÍA

Declaramos ser autores de la presente tesis y reconocemos 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 _____.

Susel M. Rosabal Colás

Juan Fuentes Bauta

Firma del Autor

Firma del Autor

Ing. Maria T. Rosales González

Ing. Dionny Cardoso Carmona

Firma del Tutor

Firma del Tutor

AGRADECIMIENTOS

Existen varias personas importantes en mi vida que me apoyaron e hicieron posible llegar hasta aquí, y realizar un sueño.

Quiero agradecer a mi mamá que ha sido mi inspiración, mi mayor apoyo, mi sostén cuando he podido caer, ella es la parte más importante de este sueño.

A mi abuelita por su silencio cuando no se necesitan palabras, solo su mirada me transmite lo que necesito saber para sentir su amor.

A mis hermanos por permitirme ser su ejemplo a pesar de mi corta edad en comparación con ellos, por sus palabras conservadoras, pero tiernas y confortables.

A mis tías y tíos, a Luis Enrique por dejarme ser su orgullo y hacerme sentir como una hija para él.

A mis primas Berenice, Mayara y a mis sobrinitas que me han dado mucha alegría, las pequeñas Daphne Valeria, Dilara Deniz y Essined.

A mi novio por su comprensión, por sus consejos, por sus críticas que me hacían más fuerte cada día, por su amor, por su entrega, por su presencia.

Agradecer a mis tutores por su apoyo incomparable, por sus palabras, por ser nuestros guías, por dejarnos saber que podíamos llegar.

A mi compañero de tesis por su entrega y por la química que nos permitió llegar al final del camino.

Susel M. Rosabal Colás

Primeramente quiero agradecerle a Dios porque si hoy he llegado hasta aquí ha sido por su gracia, su misericordia y su amor.

A mis padres por su amor, comprensión, dedicación, preocupación y apoyo en todos estos años de mi vida.

A mis hermanas Yanet y Yamile por su cariño y consejos para seguir adelante en todos los años de mis estudios.

A toda mi familia por su confianza en mí en especial a mi abuela Amalia y mi tía Zaida.

A mi otra familia en Cristo, mis hermanos de la UCI por su apoyo y compañía en estos 5 años de universidad en especial a Yerandy, Yasmery, Yasel, Yaneisi y Adyana.

A mis tutores Maria Teresa y Dionny por haberme apoyando y guiado en el desarrollo de este trabajo.

A mi compañera de tesis Susel, sin ella este trabajo no se hubiera completado.

Juan Fuentes Bauta

RESUMEN

El Sistema de Planificación de Actividades (SIPAC) es una herramienta que gestiona las actividades a todos los niveles organizacionales. Independientemente de las funcionalidades que el sistema brinda no posibilita unificar la información almacenada en las diferentes bases de datos, ni evaluar el comportamiento de los diferentes indicadores. En la presente investigación se propone una solución basada en el desarrollo de un Almacén de datos o Datawarehouse (DWH) que permita hacer análisis estadísticos históricos de los principales indicadores de las áreas de análisis con las que cuenta el SIPAC, facilitando la toma de decisiones a los directivos de las entidades que hagan uso del sistema. Para ello se realiza el análisis, diseño e implementación de los subsistemas de almacenamiento, integración y visualización, así como la aplicación de listas de chequeo y los casos de prueba con el objetivo de obtener un producto que cumpla con las necesidades del cliente. Finalmente se obtiene un DWH poblado y funcional, con una capa de inteligencia de negocio que brinda vistas de análisis actualizadas, que cumple con los requerimientos especificados y con la calidad requerida.

Palabras claves: Almacén de datos, niveles organizacionales, indicadores, áreas de análisis, toma de decisiones.

ÍNDICE

DECLARACIÓN DE AUTORÍA.....	I
AGRADECIMIENTOS	II
RESUMEN	III
INTRODUCCIÓN	1
CAPITULO 1: FUNDAMENTOS TEÓRICOS SOBRE EL DESARROLLO DE LOS ALMACENES DE DATOS.....	5
1.1 Almacenes de datos.....	5
1.1.1 Características de un almacén de datos.....	5
1.2 Estado del arte.....	6
1.2.1 Almacenes de datos en el mundo	6
1.2.2 Almacenes de datos en Cuba	8
1.2.3 Valoración	10
1.3 Procesos Básicos de un DWH	11
1.3.1 Integración de datos.....	11
1.3.2 Inteligencia de Negocios (BI).....	12
1.4 Procesamiento Analítico en Línea (OLAP)	12
1.4.1 Implementaciones OLAP.....	13
1.5 Metodologías para el desarrollo de Almacenes de Datos.....	13
1.5.1 Metodología propuesta por William H. Inmon.....	14
1.5.2 Metodología propuesta por Ralph Kimball.....	14
1.5.3 Metodología de Desarrollo de Almacenes de Datos de DATEC.	15
1.6 Herramientas seleccionadas para la solución.....	16
1.6.1 Herramienta de modelado.....	16
1.6.2 Sistema Gestor de Base de Datos	16
1.6.3 Herramientas para la integración de datos	17
1.6.4 Herramientas para inteligencia de negocio.....	18

1.6.5	Herramientas para soporte de aplicación.....	19
1.6.6	Herramienta para el control de versiones.....	19
1.7	Conclusiones del capítulo	20
CAPITULO 2: ANÁLISIS Y DISEÑO DEL ALMACÉN DE DATOS PARA EL SISTEMA DE PLANIFICACIÓN DE ACTIVIDADES (SIPAC).		21
2.1	Estudio preliminar del negocio	21
2.2	Necesidades de los usuarios.....	22
2.2.1	Requisitos de información	22
2.2.2	Requisitos funcionales	23
2.2.3	Requisitos no funcionales	24
2.3	Reglas del negocio.....	25
2.4	Diagrama de Caso de Uso del sistema	25
2.4.1	Actores del sistema.....	26
2.5	Arquitectura del sistema.....	27
2.6	Modelo Multidimensional.....	28
2.6.1	Características	29
2.6.2	Esquema.....	30
2.7	Diseño del almacén de datos	30
2.7.1	Diseño del subsistema de almacenamiento	30
2.7.2	Diseño del subsistema de integración	35
2.7.3	Diseño del subsistema de visualización	37
2.8	Política de respaldo y recuperación.....	38
2.9	Esquema de seguridad	39
2.10	Conclusiones del capítulo	41
CAPITULO 3: IMPLEMENTACIÓN Y PRUEBA DEL ALMACÉN DE DATOS PARA EL SISTEMA DE PLANIFICACIÓN DE ACTIVIDADES (SIPAC).....		42
3.1	Implementación del subsistema de almacenamiento.....	42

3.1.1	Estándares de codificación.....	42
3.1.2	Modelo de datos físico.	44
3.2	Implementación del subsistema de integración de datos.....	46
3.2.1	Gestión del cambio en las dimensiones.	47
3.2.2	Gestión de los metadatos del proceso de integración de datos.....	48
3.2.3	Captura de los cambios en los datos.....	49
3.2.4	Implementación de los hechos.	49
3.2.5	Implementación de los trabajos.....	50
3.3	Implementación del subsistema de visualización.	51
3.3.1	Cubos multidimensionales.....	51
3.3.2	Reportes.	54
3.4	Resultados obtenidos.....	55
3.5	Validación de la solución.....	55
3.5.1	Herramientas para la aplicación de las pruebas.	57
3.6	Evaluación de los resultados.....	63
3.7	Conclusiones del capítulo.....	65
	CONCLUSIONES GENERALES	66
	RECOMENDACIONES.....	67
	REFERENCIAS BIBLIOGRÁFICAS	68
	BIBLIOGRAFÍA	72
	GLOSARIO DE TÉRMINOS.....	73

INTRODUCCIÓN

El desarrollo de las herramientas de las Tecnologías de la Información y las Comunicaciones (TIC) ha pasado a ser un elemento estratégico en términos de competitividad para las organizaciones dentro de un entorno globalizado; lo que implica acciones de análisis, diseño, implementación de soluciones informáticas y una readaptación de los procesos organizacionales. Dichas acciones estimulan el constante perfeccionamiento de las empresas en cuanto a las tecnologías que utilizan y comercializan. En este entorno surgen los sistemas de información que permiten estudiar los datos históricos de cada empresa.

El reto de una organización, es integrar sus datos corporativos en un único repositorio sobre el cual los usuarios puedan realizar consultas o informes y hacer análisis de datos. Así, nacieron las bases de datos multidimensionales llamadas Almacenes de datos o Datawarehouse (DWH), como la solución para obtener un sistema capaz de dar soporte a la toma de decisiones estratégicas y tácticas (1). Los DWH surgen dada la necesidad de resolver problemas de análisis de grandes cantidades de información y para ayudar a la toma de decisiones con la capacidad de realizar análisis multidimensionales en el tiempo.

Las empresas utilizan la planificación como un esquema básico que le propiciará establecer un sistema de toma de decisiones que garantice aprovechar las oportunidades, reduciendo al mínimo los riesgos. Para garantizar una buena planificación, con el objetivo de lograr una transparencia en la gestión profesional, posibilitando priorizar, esclarecer y unificar los propósitos de las entidades surge el Sistema de Planificación de Actividades (SIPAC); herramienta para la gestión de las actividades a todos los niveles organizacionales. SIPAC cuenta con varios módulos encargados de generar las configuraciones necesarias para el seguimiento de las tareas principales en función del cumplimiento de los objetivos de cada entidad; gestión de los posibles involucrados, nomencladores y niveles de subordinación basados en reglas de la compartimentación de la información. Permite interrelacionar objetivos de trabajo y actividades en tiempo real, crear el flujo de información hacia los diferentes niveles a partir de la definición de estados y transiciones, así como la recuperación de la información de acuerdo al modelo de planificación actual. Todas estas operaciones que SIPAC permite realizar generan información estratégica u operativa, la primera es la base de la toma de decisiones a largo plazo en las entidades y organismos y la segunda es la información latente de la planificación.

Independientemente de las funcionalidades que brinda el sistema, no da la posibilidad de:

1. Unificar la información almacenada en diferentes bases de datos.
2. Analizar la información unificada para evaluar el comportamiento de los diferentes indicadores, esto se traduce en:
 - Conocer la cantidad de tareas cumplidas asociadas a un usuario en un período determinado.
 - Ver el porcentaje de las tareas cumplidas con respecto a los planes trazados para darle cumplimiento a los objetivos propuestos en un período determinado.
 - Analizar el estado de cumplimiento de los planes y objetivos trazados en un instante de tiempo.
 - Conocer la cantidad de Factores que Influyen en el Plan (FIP) para cada usuario involucrado en el proceso de planificación de actividades.
 - Conocer el porcentaje de las Áreas de Resultados Clave (ARC) asociadas a los capítulos y usuarios.

Este conjunto de factores trae como consecuencias una demora en el análisis y estudio de la información, los reportes que agrupan la totalidad de la misma se realiza de forma engorrosa provocando ineficiencia en la obtención de resultados medibles asociados a cada área de análisis, en un período determinado. Todo esto puede incidir de manera negativa en la ejecución de los procesos regulares y en la prevención de comportamientos futuros de las entidades, afectando su función como institución.

Teniendo en cuenta los elementos planteados anteriormente se identifica como **problema a resolver**: ¿Cómo visualizar la información estratégica a partir de la integración de los indicadores clave de la planificación, de manera que facilite los procesos de Ejecución y Control de la Planeación Estratégica u Operativa?

Con vista a la solución del problema se define como **objeto de estudio**: El proceso de desarrollo de almacenes de datos. La presente investigación se encuentra enmarcada en el **campo de acción**: El proceso de desarrollo de almacenes de datos para sistemas de planificación de actividades, trazándose como **objetivo general**: Desarrollar un Almacén de datos para el Sistema de Planificación de Actividades (SIPAC) para visualizar la información estratégica a partir de la integración de los indicadores clave de la planificación que facilite los procesos de Ejecución y Control de la Planeación Estratégica u Operativa. Desglosándose en los siguientes objetivos específicos:

Objetivos específicos:

1. Obtención del marco teórico para fundamentar la investigación.
2. Implementación de la solución teniendo en cuenta las necesidades del cliente.
3. Validación de la solución mediante la aplicación de técnicas de caja negra, listas de chequeo y el perfilado de los datos.

Para el desarrollo de la investigación fue necesario trazar las siguientes **tareas de la investigación:**

1. Estudiar el estado del arte referente a soluciones informáticas relacionadas con el campo de acción para seleccionar la metodología a utilizar.
2. Definir la arquitectura del almacén de datos para tener una visión general de la solución.
3. Identificar, especificar y validar los requisitos funcionales, no funcionales y de información para el buen desarrollo del análisis.
4. Obtener el diseño de la solución para un mejor entendimiento del negocio.
5. Implementar los requisitos identificados para solucionar las necesidades del cliente.
6. Validar la solución mediante técnicas de pruebas para garantizar la calidad del resultado final.

Por lo que se plantea como **hipótesis** la siguiente: Si se desarrolla un almacén de datos para el Sistema de Planificación de Actividades (SIPAC) se visualizará la información a partir de los indicadores claves de la planificación que facilite los procesos de Ejecución y Control de la Planeación Estratégica u Operativa.

La **estructura del documento** se encuentra definida en tres capítulos.

Capítulo 1: Fundamentos teóricos sobre el desarrollo de los Almacenes de datos.

En este capítulo se abordaron varios elementos teóricos sobre los almacenes de datos (DWH). Se llevó a cabo la revisión bibliográfica con relación a las metodologías de desarrollo de los DWH, seleccionando la metodología a utilizar. Además se especificaron las herramientas que se emplean para solucionar el problema en cuestión.

Capítulo 2: Análisis y diseño del Almacén de datos para el Sistema de Planificación de Actividades (SIPAC).

En este capítulo se realizó el diseño de los subsistemas de almacenamiento, integración y visualización y se llevó a cabo el análisis del negocio mediante el levantamiento de los requisitos funcionales, no funcionales y de información. Se especificaron las reglas del negocio, así como los casos de uso del sistema. Además se definió la arquitectura del sistema, el esquema de seguridad y las políticas de respaldo y recuperación.

Capítulo 3: Implementación y prueba del Almacén de datos para el Sistema de Planificación de Actividades (SIPAC).

En este capítulo se realizó la implementación de los subsistemas de almacenamiento mediante la creación del modelo de datos. Además se implementó el subsistema de integración, así como los flujos de transformación y los trabajos. También se realizó la implementación del subsistema de visualización, el cual comprende la implementación de los cubos OLAP y los reportes candidatos. En este capítulo además se presentaron las pruebas realizadas a la solución, mediante la aplicación de las listas de chequeo y los casos de prueba (CP) por Caso de Uso (CU), así como los resultados obtenidos en cada una de ellas luego de su aplicación.

CAPITULO 1: FUNDAMENTOS TEÓRICOS SOBRE EL DESARROLLO DE LOS ALMACENES DE DATOS.

En el presente capítulo se realizó una revisión conceptual de los Almacenes de Datos o Datawarehouse (DWH), abordándose varios elementos teóricos como características, procesos básicos y se llevó a cabo un estudio detallado del estado del arte de estos sistemas de información incluyendo la metodología de desarrollo seleccionada. Además se especificaron las herramientas que se seleccionaron para dar solución al problema en cuestión.

1.1 Almacenes de datos.

Los almacenes de datos son un escalón superior en la evolución de la Base de Datos (BD) hacia mayor inteligencia y mejores funcionalidades. Muchos autores han definido de diversas formas lo que representa un almacén de datos, por lo que es difícil formular una definición rigurosa. A continuación se citan algunas de ellas:

Bill Inmon fue el primero que introdujo este término, el padre de esta disciplina, define un almacén de datos como una colección de datos orientados por temas, integrados, variables en el tiempo y no volátil para el apoyo de la toma de decisiones. (2)

Ralph Kimball es otro conocido autor en el tema de los almacenes de datos y los define como una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis, es la unión de todos los mercados de datos de una entidad. (3)

Luego de haber analizado algunas de las fuentes bibliográficas se definió como **Almacén de Datos**: una colección de datos que recoge información de múltiples sistemas fuentes u operacionales dispersos, orientada al análisis de la información histórica contenida, apoyando de esta forma el análisis y la toma de decisiones en una empresa.

1.1.1 Características de un almacén de datos.

El uso de los almacenes de datos es amplio y llega a diferentes tipos de organizaciones y distintos temas de interés. Ellos tienen un conjunto elementos que lo caracterizan, dentro de las principales se encuentran las siguientes: integrado, orientado a tema, de tiempo variante y no volátil. (4)

Integrado: El aspecto más importante del ambiente del DWH es que la información encontrada en el interior esté siempre integrada. La integración de datos se muestra de muchas maneras:

Capítulo 1: Fundamentos teóricos sobre el desarrollo de los almacenes de datos

en convenciones de nombres consistentes, en la medida uniforme de variables, en la codificación de estructuras consistentes, en atributos físicos de los datos consistentes, fuentes múltiples y otros. Se construye mediante la integración de datos múltiples y heterogéneos, así como BD relacionales, ficheros y registros de transacciones on-line. (4)

Orientado a tema: Los datos son organizados y resumidos por temas, tales como Objetivos, Actividades, Áreas de Resultados Claves. La información se clasifica en base a los aspectos que son de interés para la empresa, proporcionando una vista simple y concisa de las materias de negocio excluyendo aquellos datos que no se necesitan para el proceso de toma de decisiones. (4)

De tiempo variante: El DWH almacena datos relativos a un período de tiempo y estos datos se incrementan periódicamente. Por esto se dice que la única dimensión que está presente en todo DWH es la dimensión *Tiempo*, ya que es la que se encarga de controlar la fecha en la que ocurre cada suceso. (4)

No volátil: La actualización (insertar, borrar y modificar), se hace regularmente en el ambiente operacional sobre una base de registro por registro, pero la manipulación básica de los datos que ocurre en el DWH es mucho más simple. Hay dos únicos tipos de operaciones: la carga inicial de datos y el acceso a los mismos. No hay actualización de datos (en el sentido general de actualización) en el depósito, como una parte normal de procesamiento. (4) (5)

1.2 Estado del arte

A través de los años los DWH se han convertido en un recurso importante de apoyo a la toma de decisiones en un amplio número de organizaciones a nivel internacional y nacional. En este acápite se da una valoración de cómo se comporta el uso de los DWH en el mundo y específicamente en Cuba.

1.2.1 Almacenes de datos en el mundo

En la actualidad la mayoría de las organizaciones cuentan con un sistema de información que soporta gran parte de las actividades diarias propias del sector de negocios en donde se esté desempeñando. Este sistema puede ser sencillo o robusto todo depende de las exigencias del negocio y con el transcurso del tiempo estas aplicaciones llegan a tener la historia de la organización. Los DWH aseguran una vista única de los datos, que pueden provenir de diversas fuentes.

DWH para la Corett Colima

Fortalecer el trabajo que realiza en sus oficinas el Comité para la Regularización de la Tenencia de Tierra (Corett¹) requiere de una mejor organización y un mejor uso de sus recursos, principalmente los informáticos. Por ende se desarrolló el DWH Corett para mejorar los servicios que el comité ofrece a sus usuarios en el trámite de regularización de sus pisos y el ordenamiento cartográfico que se requiere para acceder a los servicios municipales en México. (6)

Este sistema se implementa como un solo software que integra las 5 áreas que intervienen en el proceso de regularización, aportando a la investigación un elemento más a priorizar. Sin embargo resulta algo lento para cargar toda la estructura del almacén de datos y no implementa un archivo digital para los documentos que forman el expediente físico, que se genera después del trámite, es un documento de consulta que sería muy deseable poder consultar digitalmente y que le daría una mayor atención a la información manejada en el sistema.

DWH para los Laboratorios Farmacéuticos de San Salvador

Este DWH está dirigido a los mandos medios de los Laboratorios Farmacéuticos ubicados en la Zona Metropolitana de San Salvador, los cuales se dedican constantemente a la toma de decisiones oportunas tomando de referencias la información de los reportes e informes que son elaborados. El DWH para los Laboratorios Farmacéuticos utiliza la metodología que propone Kimball, pues ellos concentran la información específica de los departamentos o áreas de negocio en los almacenes más pequeños de información de un área en específico, conocidos como mercados de datos o datamarts, basándose en una estrategia de afuera hacia adentro es decir utilizar información diversa que permita llegar a un análisis específico. Este fue un aspecto a tener en cuenta a la hora de diseñar el almacén de datos para SIPAC. (7)

Este DWH tiene algunos elementos que aportan a la investigación de manera general, pero deja algunas dudas en cuanto a la granularidad, elemento indispensable en todo sistema informático que permite la reutilización de la información. Esto puede verse orientado al tamaño de las bases de datos, la complejidad con las que son desarrolladas las consultas y el número de usuarios que accederán a la información.

¹ *Organismo público descentralizado de Poder Ejecutivo Federal de México que regulariza la tenencia de tierras en asentamientos humanos irregulares ubicados en zonas de origen social y de propiedad federal.*

Programa Fast Track DWH de Microsoft SQL Server

Con Microsoft SQL Server 2012, Microsoft continúa siendo el líder en cantidad y calidad al ofrecer capacidades completas para admitir aplicaciones de almacenamiento de datos. Este ha introducido el programa Fast Track DWH de Microsoft SQL Server para permitir que las organizaciones implementen rápidamente infraestructuras de almacenes de datos empresariales predecibles. Fast Track 4.0 de Microsoft SQL Server, la cuarta generación de programas de esta iniciativa, se ha diseñado para posibilitar que las organizaciones implementen sin retrasos y con confianza una solución de sistema completa, en la que se incluye el servidor, el almacenamiento y la conectividad para los almacenes de datos basados en Microsoft SQL Server 2012. (8)

Fast Track es una excelente oportunidad para las organizaciones de implementar configuraciones probadas mediante tecnologías demostradas que ofrecen soluciones equilibradas de un modo rápido y fiable. Sin embargo, su configuración necesita tecnologías que no son propios del entorno en que se desarrollará el DWH para SIPAC, ni aun la que propone para las medianas empresas, queda fuera del alcance económico y del entorno tecnológico con el que se cuenta.

1.2.2 Almacenes de datos en Cuba

DWH para los Ensayos Clínicos que se gestionan en el Centro de Inmunología Molecular (CIM).

El DWH para los ensayos clínicos que se gestionan en CIM tiene como objetivo almacenar la mayor cantidad de información y con ello viabilizar la integración de los datos de la institución que posteriormente serán analizados. Se emplea como metodología de arquitectura la propuesta por Kimball. Las principales áreas de análisis son tratamiento, datos demográficos (edad, sexo, raza, tratamientos previo, etc.), seguridad, eficacia, respuesta. Se realiza el análisis de medidas como cantidad de respuestas, tiempo de supervivencia, cantidad de pacientes, entre otros. (9)

Sistema DWH comercial de la Corporación CIMEX

El sistema DWH de 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

Capítulo 1: Fundamentos teóricos sobre el desarrollo de los almacenes de datos

con la calidad requerida, aumentando las ganancias o utilidades de la empresa”. Dentro de las áreas de análisis que presenta se encuentran las siguientes: venta, inventarios, compra, transacciones de ajuste y de transferencia. Los indicadores con que cuenta son totales, promedios, valores medios, mínimo, máximo, etc. También realiza el análisis de otras medidas importante del negocio, aunque no primarias sino obtenidas como resultado de previo cálculos, tales como utilidad, precio de venta, días de existencia, uso, rotación, entre otras. (10)

DWH del Sistema Estadístico Nacional (SEN) para la Oficina Nacional de Estadísticas (ONE)

El DW del Sistema Estadístico Nacional para la ONE tiene como objetivo maximizar la eficiencia de cómo llevar a los usuarios reportes personalizados, tablas y resúmenes que se personalizan por cada una de las áreas en un menor costo de tiempo. Cuenta con áreas como son censos y encuestas, registros administrativos e informaciones que suplen los diferentes sectoriales. Hace empleo de una plataforma tecnológica de alto nivel como es el caso de Oracle. (11)

A continuación se muestra una tabla descriptiva que aborda elementos importantes sobre los DWH anteriormente mencionados:

Tabla 1. Descripción de almacenes de datos estudiados

Almacenes de datos	País	Tecnología	Indicadores y perspectivas de análisis
Corret Colima	México	Microsoft Visual Basic Express AcccDb ² ODBMS ³ VB.net ⁴ Microsoft.net DAO ⁵ Hamachi ⁶ .	Máximos y mínimos. Inventarios, ventas, alquileres, ingresos.

² Formato de archivo de Microsoft Access 2007, denominado motor de base de datos de Access, específica para Office de Jet.

³ Sistema de base de datos orientada a objetos que han evolucionado para almacenar y procesar estructuras de datos OOP (Programación Orientada a Objetos).

⁴ Lenguaje de programación dirigido por eventos.

⁵ Modelo de acceso a base de datos de Microsoft que coexiste con Access.

Capítulo 1: Fundamentos teóricos sobre el desarrollo de los almacenes de datos

		Google ST. View	
Laboratorios Farmacéuticos	San Salvador	Microsoft Visual FoxPro 6.0 Microsoft Visual Basic PostgreSQL 7.0.	Promedios, máximos y mínimos. Existencias, salidas, despachos, movimiento de entradas y salidas, exportación, despacho y código a muestreo.
Fast Track	Estados Unidos	Cisco ⁷ Microsoft SQL Server 2012	Hereda los indicadores y perspectivas de análisis específicos de la empresa donde se desplace.
Ensayos Clínicos	Cuba	Oracle	Máximos y mínimos. Datos demográficos, seguridad, eficacia, respuesta.
CIMEX	Cuba	Oracle	Promedios, valores medios, máximos, mínimos, totales. Venta, inventarios, compra.
ONE	Cuba	Oracle	Valores medios, totales. Censos, encuestas, registros administrativos.

1.2.3 Valoración

Luego de haber realizado un análisis del uso de los DWH tanto en el ámbito internacional y nacional se llegó a la conclusión que hoy en día es de vital importancia el empleo de estos

⁶ Software que crea una red virtual que permite establecer una conexión segura y fiable entre 2 o más PC's conectadas a internet

⁷ Empresa global dedicada a la fabricación, ventas, mantenimiento y consultoría de equipos de telecomunicaciones tales como routers (enrutadores), switches (conmutadores), firewall (cortafuegos), así como productos de telefonía IP.

sistemas para el apoyo a la toma de decisiones en una organización donde la cantidad y diversidad de la información es alta. Dado los resultados obtenidos de su uso, se logra maximizar la eficiencia y mostrar de una forma agradable la información analizada, que puede estar dirigida a períodos o áreas determinadas de la institución. Independientemente que estos sistemas analizados centraron su atención al área del comercio y el marketing, presentan su información por indicadores como: totales, promedios, mínimo, máximo, tales indicadores pueden ser aplicados en el desarrollo del DWH para SIPAC. Es de vital importancia implementar un DWH que sea capaz de trabajar en conjunto y, por lo tanto, aumentar el valor operacional del SIPAC y hacer más fácil el acceso a una variedad de datos a los usuarios finales.

1.3 Procesos Básicos de un DWH

Independientemente de la metodología que se seleccione, el desarrollo de un DWH consta de dos etapas fundamentales para su realización, las cuales serán descritas a continuación.

1.3.1 Integración de datos

La integración de datos es el proceso que organiza el flujo de la información entre diferentes sistemas en una organización y aporta los métodos y herramientas necesarias para mover datos desde múltiples fuentes a un DWH, limpiarlos y cargarlos en otra base de datos. (12)

Una de las formas en que puede realizarse es mediante el proceso de Extracción, Transformación y Carga, el cual ha sido seleccionado para la solución a desarrollar.

Extracción, Transformación y Carga de los datos

El proceso de Extracción, Transformación y Carga (*Extract, Transform and Load - ETL*) constituye el cimiento del almacén de datos. El ETL permite extraer los datos desde los sistemas fuentes, garantiza la calidad de los mismos y hace cumplir las reglas del negocio (RN) definidas para ello. Luego hace posible que datos de distintas fuentes puedan ser usados juntos y por último entrega los datos en un formato listo para ser presentado, lo que posibilita a los desarrolladores implementar sus aplicaciones y a los usuarios finales tomar decisiones. (13)

Extracción: La extracción es el primer paso de la obtención de datos en el medio ambiente de almacenamiento de datos. Medios de extracción de leer y comprenderlos datos de origen, y la copia de las piezas que se necesitan para el área de preparación de datos para futuros trabajos. (14)

Transformación: Una vez que los datos son extraídos, el proceso de transformación se encarga de preparar los datos de la manera adecuada para integrarlos en el almacén. Hay distintos pasos de transformación, como la limpieza de los datos, selección de los campos necesarios para la carga del almacén, también se pueden combinar distintas fuentes de datos y realizar otras operaciones. (14)

Carga: Al final del proceso de transformación, los datos están en la forma para ser cargados en el almacén. Después de la carga inicial se procede a mantener el almacén actualizándolo periódicamente. (14)

1.3.2 Inteligencia de Negocios (BI)

La inteligencia de negocios (Business Intelligence - BI) es la habilidad para transformar los datos en información, de forma que se pueda optimizar el proceso de toma de decisiones en los negocios. Es un conjunto multifuncional de metodologías, procesos y tecnologías que brinda la posibilidad de generar estructuras de datos para realizar análisis de la información a partir de un conjunto de datos almacenados en uno o varios sistemas. (15)

En un sistema de BI se suelen encontrar los siguientes componentes o funcionalidades:

- **Cubos OLAP** (del inglés On Line Analytical Processing): es una herramienta de visualización multidimensional que permite analizar los datos provenientes de un DWH y obtener acceso a datos organizados y agregados de orígenes de datos empresariales. Esta herramienta soporta el análisis interactivo de la información de resumen, soportando muchas tareas de agrupación de datos.
- **Cuadro de Mando Integral (CMI)**: también conocido como Balanced Scorecard (BSC) o dashboard, es una herramienta de visualización de datos que muestra el estado actual de las métricas e indicadores claves del rendimiento de una empresa o área de negocio.
- **Reportes y consultas** (del inglés Queries and reports): generación de información con alto nivel de detalle orientado a un nivel gerencial que es utilizado para la toma de decisiones.

1.4 Procesamiento Analítico en Línea (OLAP)

Los sistemas OLAP se encargan del procesamiento analítico de los datos y están enfocados a apoyar la toma de decisiones en determinada entidad. OLAP trabaja, entre otros elementos,

con funciones de agrupación, entre las que se encuentran las ya conocidas COUNT, AVG, SUM, etc.; así como otras que han sido creadas específicamente para este análisis.

1.4.1 Implementaciones OLAP

Existen dos tendencias para la implementación física del modelo multidimensional: Multidimensional OLAP (MOLAP) y Relacional OLAP (ROLAP). A continuación se realiza una descripción de sus principales características.

Procesamiento Analítico Relacional en Línea (ROLAP por sus siglas en inglés): almacena los datos en un motor relacional de forma detallada en tablas normalizadas. Comúnmente se utilizan esquemas en estrella o copo de nieve para esta implementación, pero esta además permite trabajar sobre cualquier BD relacional, convirtiéndose en la principal ventaja de esta arquitectura que permite el análisis de cantidad de datos. (16) (17)

Procesamiento Analítico Multidimensional en Línea (MOLAP por sus siglas en inglés): almacena físicamente los datos en estructuras multidimensionales de modo que las representaciones externas e internas concuerden. Presenta algunas desventajas, partiendo de que la implementación no es adecuada cuando se necesita trabajar con cantidad de tablas de dimensiones y también que las herramientas diseñadas son restringidas y propietarias. (16) (17)

Para desarrollar la solución se utilizó ROLAP porque es el más apropiado para analizar volúmenes grandes de información. Además esta arquitectura propone acceder directamente al DWH implementada sobre una base de datos relacional y las operaciones sobre los datos se traducen a consultas SQL, manejadas por un gestor OLAP propio. Se utilizó como gestor de BD PostgreSQL, el cual solo soporta el almacenamiento relacional.

1.5 Metodologías para el desarrollo de Almacenes de Datos

Una metodología es aquella guía que se sigue a fin de realizar las acciones propias de una investigación. En términos más sencillos, se trata de la guía que va indicando qué hacer y cómo actuar cuando se quiere obtener algún tipo de investigación. Toda metodología debe referirse a un procedimiento ordenado que permita encaminar los pasos hacia la consecución de los objetivos propuestos dentro de la investigación. (18)

En la actualidad han surgido diferentes metodologías para el desarrollo e implementación de almacenes de datos que se enmarcan en dos enfoques principales bien definidos que han

servido de guía para su creación. Las mismas son la metodología de Kimball en honor a su creador Ralph Kimball (creador del enfoque multidimensional para el diseño de almacenes de datos) y la metodología de Inmon debido a su creador Bill H. Inmon (creador del término almacenes de datos y conocido como el padre de dicha disciplina).

En este epígrafe se realizó un análisis de algunas de estas metodologías estudiadas, con el objetivo de seleccionar la más adecuada para la arquitectura y desarrollo del DWH para SIPAC.

1.5.1 Metodología propuesta por William H. Inmon

Inmon defiende una metodología descendente (top-down) a la hora de diseñar un almacén de datos. En esta metodología los Mercados de Datos (DataMarts) se crearán después de haber terminado el DWH completo de la organización. Tiene como inconveniente que puede llevar una implementación lenta y costosa, debido a la cantidad de datos y su poca flexibilidad. Es recomendada cuando se hace difícil representar el modelo a través de dimensiones y la complejidad de la solución se hace demasiada grande. (19)

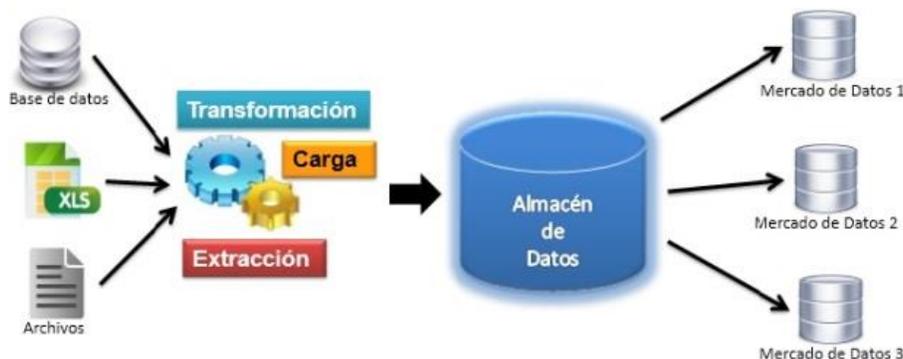


Figura 1. Arquitectura descendente

1.5.2 Metodología propuesta por Ralph Kimball

Kimball defiende una metodología ascendente (bottom-up) a la hora de diseñar un almacén de datos. Plantea que primero se deben crear Mercados de Datos (DataMarts) por cada departamento independiente orientados a temas relacionados con él y que la unión de todos ellos conformaría el almacén. Se basa en dividir el mundo de inteligencia de negocio entre los hechos y dimensiones, las cuales contienen toda la información cualitativa y cuantitativa de los indicadores, permitiendo una solución completa en un menor período de tiempo que el enfoque de Inmon. Presenta como problema que no explica detalladamente los elementos necesarios para la etapa de diseño, ni cómo se debe desarrollar y no tiene fase de prueba. (20)



Figura 2. Arquitectura ascendente

1.5.3 Metodología de Desarrollo de Almacenes de Datos de DATEC.

El Centro de Tecnologías de Gestión de Datos (DATEC) propone *La Metodología de Desarrollo de Almacenes de Datos* que toma como base la metodología de Kimball para definir los aspectos específicos del desarrollo de los DWH. La metodología propuesta por DATEC está dividida en ocho fases, las mismas generan un conjunto de artefactos que propician obtener un resultado físico del proyecto. Ellos contienen todo lo que se necesita para poder utilizar, mantener y seguir desarrollando el almacén. (21) A continuación se describen las fases de la metodología de DATEC que se utilizaron y los artefactos que generan:

- **Estudio preliminar y planeación:** se realiza un estudio de la entidad cliente donde se recoge en un diagnóstico integral todo lo referente para la construcción, condiciones existentes para el desarrollo y montaje de la misma. Además se llevan a cabo las tareas de planeación del proyecto, generando como artefactos el cronograma que contiene las actividades para llevar a cabo el desarrollo del almacén y el glosario de términos.
- **Requisitos:** se realiza el levantamiento de los requisitos de información para su posterior análisis y se determinan los requisitos funcionales y no funcionales de la solución. Genera como artefactos el diccionario de datos, perfilado de datos, especificación de caso de uso, descripción de requisitos, reglas del negocio y glosario de términos.
- **Arquitectura:** se definen las vistas arquitectónicas de la solución y aspectos como: los subsistemas y componentes, la seguridad, la comunicación y la tecnología a utilizar.
- **Diseño e Implementación:** se define el diseño de cada uno de los subsistemas (repositorio de datos, integración de datos y presentación de datos) y luego se lleva a cabo la implementación de cada uno de ellos. Genera como artefactos el código fuente

(script de BD, transformación, trabajos, esquemas de cubos OLAP, reportes OLAP), reportes candidatos y manual de usuario.

- **Prueba:** se realizan las pruebas que validan la calidad del producto (Pruebas de Unidad llevadas, Pruebas de Integración y Pruebas de Aceptación con el cliente final). En esta fase no es la única donde se realizan pruebas, ya que en todas las fases hay aseguramiento de la calidad, generando como artefacto la carta de aceptación.

Se decidió aplicar este modelo debido a que el ciclo de vida de la metodología está organizado por fases y algunas de ellas podrán ser implementadas de forma paralela, lo que proporciona que dicha metodología tenga un desarrollo ágil y ajustado al modelo de Línea de Producción de Software (LPS). Además el desarrollo del DWH a partir de la construcción de mercados de datos permitió ir avanzando de manera evolutiva, obteniendo resultados parciales que son validados por los clientes hasta obtener la solución final.

1.6 Herramientas seleccionadas para la solución

A continuación se hace referencia a las herramientas elegidas para el desarrollo de la solución, teniendo en cuenta la arquitectura definida por DATEC.

1.6.1 Herramienta de modelado.

La industria del software evoluciona constantemente y esto genera la necesidad de utilizar herramientas para el modelado de sistemas que le faciliten el trabajo a los analistas y diseñadores de software. Como herramienta de modelado se utilizará *Visual Paradigm for UML 8.0 Enterprise Edition*.

Visual Paradigm for UML 8.0 Enterprise Edition

Es una herramienta CASE (del inglés Computer Aided Software Engineering) que soporta el modelado mediante UML⁸ y proporciona asistencia a los analistas, ingenieros de software y desarrolladores durante todos los pasos de ciclo de vida de desarrollo de un software. En general constituye una aplicación muy completa y con abundante facilidades, su utilización contribuye a la creación rápida de aplicaciones de calidad y con un costo muy económico. (22)

1.6.2 Sistema Gestor de Base de Datos

Un Sistema de Gestión de Bases de Datos (SGBD) es un conjunto de programas que permiten el almacenamiento, modificación y extracción de la información en una base de datos, además

⁸UML: lenguaje gráfico para visualizar, especificar, construir y documentar un sistema.

de proporcionar herramientas para añadir, borrar, modificar y analizar los datos (23). Para la solución se hizo uso de PostgreSQL 9.1.

PostgreSQL 9.1

Constituye el SGBD objeto-relacional de código abierto más desarrollado del mundo y opera bajo licencia BSD⁹, es multiplataforma, ofrece una documentación bien organizada, pública y libre, presenta conectores de datos foráneos que permite añadir y consultar fuentes de datos externas desde PostgreSQL. Alcanza un conjunto de características tales como: herencia, tipos de datos, funciones, restricciones, disparadores, reglas e integridad transaccional. (24)

PgAdminIII 1.14.0

Constituye la más completa y popular herramienta de código abierto para la administración de BD PostgreSQL. Hace más sencilla la administración y está diseñado para satisfacer las necesidades de todos los usuarios, teniendo en cuenta la realización de consultas SQL desde la más sencilla, hasta el desarrollo de BD de alta complejidad. Presenta una interfaz administrativa gráfica y es una herramienta de consulta SQL. (25)

1.6.3 Herramientas para la integración de datos

Las herramientas de integración de datos brindan el soporte necesario para garantizar las actividades de extracción de los datos desde las diversas fuentes, transformación de ellos mediante un conjunto de procesos que suceden a partir de la obtención de datos optimizados para el análisis y la presentación de reportes, y carga de los datos en el almacén. (26) Para la realización de la solución se utilizarán el *Pentaho Data Integration (PDI) 4.2.1* y el *DataCleaner 1.5.4*.

Pentaho Data Integration (PDI) 4.2.1

Herramienta que incluye componentes que facilitan el modelado y la ejecución de transformaciones sobre flujos de datos. Entre las principales ventajas del PDI sobresale que es multiplataforma, además soporta un entorno de ejecución de Java 1.4 o superior, utiliza *JavaScript* en la realización de transformaciones avanzadas consistentes en cálculos y fórmula, mediante una interfaz empotrada. (27)

⁹*Berkeley Software Distribution (BSD)*, sólo requiere que el código fuente licenciado mantenga la información de derechos de autor y licenciamiento.

Su diseño es orientado a metadatos, haciendo uso de la herramienta *Spoon* para el diseño gráfico de trabajos y transformaciones utilizando componentes que se entrelazan entre sí mediante saltos, estos últimos establecen el flujo de datos entre los diferentes componentes. (28)

DataCleaner 1.5.4

Herramienta que permite la valoración del nivel de calidad de los datos comprendidos en el sistema de información preparando estos para futuras transformaciones en el proceso ETL. La misma requiere un entorno de ejecución de Java 5.0 o una versión superior y de drivers JDBC¹⁰. Es de fácil utilización, genera sofisticados gráficos e informes que permiten a los usuarios identificar y analizar la estructura del origen de datos y combinar resultados y gráficos, estableciendo vistas fáciles de interpretar para evaluar la calidad de los datos. (29)

1.6.4 Herramientas para inteligencia de negocio

Implementar herramientas de BI dentro de la organización permite soportar las decisiones que se toman y al nivel interno ayuda en la gestión del personal. Los sistemas BI son un tipo de software en crecimiento dentro del mercado mundial, ellos brindan una amplia gama de capacidades funcionales para realizar todos los procesos de esta etapa. (30) Las herramientas seleccionadas para de la solución son *Mondrian OLAP server 3.2*, *Pentaho Schema Workbench 3.2.1* y *Pentaho Business Intelligence 4.8.0*.

Mondrian OLAP server 3.2

Motor OLAP desarrollado en Java que gestiona la comunicación entre una aplicación OLAP y las fuentes de datos. La facilidad para el análisis de grandes volúmenes de información que se encuentran en la BD es una de sus características más significativas. Esta herramienta permite la realización de consultas al DWH y brinda la posibilidad de mostrar los resultados a través de un navegador de forma tal que el usuario pueda realizar actividades típicas de navegación sobre cubos OLAP. Se ha convertido en una de las aplicaciones más importantes de la Suite de Pentaho BI.

Pentaho BI Server 4.8.0

¹⁰JDBC: *Java Database Connectivity* es una aplicación para trabajar con BD desde Java.

La plataforma *Pentaho BI* proporciona la arquitectura y la infraestructura necesaria para crear soluciones de BI. Dentro de esta plataforma la aplicación más conocida es el BI server, diseñado para integrarse fácilmente en cualquier proceso de negocio, cuyo funcionamiento web se basa en un sistema de gestión de reportes, servidores de integración de aplicaciones y en el motor de flujo de trabajo ligero. (31) Este servidor cuenta con una interfaz de usuario donde se encuentran disponibles todos los reportes, vistas OLAP y cuadros de mando; además tiene una consola de administración que permite gestionar y supervisar la aplicación, los usuarios de la misma, los reportes que revisa cada usuario, cuándo se accedió a cada reporte, entre otras funciones.

Pentaho Schema Workbench 3.2.1

Es una herramienta para el desarrollo y prueba de cubos OLAP de forma visual. La definición del XML no es extremadamente compleja, pero en la práctica resulta engorroso recordar cada uno de los elementos junto a sus atributos y sub-elementos tal y como se encuentran en el almacén. Con esta aplicación, se puede configurar una conexión con el modelo físico, para luego elaborar el esquema lógico de manera simple y efectiva. Para ello la herramienta ofrece un editor de esquemas con la fuente de datos subyacente para su validación. (31)

1.6.5 Herramientas para soporte de aplicación.

Las herramientas de soporte web se utilizan para cargar archivos y brindar contenido estático en el navegador, realizando el intercambio mediante el protocolo HTTP (*Hypertext Transfer Protocol*).

Apache Tomcat 6.0

Apache Tomcat es un servidor web de software libre bajo licencia Apache en su versión 2.0, gestionado por la fundación Apache Tomcat, es el más utilizado a la hora de trabajar con Java en entornos web. Fue implementado en Java, esto propicia que trabaje en cualquier sistema operativo que tenga instalado *Java Development Kit* (JDK) (26). Fue escogida esta herramienta fundamentalmente porque la plataforma de *Pentaho BI* a utilizar la instala por defecto.

1.6.6 Herramienta para el control de versiones

Los controles de versiones son un conjunto de métodos y herramientas para controlar los aspectos relacionados con los cambios de un archivo en el tiempo. Es importante para el

Capítulo 1: Fundamentos teóricos sobre el desarrollo de los almacenes de datos

desarrollo del sistema emplear un correcto control de versiones. Para esta actividad en la realización de la solución se seleccionó *Subversion 1.5.4*.

Subversion (SVN) 1.5.4

Subversion soporta conexiones simultáneas al repositorio, permitiendo también que cada uno de los usuarios pueda leer o escribir en esos archivos. Es un sistema de control de versiones libre y de código abierto. Tiene un repositorio como elemento fundamental que constituye un almacén central de datos. (32)

1.7 Conclusiones del capítulo

Luego de culminar el capítulo y de realizar un estudio preliminar sobre los elementos básicos de los almacenes de datos se concluyó que:

- Se dedujo la necesidad de implementar un DWH para SIPAC partiendo de los principales conceptos relacionados con el tema.
- Se escogió la metodología propuesta por DATEC para el desarrollo de almacenes de datos debido a que comprende todas las fases de la construcción de un DWH y las tecnologías y herramientas que propone, pues ofrecen numerosas ventajas que facilitan su desarrollo.

CAPITULO 2: ANÁLISIS Y DISEÑO DEL ALMACÉN DE DATOS PARA EL SISTEMA DE PLANIFICACIÓN DE ACTIVIDADES (SIPAC).

En el presente capítulo se realizó el análisis y diseño de la propuesta de solución, mostrándose los artefactos generados. Para un mejor entendimiento del negocio se especificaron las características del mismo, definiéndose las necesidades de información, así como las reglas del negocio, casos de uso y la arquitectura del sistema. Se realizó el diseño del subsistema de almacenamiento a partir de las dimensiones, hechos y medidas identificadas, construyendo la matriz multidimensional y posteriormente el modelo de datos. También se llevó a cabo el diseño del subsistema de integración mediante el perfilado de los datos y diseño de las transformaciones. Finalmente se diseña el subsistema de visualización donde se define la arquitectura de información y el diseño de los reportes candidatos.

2.1 Estudio preliminar del negocio

La planificación estratégica y operativa constituye un nuevo modelo de planificación que tiene como principal función planificar y organizar los procesos de: establecimiento, aprobación y puntualización de los planes, sus sistemas de control y factores que influyen en dichos planes. Este nuevo modelo se va a ocupar de establecer resultados finales hacia los cuales se dirigen las actividades organizacionales e individuales, en aras de cumplir los objetivos económicos y sociales que demanda el desarrollo integral de la sociedad cubana. (33)

En función de esto se desarrolló el Sistema de Planificación de Actividades (SIPAC) en la Universidad de las Ciencias Informáticas (UCI), el cual se encarga de informatizar el proceso de planificación estratégica y operativa adoptado por las entidades cubanas. Se basa en la Instrucción no.1 del Presidente de los Consejos de Estado y de Ministros para la Planificación de los objetivos y actividades en los órganos, Organismos de la Administración Central de Estado, entidades nacionales y Administraciones locales del Poder Popular.

Dicho sistema está destinado a facilitar la gestión de las actividades a todos los niveles organizacionales, permite interrelacionar objetivos de trabajo y actividades en tiempo real; garantizando el seguimiento del desarrollo y cumplimiento de los objetivos y tareas principales en las entidades. Informatiza los procesos de Ejecución y Control de la Planeación Estratégica (definición de los objetivos a largo plazo y estrategias para alcanzarlos) y Operativa (puntualización de las actividades que debe efectuar cada individuo a corto plazo). Es

Capítulo 2: Análisis y diseño del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

catalogado como una solución integral para la gestión de elementos de la planeación estratégica y operativa basada en actividades, objetivos, y planes, diseñada sobre las bases de la compartimentación de la información.

2.2 Necesidades de los usuarios

Para el buen desarrollo del análisis en el proceso de negocio es preciso conocer qué es lo que necesitan los usuarios. La implicación de los mismos durante el ciclo de vida del producto es de gran importancia, de ahí proviene la posibilidad de que los resultados sean satisfactorios o insatisfactorios en correspondencia a sus necesidades. Para el caso de los DWH a los requisitos funcionales (RF) y a los requisitos no funcionales (RNF) son adicionados las necesidades de información, teniendo en cuenta que las mismas constituyen la base para un correcto diseño del almacén de datos.

2.2.1 Requisitos de información

Los requisitos de información son las principales funcionalidades que el sistema debe tener disponible a la hora de realizar análisis sobre los datos. Constituyen la entrada fundamental para el proceso de Inteligencia de Negocio (BI) y las reglas del negocio con los elementos disponibles en las fuentes. En la investigación fueron identificados 10 requisitos de información (RI) pertenecientes a cada área de análisis:

A continuación se muestran los RI identificados en la investigación:

Actividades

RI 1: Total de actividades por estado cumplimiento, plan, plan_arc, objetivos, grupo de usuarios, FIP, categoría de la actividad, tipo de actividad, participantes, dirigentes, involucrados, lugar, principal, extraplan, porcentaje de cumplimiento y tiempo.

RI 2: Porcentaje de actividades por estado cumplimiento, plan, plan_arc, objetivos, grupo de usuarios, FIP, categoría de la actividad, tipo de actividad, participantes, dirigentes, involucrados, lugar, principal, extraplan, porcentaje de cumplimiento y tiempo.

Objetivos

RI 3: Total de objetivos por plan_arc, FIP, estado cumplimiento, grupo de usuarios, involucrados, porcentaje de cumplimiento y tiempo.

Capítulo 2: Análisis y diseño del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

RI 4: Porcentaje de objetivos por plan_arc, FIP, estado cumplimiento, grupo de usuarios, involucrados, porcentaje de cumplimiento y tiempo.

Plan

RI 5: Total de planes por categoría del plan, estado de cumplimiento, porcentaje de cumplimiento y tiempo.

RI 6: Porcentaje de planes por categoría del plan, estado de cumplimiento, porcentaje de cumplimiento y tiempo.

Áreas de Resultados Clave (ARC)

RI 7: Total de ARC por capítulos, plan, involucrados, grupo de usuarios y tiempo.

RI 8: Porcentaje de ARC por capítulos, plan, involucrados, grupo de usuarios y tiempo.

Factores que Influyen en el Plan (FIP)

RI 9: Total de FIP por nivel de acceso al FIP, tipo de FIP, involucrados, grupo de usuarios, participantes, dirigentes, emite y tiempo.

RI 10: Porcentaje de FIP por nivel de acceso al FIP, tipo de FIP, involucrados, grupo de usuarios, participantes, dirigentes, emite y tiempo.

2.2.2 Requisitos funcionales

Los RF especifican en un nivel adecuado de detalle lo que debe hacer o no una organización, para lo cual se es necesario expresarlos en lenguaje natural y deben estar individualmente centrados en ese negocio. (34) Los mismos deben estar orientados a las necesidades de los usuarios finales y garantizados por la plataforma de visualización BI server 4.8.0, la cual brinda muchas de las funcionalidades identificadas como RF. Algunas de esas funcionalidades son Gestionar usuario, Gestionar rol y Gestionar reporte, y de esta manera dar respuesta a los RI. A continuación se muestran los RF identificados para la presente investigación:

RF 1. Autenticar usuario.

RF 2. Adicionar usuario.

RF 3. Modificar usuario.

RF 4. Eliminar usuario.

RF 5. Visualizar usuario.

RF 6. Adicionar rol.

Capítulo 2: Análisis y diseño del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

- RF 7. Modificar rol.
- RF 8. Eliminar rol.
- RF 9. Visualizar rol.
- RF 10. Adicionar reporte.
- RF 11. Modificar reporte.
- RF 12. Eliminar reporte.
- RF 13. Visualizar reporte.
- RF 14. Extraer datos de la fuente.
- RF 15. Personalizar reporte.
- RF 16. Exportar a otro formato.
- RF 17. Mostrar gráfico.
- RF 18. Configurar gráfico.
- RF 19. Mostrar propiedades.
- RF 20. Suprimir filas y columnas vacías.
- RF 21. Realizar transformación de los datos.
- RF 22. Cargar los datos al almacén de datos del Sistema de Planificación de Actividades (SIPAC).

2.2.3 Requisitos no funcionales

Los Requisitos no funcionales (RNF) son restricciones de los servicios o funciones ofrecidas por el sistema. Incluyen restricciones de tiempo sobre el proceso de desarrollo y estándares. Los requisitos no funcionales a menudo se aplican al sistema en su totalidad. (35) Es decir son aquellos que no tienen que ver directamente con las funcionalidades del sistema, pero garantiza su correcto funcionamiento. Para la investigación se identificaron los siguientes RNF:

1. Usabilidad
 - 1.1 El sistema podrá ser usado por personas con conocimientos básicos en el manejo de computadoras.
 - 1.2 Agilizar el acceso a los reportes del almacén de datos mediante la distribución de la información por áreas de análisis.
2. Seguridad
 - 2.1 La información que se maneje en el sistema está protegida de acceso no autorizado, a partir de los diferentes roles de los usuarios que empleen el sistema.
 - 2.2 Si se desea eliminar o modificar un elemento de la base de datos que está siendo utilizado por otro elemento que depende de él, el sistema no permitirá que este elemento sea eliminado.
3. Confiabilidad
 - 3.1 Garantizar la persistencia de la información. Se debe realizar un respaldo total de los datos del almacén con una frecuencia mensual.
4. Restricciones de diseño.

Capítulo 2: Análisis y diseño del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

- 4.1 Lograr la homogeneidad de la estructura de los elementos definidos en el almacén.
Las estructuras del almacén de datos deben tener un nombre estándar teniendo en cuenta el tipo de estructura que sea.

2.3 Reglas del negocio

Una regla de negocio es una declaración que define o limita algún aspecto del negocio. Su intención es valorar la estructura del negocio, o controlar, o influenciar el comportamiento del mismo. (36) A continuación se muestran las reglas del negocio que fueron identificadas en la investigación:

RN 1. Una vez que los datos estén cargados en el almacén de datos, no pueden existir campos nulos.

RN 2. Los elementos de la planificación que tengan los campos “dirigente y participante” vacíos se reemplazará por “no definido”.

RN 3. Las actividades que contengan el campo “lugar” vacío se reemplazará por “no definido”.

RN 4. El campo “esprincipal” de la tabla dat_actividades estará definido de la siguiente manera: Sí (cuando su valor sea 1) y No (cuando su valor sea 0).

RN 5. El campo “extraplan” de la tabla dat_actividades estará definido de la siguiente manera: Sí (cuando su valor sea 1) y No (cuando su valor sea 0).

RN 6. La tabla dat_fip que tenga el campo “emite” vacío se reemplazará por “No definido”.

RN 7. Los valores que indiquen total tienen que ser mayores o iguales que cero.

RN 8. Los valores que indiquen por ciento tienen que estar en un rango de cero a 100.

RN 9. El campo “dnelemento” de la tabla dat_elementos y “dengrupo” de la tabla dat_grupo se renombrarán a denominación.

2.4 Diagrama de Caso de Uso del sistema

Un diagrama de CUS es modelado por la interacción entre los CU y los actores del sistema. Para su elaboración se tomaron como base los RI y RF agrupados e identificados anteriormente en 12 CU, 5 de ellos son CU de información agrupados por las áreas de análisis definidas con anterioridad:

Capítulo 2: Análisis y diseño del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

CU Visualizar información de actividades.

CU Visualizar información de objetivos.

CU Visualizar información de plan.

CU Realizar extracción de los datos.

CU Gestionar rol.

CU Autenticar usuario.

CU Visualizar información de ARC.

CU Visualizar información de FIP.

CU Realizar operaciones sobre reportes.

CU Gestionar usuario.

CU Gestionar reporte.

CU Realizar transformación y carga de los datos.

Tres de ellos son agrupados por el patrón CRUD que plantea agrupar las funcionalidades crear, leer, modificar y eliminar tales como: CU Gestionar usuario, CU Gestionar rol y CU Gestionar reporte. Otro patrón de CU que se refleja en el diagrama es concordancia por adición, el cual se pone de manifiesto en la relación de extensión que existe entre los CU de información y el CU *Realizar operaciones sobre reportes*. A continuación se presenta el DCUS del DWH para SIPAC:

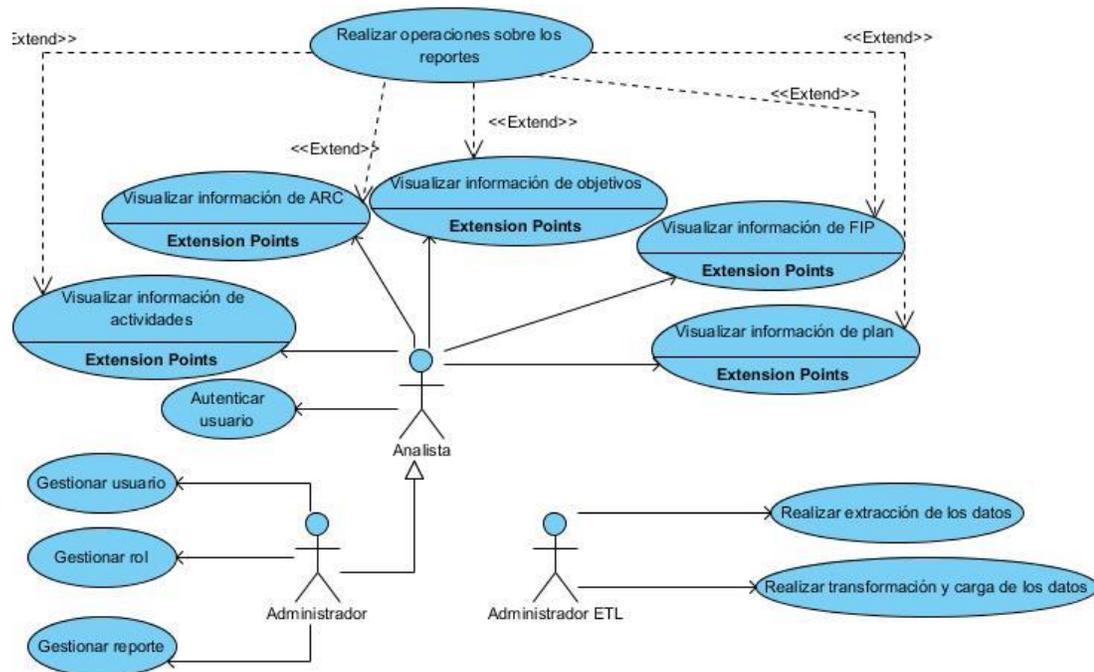


Figura 3. Diagrama de Caso de Uso del Sistema

2.4.1 Actores del sistema

Los actores del sistema son las personas que interactúan con la aplicación y se benefician de ella. Los actores del Almacén de Datos para SIPAC son los siguientes:

Capítulo 2: Análisis y diseño del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

Tabla 2. Actores del sistema

Actor	Responsabilidad en el sistema
Administrador	Realiza todas las actividades de administración del DWH. Además se puede comportar como el actor analista debido a la relación de generalización/especialización que ellos tienen, evidenciándose de esta manera el patrón múltiples actores rol común.
Administrador ETL	Realiza las actividades de extracción de los datos de las fuentes, así como la transformación y la carga de estos hacia el DWH.
Analista	Consulta la información de los reportes.

2.5 Arquitectura del sistema

Una Arquitectura DWH (Datawarehouse Architecture - DWHA) es una forma de representar la estructura total de los datos, comunicación, procesamiento y presentación, que existe para los usuarios finales que disponen de una computadora dentro de la empresa. (37)

Para tener una visión general del sistema se explica a continuación la arquitectura del mismo, detallando los tres subsistemas que la conforman: integración de datos, almacenamiento y visualización.

El subsistema de integración es el encargado de realizar el proceso de extracción, transformación, integración, limpieza y estandarización de la información proveniente de la fuente de datos, en este caso de la base de datos SIPAC, para su posterior carga hacia la fuente destino. Todo este proceso se realiza mediante las herramientas DataCleaner 1.5.4 y Pentaho Data Integration 4.2.1.

El subsistema de almacenamiento es el que contiene las tablas de hechos identificadas en el negocio, las dimensiones seleccionadas para la solución y las relaciones que existen entre ellas. Este subsistema se nutre de los datos manipulados por el subsistema de integración,

Capítulo 2: Análisis y diseño del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

mediante el protocolo de comunicación TCP/IP, quedando implementado el almacén de datos, apoyado por el SGBD PostgreSQL 9.1 y administrado por la herramienta pgAdmin III 1.14.0.

El subsistema de visualización es el encargado de organizar los reportes por áreas de análisis de los datos almacenados en el subsistema de almacenamiento, utilizando el protocolo TCP/IP, facilitando al usuario una búsqueda rápida de información, a través de la herramienta Pentaho Business Intelligence 4.8.0.

Existe un cliente para la administración de cada uno de los subsistemas el cual accede mediante el protocolo TCP/IP y un cliente para realizar operaciones sobre los reportes encontrados en el subsistema de visualización, mediante el protocolo HTTP.

A continuación se muestra la figura correspondiente a la arquitectura propia del sistema:

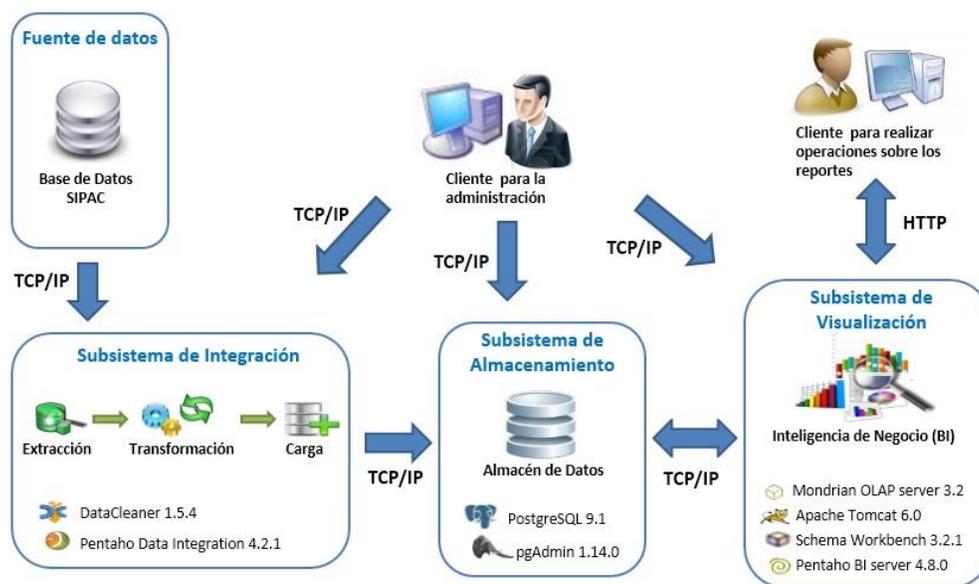


Figura 4. Arquitectura de la solución

2.6 Modelo Multidimensional

La tecnología de los almacenes de datos debido a su orientación analítica, impone un procesamiento y pensamiento distinto, la cual se sustenta por un modelo de bases de datos propio, conocido como Modelo Multidimensional, el cual busca ofrecer al usuario su visión respecto de la operación del negocio. El Modelo Multidimensional es una técnica para modelar bases de datos simples y entendibles al usuario final, con la idea fundamental de visualizar fácilmente la relación que existe entre los distintos componentes del modelo. (38)

2.6.1 Características

En general, la estructura básica de un DWH para el modelo multidimensional está definida por dos elementos: esquemas y tablas. En el modelo multidimensional la información está estructurada en dos tipos básicos de tablas, los hechos y dimensiones. Esta estructura de almacenamiento tiene forma de cubo multidimensional donde las aristas representan las dimensiones y las intersecciones entre estas aristas son los hechos. Mientras mayor sea el número de dimensiones vinculadas a un hecho, mayor es el nivel de detalle del mismo.

Cubo: es una estructura multidimensional que contiene información con fines analíticos, sus componentes principales son las dimensiones y las medidas. Las celdas del cubo se definen por la intersección de miembros de dimensión y contienen los valores agregados de las medidas en esa intersección concreta. (39)

Tabla de hechos: es la tabla central en un esquema dimensional. Es en ella donde se almacenan las mediciones numéricas del negocio. Cada medida es tomada de la intersección de las dimensiones que la definen. Se considera el evento específico que constituye la unidad fundamental de análisis de datos, para la toma de decisiones. (38)

Tablas de dimensiones: es una entidad de negocio respecto a la cual se deben calcular las métricas. Estas tablas son las que se conectan y alimentan a la tabla de hechos. En ella existe el detalle de los valores de la dimensión respectiva. (38)

Medidas: son aquellas características del negocio que pueden ser cuantificadas y son seleccionadas para el análisis. Se corresponden con los datos numéricos. Los valores que toman son el resultado de las diferentes combinaciones posibles de los miembros de las dimensiones sobre las que se definen. (40)

Otros de los conceptos fundamentales en el modelo multidimensional son las jerarquías y la granularidad. Las jerarquías representan una relación lógica entre los datos de una dimensión e implica una organización por niveles dentro de la misma y la granularidad representa el nivel de detalle en que se almacena la información. (38) La granularidad del DWH para SIPAC está dada por la dimensión Actividad y tiene como valor 7. Por otra parte, la colección de tablas en el almacén se conoce como esquema. Los esquemas se pueden agrupar dentro de las tres categorías básicas: esquemas estrellas, esquemas copo de nieve y esquema constelación de hechos.

2.6.2 Esquema

Para la solución se empleó el modelo multidimensional, basándose en el esquema constelación de hechos, debido a que se adaptó mejor a los requerimientos y necesidades del usuario. Este modelo está compuesto por una serie de esquemas en estrella, formado por una tabla de hechos principal y una o más tablas de hechos auxiliares, posibilitando que sea fácilmente modificable de acuerdo a nuevas características o necesidades que surjan. Se está en presencia de una constelación de hechos cuando múltiples tablas de hechos comparten las mismas dimensiones. (41) A continuación se muestra una representación del mismo:

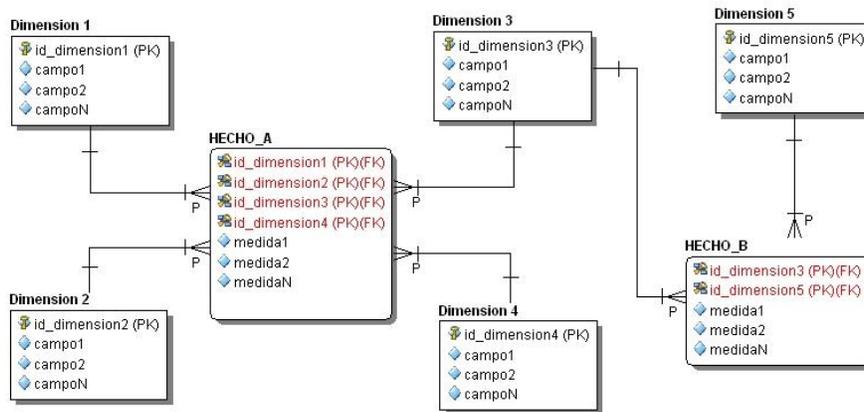


Figura 5. Esquema constelación de hechos

2.7 Diseño del almacén de datos

Se realizó el diseño del subsistema de almacenamiento donde a partir de los resultados de la fase de requisitos se definieron las estructuras dimensionales del almacén de datos (hechos y dimensiones) para la creación de la base de datos dimensional. Luego en el subsistema de integración se definieron las estrategias que deben seguirse para el desarrollo de los procesos de integración de datos, donde se tomaron en cuenta los resultados obtenidos durante el estudio de las fuentes y los requisitos funcionales identificados. Finalmente, en el diseño del subsistema de visualización se diseñaron los cubos OLAP para el análisis multidimensional de la información y se definieron los reportes que se implementados como parte de la solución.

2.7.1 Diseño del subsistema de almacenamiento

En el diseño del subsistema de almacenamiento se identificaron 5 hechos, 18 dimensiones y 10 medidas a partir de los requisitos de información extraídos de las necesidades del cliente. Se

Capítulo 2: Análisis y diseño del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

procedió luego a la creación de la matriz bus y posteriormente al modelo de datos de la solución.

Cada dimensión presenta un conjunto de atributos organizados en jerarquías, para un análisis más detallado a continuación se presenta la descripción de las 18 dimensiones identificadas con sus respectivos niveles jerárquicos:

Tabla 3. Descripción de las dimensiones

Dimensión	Jerarquía	Descripción
dim_actividad	denominacion_actividad lugar principal extraplan	Indica la denominación de cada actividad y su código, además del lugar donde se va a ejecutar la actividad, si es principal o no y si es extraplan o no.
dim_categoria_actividad	categoria_actividad	Registra la categoría de la actividad planificada y su código.
dim_tipo_actividad	denominacion_tipo_actividad	Contiene el tipo de actividad planificada y su código.
dim_dirigente	nombre_dirigente	Indica el nombre del dirigente en una actividad planificada o en un FIP, además de contar con su respectivo código.
dim_participante	nombre_participante	Indica el nombre del participante en una actividad o en un FIP, además de contar con su respectivo código.
dim_tipo_fip	denominacion_tipo_fip	Contiene el tipo de FIP y su código.
dim_objetivo	denominacion_objetivo	Indica el objetivo de cada actividad y su código.
dim_nivel_acceso_fip	denominacion_nivel_acceso_fip	Contiene el nivel de acceso de FIP y su código.
dim_cumplimiento	estado_cumplimiento	Contiene estado de cumplimiento

Capítulo 2: Análisis y diseño del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

	porciento_cumplimiento	y el porciento de cumplimiento de las actividades, objetivos y planes.
dim_plan	denominacion_plan	Indica el plan trazado en la planificación relacionado con los objetivos, además de contar con su respectivo código.
dim_involucrado	denominacion_involucrado	Registra los involucrados relacionados en las actividades, objetivos, ARC y FIP, además de contar con su respectivo código.
dim_tipo_involucrado	tipo_involucrado	Contiene el tipo de involucrado (si es un grupo o un usuario en específico) relacionados en las actividades, objetivos, ARC y FIP, además de contar con su respectivo código.
dim_tiempo_mes	tiempo	Contiene un conjunto de atributos relacionados con la fecha en la cual se realiza algún elemento de la planificación.
dim_fip	denominación_fip	Indica el FIP relacionado en las actividades y objetivos, además de contar con su respectivo código.
dim_arc	denominacion_arc	Indica las ARC relacionadas con los objetivos y las actividades, además de contar con su respectivo código.
dim_emite	emite	Registra el nombre del jefe o la entidad que emite la orden del FIP, además de contar con su

Capítulo 2: Análisis y diseño del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

		respectivo código.
dim_categoria_plan	categoria_plan	Registra la categoría del plan planificado y su código.
dim_capitulo	denominacion_capitulo	Registra el nombre del capítulo relacionado con las ARC y su código.

A continuación se presenta la descripción de los 5 hechos identificados:

Tabla 4. Descripción de los hechos

Hechos	Descripción
hech_actividad	Es objeto de análisis del negocio. Comprende todo lo relacionado con las actividades planificadas.
hech_fip	Es objeto de análisis del negocio. Comprende todo lo relacionado con los FIP.
hech_objetivo	Es objeto de análisis del negocio. Comprende todo lo relacionado con los objetivos planificados.
hech_arc	Es objeto de análisis del negocio. Comprende todo lo relacionado con las ARC.
hech_plan	Es objeto de análisis del negocio. Comprende todo lo relacionado con los planes.

En la tabla 3 se detallan las 10 medidas correspondientes a los hechos anteriormente descritos. El tipo de dato de las mismas es *integer*.

Tabla 5. Descripción de las medidas

Medidas	Descripción
total_actividades	Muestra el total de actividades planificadas.
porciento_actividades	Muestra el porciento de actividades planificadas.
total_fip	Muestra el total de FIP.
porciento_fip	Muestra el porciento de FIP.

Capítulo 2: Análisis y diseño del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

total_objetivos	Muestra el total de objetivos planificados.
porciento_objetivos	Muestra el porciento de los objetivos planificados.
total_planes	Muestra el total de planes.
porciento_planes	Muestra el porciento de planes.
total_arc	Muestra el total de ARC.
porciento_arc	Muestra el porciento de ARC.

Matriz dimensional

La matriz bus o matriz dimensional tiene como objetivo mostrar la relación que existe entre los hechos y las dimensiones y de esta manera garantizar que no ocurra un solapamiento entre las tablas de hechos, es decir, que no existan dos o más hechos que compartan exactamente las mismas dimensiones. Además permite validar el correcto diseño del modelo de datos y determinar el impacto que provocaría un cambio durante el desarrollo del sistema.

Las columnas de la matriz representan las dimensiones y filas de los hechos identificados. Las celdas marcadas con X indican que la fila del hecho está relacionada con la columna de la dimensión. Al realizar el análisis se arrojaron como resultados que las dimensiones dim_dirigente, dim_participante, dim_objetivo, dim_cumplimiento, dim_plan, dim_involucrado, dim_tipo_involucrado, dim_tiempo_mes, dim_fip y dim_arc son compartidas por más de un hecho. Además se evidencia que no existen hechos solapados, por lo que se puede afirmar que el diseño del modelo de datos es correcto. A continuación se muestra la matriz dimensional correspondiente al Almacén de datos para el Sistema de Planificación de Actividades.

hech/dim	dim_1	dim_2	dim_3	dim_4	dim_5	dim_6	dim_7	dim_8	dim_9	dim_10	dim_11	dim_12	dim_13	dim_14	dim_15	dim_16	dim_17	dim_18
hech_1	x	x	x	x	x		x		x	x	x	x	x	x	x			
hech_2				x	x	x		x			x	x	x	x		x		
hech_3							x		x	x	x	x	x	x	x			
hech_4								x	x			x					x	
hech_5									x	x	x	x		x				x

hech_1: hech_actividad
hech_2: hech_fip
hech_3: hech_objetivo
hech_4: hech_plan
hech_5: hech_arc

dim_1: dim_categoria_actividad
dim_2: dim_tipo_actividad
dim_3: dim_actividad
dim_4: dim_dirigente
dim_5: dim_participante
dim_6: dim_tipo_fip

dim_7: dim_objetivo
dim_8: dim_nivel_acceso_fip
dim_9: dim_cumplimiento
dim_10: dim_plan
dim_11: dim_involucrado
dim_12: dim_tipo_involucrado

dim_13: dim_tiempo_mes
dim_14: dim_fip
dim_15: dim_arc
dim_16: dim_emite
dim_17: dim_categoria_plan
dim_18: dim_capitulo

Figura 6. Matriz dimensional

Capítulo 2: Análisis y diseño del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

En la realización del perfilado de los datos al sistema fuente del SIPAC se generaron los reportes estándares de medidas, análisis de cadena y análisis numérico arrojando los siguientes resultados (ver Anexos 3, 4 y 5):

- Las fechas se encontraban con tipo de datos *DATE* y se transformaron al tipo de dato *VARCHAR* para poder hacer más fácil su uso y poder analizar las actividades por año, mes, semestre y trimestre.
- Existían varias columnas de tipo *VARCHAR* admitiendo como tamaño máximo 255 caracteres, las cuales se transformaron a tipo de dato *TEXT* permitiendo que no existan límites en los caracteres a almacenar.
- Existía presencia de valores nulos y vacíos en la mayoría de las tablas analizadas.

A estos resultados se le dio tratamiento mediante el cumplimiento de las reglas de negocio y transformación.

Diseño del proceso de integración

Después del análisis de los datos y conociendo la estructura de la base de datos se procedió a la realización del proceso de integración de datos. A continuación se presenta el diseño general correspondiente a dicho proceso:

Capítulo 2: Análisis y diseño del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

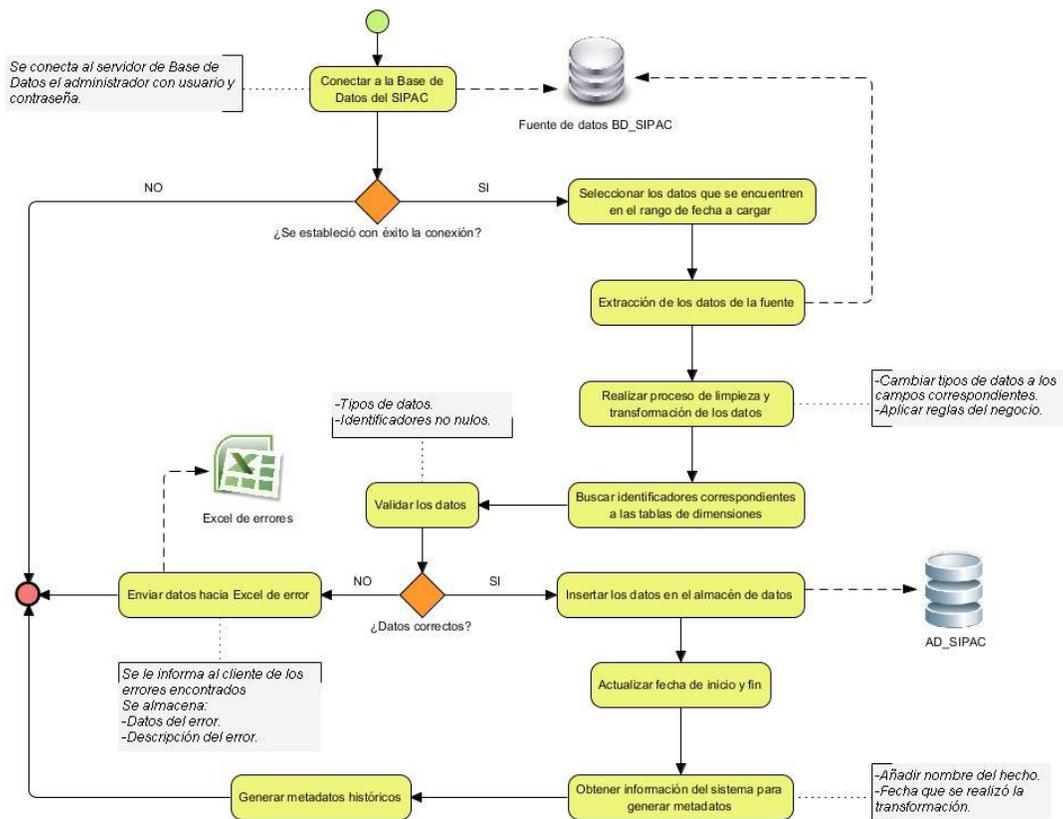


Figura 8. Diseño general del proceso de integración del DWH para SIPAC.

2.7.3 Diseño del subsistema de visualización

En el diseño del subsistema de visualización se organizaron los reportes por áreas de análisis, facilitando al usuario final una búsqueda rápida de información.

Arquitectura de información

Para la arquitectura de información se definió la estructura de navegación conformada por un área de análisis general (A.A.G), 5 áreas de análisis (A.A) y 2 libros de trabajo (L.T) por cada área de análisis para un total de 10 libros de trabajo. Cada libro de trabajo contiene las vistas de reportes, sumando un total de 80. A continuación se presenta la arquitectura de información correspondiente al Almacén de datos para el Sistema de Planificación de Actividades:

Capítulo 2: Análisis y diseño del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

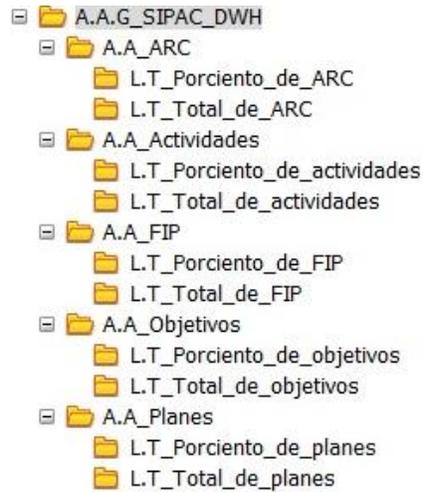


Figura 9. Arquitectura de información.

Diseño de los cubos OLAP

En el Almacén de datos para el Sistema de Planificación de Actividades se diseñaron 5 cubos multidimensionales, un cubo por cada tabla de hechos.

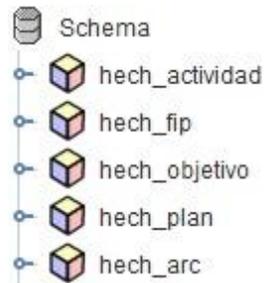


Figura 10. Cubos OLAP.

2.8 Política de respaldo y recuperación

Para garantizar la persistencia de la información, se estableció una política de respaldo y recuperación que comprende tres elementos fundamentales:

- Periodicidad de salvadas del sistema: las salvadas se llevarán a cabo durante todo el año, en un período aproximado de 30 días (copias parciales) y en los meses de junio y diciembre (copia total).

Capítulo 2: Análisis y diseño del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

- Tablas involucradas: las tablas que normalmente se involucran en la realización son los 5 hechos con los que cuenta el almacén de datos. Además las tablas de dimensiones relacionadas estarán implicadas en caso de cambiar.
- Salvas existentes: cada administrador de los servidores definirá dónde se almacenará la información guardada. Se realizará el chequeo de su estado mensualmente.

2.9 Esquema de seguridad

La seguridad es un aspecto crucial para el desarrollo de almacenes de datos. La información que ellos contienen es un requisito serio que debe ser considerado, no como un aspecto aislado, sino como algo presente en todas las etapas del ciclo de vida de desarrollo, desde el análisis de los requerimientos hasta la implementación y mantenimiento. (42)

La seguridad del Almacén de datos para el Sistema de Planificación de Actividades está dada en gran medida por los niveles de acceso al sistema, apoyándose específicamente en los roles definidos para la interacción de los usuarios con el mismo.

Seguridad en la base de datos

Para el acceso a la base de datos se definieron los roles que se describen en la tabla 5. Estos manejan el acceso a las funcionalidades que le corresponde a cada rol para la interacción con la base de datos.

Tabla 6. Usuarios en la base de datos

Usuario	Descripción
Administrador ETL	Tiene permiso solo para realizar las funciones del proceso ETL.
Administrador	Tiene acceso total a la base de datos. Realiza las operaciones de administración en la misma.
Analista	Tiene acceso de solo lectura a la base de datos. Accede a la misma solo mediante la aplicación.

Seguridad en el proceso ETL

Para la seguridad en el proceso ETL se garantizó que los datos de las fuentes, las transformaciones y los trabajos no sean manipulados mediante la seguridad que ofrece el

Capítulo 2: Análisis y diseño del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

sistema operativo. Teniendo en cuenta que el almacén es multiplataforma a continuación se describe cómo se garantizó la seguridad en el proceso ETL tanto para el sistema operativo Windows como para Linux.

En Windows se marca el atributo de sólo lectura en las propiedades de la carpeta donde se almacenan todos ellos, de tal forma que solo el usuario que tenga permisos de escritura pueda modificar los ficheros. (Ver Anexo 1)

En el caso del sistema operativo Linux para mantener la seguridad en el proceso ETL se utilizó a través de la consola el comando *chmod 755* que indica asignarle permisos al *propietario* de lectura, escritura y ejecución, en el caso del *grupo* y *otros* solo de lectura y ejecución; indicando seguidamente la dirección donde se encuentra el fichero. (Ver Anexo 2)

Seguridad en la aplicación

Es de gran importancia que se defina un esquema de seguridad que permita la sostenibilidad en la gestión en este servidor, debido a la cantidad de usuarios que demandan acceso a ella. A continuación se muestran los roles que se definieron para que el acceso a la aplicación sea de forma organizada y con los permisos correspondientes.

Tabla 7. Usuarios en la aplicación

Usuario	Descripción
Administrador	Tiene acceso total al Sistema de Planificación de Actividades (SIPAC). Realiza operaciones de administración sobre la misma.
Analista	Tiene acceso de solo lectura al Almacén de Datos. Analiza la información contenida en los reportes.

La herramienta Pentaho BI server 4.8.0 contiene su propia seguridad estableciendo tres aspectos fundamentales (43):

- Seguridad de acceso a datos de objetos: incluye usuarios, contraseñas, autorizaciones permitidas, recursos web y protección a datos.
- Autenticación: tiene que ver con el procesamiento de información interactiva de inicio de sección (por ejemplo de nombre de usuario y contraseña) comparándola con la información recuperada del almacén de datos de seguridad.

Capítulo 2: Análisis y diseño del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

- Autorización de recursos web (URL): brinda protección a las URL para indicar a cada usuario si puede o no acceder a una determinada página.
- Autorización a objetos del dominio: en el sistema, los únicos objetos del sistema protegidos por la plataforma son los objetos de repositorio otorgados al usuario autenticado. Es responsabilidad de los objetos del dominio autorizar las operaciones solicitadas por este.

2.10 Conclusiones del capítulo

Después de realizar los procesos de análisis y diseño del DWH para SIPAC se concluye que:

- Con el levantamiento de requisitos se obtuvo como principal resultado la identificación de 10 RI, 18 RF y 6 RNF, permitiendo a partir de ellos la elaboración del DCUS.
- Se identificaron 9 RN, posibilitando de esta forma la definición de las reglas de transformación que fueron utilizadas en los procesos de integración de datos.
- Se efectuaron la descripción de los CU del DWH permitiendo describir las funcionalidades del sistema.
- Con la definición de la arquitectura del DWH, se posibilitó identificar los elementos y subsistemas involucrados en el desarrollo de la solución.
- Se diseñó el modelo de datos, que permite representar las relaciones entre las 5 tablas de hechos, las 10 medidas y las 18 dimensiones identificadas para el DWH.
- Con el diseño de los subsistemas de integración de datos y visualización, se logró definir el diseño general del proceso de ETL, la arquitectura de información del DWH y el diseño de los reportes candidatos, todo esto permitió guiar la implementación de dichos subsistemas.
- Mediante las políticas de respaldo y recuperación creadas, se definió la periodicidad de las salvadas de forma mensual y las tablas de la BD implicadas en esto.
- Los roles y permisos definidos en el DWH, se permite colaborar con la seguridad del sistema.

Capítulo 3: Implementación y prueba del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

CAPITULO 3: IMPLEMENTACIÓN Y PRUEBA DEL ALMACÉN DE DATOS PARA EL SISTEMA DE PLANIFICACIÓN DE ACTIVIDADES (SIPAC).

Una vez realizado el diseño del DWH y teniendo como guía la metodología utilizada, en el presente capítulo se procedió a realizar la implementación de cada uno de los subsistemas que integran el almacén de datos: almacenamiento, integración y visualización. En este capítulo además se presentaron las pruebas realizadas a la solución, mediante la aplicación de las listas de chequeo y los casos de prueba (CP) por Caso de Uso (CU), así como los resultados obtenidos en cada una de ellas luego de su aplicación.

3.1 Implementación del subsistema de almacenamiento.

Para la implementación del subsistema de almacenamiento se definieron estándares de codificación de las estructuras de datos para una mayor comprensión y se desarrolló además el modelo o estructura físico del almacén.

3.1.1 Estándares de codificación.

Con el objetivo de organizar cómo se van a denominar las estructuras del DWH, se formaliza un modelo, norma, patrón o estándar de codificación. Esta acción permite a los desarrolladores entender cada una de las estructuras de los mercados. Los estándares de codificación utilizados en el Almacén de datos para SIPAC se muestran a continuación:

Tabla 8. Estándares de codificación.

Tipo de objeto	Función	Nomenclatura	Descripción
Esquema	Dimensiones compartidas	dwh_compartidas	Esquema donde se encuentran las dimensiones compartidas por varios hechos.
	Metadatos	metadatos	Esquema donde se encuentran las tablas que almacenan los metadatos.
	Dimensiones	dwh_dimensiones	Esquema donde se almacenan las tablas de hechos y dimensiones definidas para recoger los datos

Capítulo 3: Implementación y prueba del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

			asociados a cada área.
Tablas	Tablas de dimensiones	dim_[nombre dimensión]	Todas las tablas de dimensiones tienen una cadena (dim_) y el concepto que describen.
	Tablas de hechos	hech_[nombre hecho]	Todas las tablas de hechos tienen una cadena (hech_) y el concepto que describen.
	Tablas de metadatos	[nombre metadatos]	Todas las tablas de metadatos tienen el concepto que describen.
Contrains	Llaves primarias	pk_[nombre tabla]_id	Todas las llaves primarias tienen una cadena que la identifica (pk_) y el nombre de la tabla a la que pertenecen, y además tienen el identificador (_id), serán de tipo numérico y autoincremental.
	Llaves foráneas	fk_[nombre tabla]_id	Todas las llaves foráneas tienen una cadena que las identifica (fk_) y el nombre de la tabla a la que pertenecen, y además tienen el identificador (_id), serán de tipo numérico y autoincremental.
Índices	Índices	ind_[nombre atributo]	Los índices se nombran con una cadena que los identifique (ind_) y a continuación el nombre del atributo de la tabla que se hace referencia.

Capítulo 3: Implementación y prueba del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

3.1.2 Modelo de datos físico.

El modelo de datos físico es una colección integrada de conceptos que permite describir los datos y las relaciones que existen entre ellos. Se obtiene a partir del modelo lógico dimensional, permitiendo describir las estructuras de datos, las restricciones de integridad y las operaciones de manipulación de los datos. Recoge las relaciones existentes entre las tablas de hechos y las dimensiones que conforman el almacén. Además se utiliza para describir los datos a un nivel interno, proporcionando detalles de cómo son almacenados los datos en el ordenador. En la solución de este modelo se utilizaron 20 secuencias, 22 índices y 27 tablas.

Índices

Un índice es una estructura de disco asociada con una tabla o una vista que se utiliza, principalmente, para mejorar el rendimiento de una base de datos. Contiene claves generadas a partir de una o varias columnas de la tabla o la vista. Se utilizan principalmente en los campos sobre los cuales se hacen búsquedas con más frecuencia. (44) Para garantizar un acceso más rápido a las consultas realizadas en el almacén se implementaron los índices, los cuales fueron desarrollados utilizando la estrategia btree (Árboles balanceados), permitiendo que cualquier búsqueda que se realice haciendo uso de las llaves se optimizará mediante este método.

Esquemas y tablas

Para facilitar una mejor organización de las tablas en el DWH para SIPAC se crearon 3 esquemas: “*compartidas*”, “*dimensiones*” y “*metadatos*”. La función de cada uno de ellos se describe en los estándares de codificación y las tablas contenidas en cada uno de ellos. A continuación se mencionan:

Tabla 9. Modelo de datos físico.

Esquemas	Tablas
 dwh_compartidas	 dim_arc  dim_cumplimiento  dim_dirigente  dim_fip  dim_involucrado  dim_objetivo

Capítulo 3: Implementación y prueba del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

	<ul style="list-style-type: none">  dim_participante  dim_plan  dim_tiempo_mes  dim_involucrado
<ul style="list-style-type: none">  dwh_dimensiones 	<ul style="list-style-type: none">  hech_actividad  hech_arc  hech_fip  hech_objetivo  hech_plan  dim_actividad  dim_capítulo  dim_categoría_actividad  dim_categoría_plan  dim_emite  dim_nivel_acceso_fip  dim_tipo_actividad  dim_tipo_fip
<ul style="list-style-type: none">  metadatos 	<ul style="list-style-type: none">  transformación  trabajo  gestión_cambio_datos_fuente  gestión_histórico_datos_fuente

A continuación se presenta cómo quedó estructurada la BD en la herramienta PgAdminIII:

Capítulo 3: Implementación y prueba del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

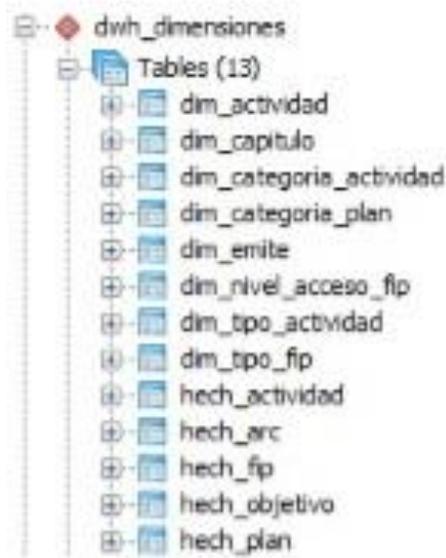


Figura 11. Estructura de la BD.

3.2 Implementación del subsistema de integración de datos.

El proceso de integración de los datos consta de 3 etapas fundamentales relacionadas entre sí: extracción, transformación y carga de datos. Para identificar y corregir los problemas se realizó la limpieza de los datos que permite llenar valores ausentes y corregir errores. Una vez que los datos son transformados se cargaron, poblando las dimensiones y los hechos que conforman la estructura del DWH para SIPAC. Para llevar a cabo el proceso de integración de los datos se utilizó un conjunto de subsistemas ETL de los 34 que Kimball propone:

Subsistema de extracción: se utiliza para obtener los datos de la fuente de datos.

Subsistema de limpieza de datos: se realizan un conjunto de pasos correspondientes a la limpieza y corrección de los datos.

Subsistema de rastreo de eventos de errores: permite capturar los datos erróneos al realizar las transformaciones y ser enviados luego a un archivo de tipo Excel.

Subsistema de control de versiones: se realiza la gestión histórica de los datos de la fuente, registrando cuándo ocurrió la última carga, aplicándose mediante la tabla metadatos gestión_histórico_datos_fuente.

Subsistema de repositorio de metadatos: obtiene datos importantes de la transformación y posibilita la gestión de la captura del cambio de los datos.

Capítulo 3: Implementación y prueba del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

Subsistema de Dimensiones Lentamente Cambiantes (SCD): implementa la lógica para crear atributos de variabilidad lenta a lo largo del tiempo. Se utiliza la SCD de tipo 2 para almacenar la información de varios años correspondiente al SIPAC.

Subsistema de captura del cambio de los datos: detecta los cambios correspondientes a las fechas que permitirán realizar las cargas mensualmente. Se utiliza para detectar los cambios de la última extracción.

Subsistema de perfilado de datos: explora los datos de la fuente para verificar su calidad y el cumplimiento de los estándares conforme a los requerimientos.

Subsistema de tablas de hecho: permite crear tablas de hechos.

3.2.1 Gestión del cambio en las dimensiones.

Durante la implementación del subsistema de integración, se procedió primeramente a realizar un total de 23 transformaciones correspondientes a la carga de las dimensiones. Estas dimensiones son lentamente cambiantes, exceptuando las dimensiones que están relacionadas con las *actividades*, *planes* e *involucrados* debido a que cada vez que se ejecuta se obtienen nuevos datos que cargar o datos que actualizar.

Las dimensiones lentamente cambiantes o SCD (Slowly Changing Dimensions) son dimensiones en las cuales sus datos tienden a modificarse a través del tiempo, ya sea de forma ocasional o constante, o implique a un solo registro o a la tabla completa. Cuando ocurren estos cambios, se puede optar por seguir alguna de estas dos grandes opciones: (45)

- Registrar el historial de cambios.
- Reemplazar los valores que sean necesarios.

La estrategia a seguir específicamente en el DWH para SIPAC es *Registrar el historial de cambios*. A continuación se especifican los tipos de estrategia a seguir cuando se tratan las SCD:

SCD tipo 1: Sobrescribir.

SCD tipo 2: Añadir fila.

SCD tipo 3: Añadir columna.

SCD tipo 6: Híbrido.

SCD tipo 4: Tabla de Historia separada.

Capítulo 3: Implementación y prueba del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

Las dimensiones que pertenecen al DWH para SIPAC se clasifican como SCD tipo 2. Para dar garantía de esto se utilizó el componente *Insertar\Actualizar*, marcando la opción de no realizar actualizaciones (ver figura 15), a través de esto se aseguró que no se realicen actualizaciones en el almacén sino que se inserte una nueva fila en la tabla.

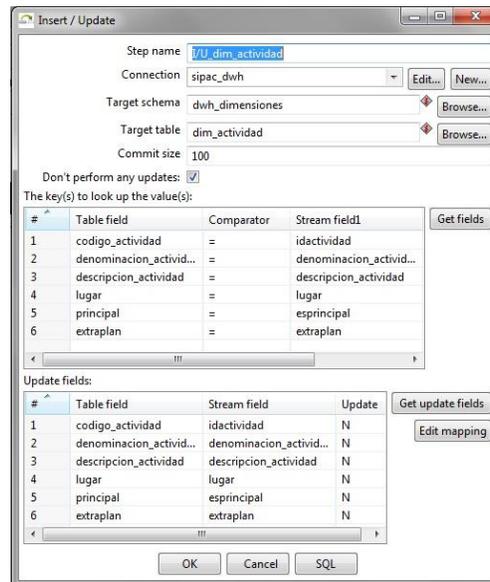


Figura 12. Componente Insertar\Actualizar.

3.2.2 Gestión de los metadatos del proceso de integración de datos.

Existen diversos tipos de metadatos, cada uno con su propio formato para describirlos. Según la función que proporcionan se clasifican en:

- Metadatos descriptivos: describen e identifican recursos de información, permitiendo a los usuarios la búsqueda y recuperación de la información.
- Metadatos estructurales: facilitan la navegación y la presentación de los recursos, proporcionando información sobre la estructura interna de los documentos, así como la relación entre ellos.
- Metadatos de proceso: utilizados para capturar la información de los procesos en que se ejecutan.
- Metadatos administrativos: facilitan la gestión de conjuntos de recursos, incluyendo la gestión de derechos y control de acceso y uso.

Capítulo 3: Implementación y prueba del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

Para esta investigación fueron utilizados los metadatos de proceso con el objetivo de obtener la información de las transformaciones y los trabajos que pertenecen a los subprocesos del ETL.

3.2.3 Captura de los cambios en los datos.

El conocimiento del flujo de información dentro de las diferentes entidades que utilizan el SIPAC es de vital importancia para determinar la forma en que se realizó la captura de los cambios. La captura de la información perteneciente a estas entidades se llevó a cabo desde los departamentos con los que cuenta cada una de ellas, donde se realizan las operaciones pertinentes a los especialistas de cada departamento.

Teniendo en cuenta lo descrito previamente se definió la obtención de los cambios en los datos en el almacén anualmente, este cambio se gestiona mediante la tabla de metadatos *gestión_cambio_datos_fuente*. Esta tabla almacena las fechas que corresponde cargar de cada uno de los hechos. Al realizar esta carga, se actualizan las fechas de inicio y fin; posibilitando así conocer la fecha en que se realizó la última carga al DWH. El proceso de la carga gestiona además el cambio histórico de los datos mediante la tabla *gestión_histórico_datos fuente*.

3.2.4 Implementación de los hechos.

En la carga de los hechos se realizaron un total de 5 transformaciones, luego de haber cargado ya las dimensiones. A continuación se describe la transformación del *hech_objetivo*:

A través del componente *entrada tabla* el cual permite leer información desde una BD utilizando código SQL, se obtuvieron haciendo uso del subsistema de extracción, los datos referentes a los objetivos, así como la relación con los otros elementos de la planificación. Luego de la extracción se realizaron un conjunto de pasos correspondientes a la limpieza y transformación de los datos, donde se aplicaron las reglas del negocio identificadas durante del diseño del almacén. Para llevar a cabo este proceso se utilizaron varios componentes dentro de los que se encuentran *seleccionar/renombrar*, *búsqueda en tabla*, *mapeo de datos*, *JavaScript*, *adicionar constantes* y *corte de cadena*. Mediante el subsistema rastreo de errores se capturaron los datos erróneos al realizar las transformaciones y fueron enviados luego a un archivo de tipo Excel, el cual le informará al cliente los campos que tuvieron irregularidades para un posterior tratamiento por parte del equipo de soporte del DWH para SIPAC. Finalmente se cargan los datos ya transformados e integrados al almacén en la tabla del hecho correspondiente, para ello se utilizó el componente *insertar/actualizar* especificándole que no realice actualizaciones sobre los datos. Mediante el control de versiones se realizó la gestión histórica de los datos de la

Capítulo 3: Implementación y prueba del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

fuelle registrando cuándo ocurrió la última carga, la cual se aplicó a través de la tabla metadatos *gestión_historico_datos_fuente*; utilizando además el subsistema de repositorio de metadatos para obtener datos importantes de la transformación y posibilitar la gestión de la captura del cambio de los datos.

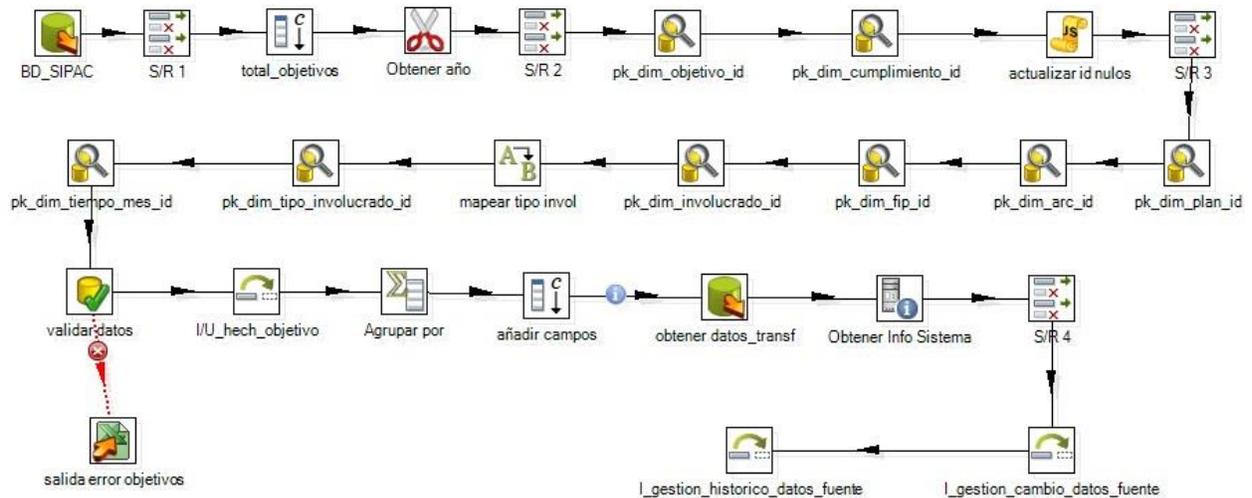


Figura 13. Transformación del hech_objetivo.

3.2.5 Implementación de los trabajos.

Un trabajo o “Job” es un conjunto de tareas con el objetivo de realizar una acción determinada. Ellos permiten ejecutar varias transformaciones o trabajos previamente diseñados y organizar una secuencia de ejecución de los mismos, ejemplo: primero las dimensiones y después los hechos. En el caso del DWH para SIPAC se creó un trabajo para llenar las dimensiones compartidas por varios hechos, 5 para cargar los hechos, los cuales ejecutan las transformaciones propias de cada hecho y posteriormente el hecho y un trabajo general que es el encargado de ejecutar el resto de los trabajos (ver figura 14). Con lo anteriormente mencionado se garantizó que si no se cargan las dimensiones compartidas, el trabajo se aborte y además que se pueda cargar algunos hechos y otros no, dependiendo si en alguna de sus dimensiones surge algún error.

Capítulo 3: Implementación y prueba del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)



Figura 14. Trabajo general.

3.3 Implementación del subsistema de visualización.

Para definir qué información se presenta a los usuarios y cómo se muestra para el análisis se llevó a cabo la implementación del subsistema de visualización. Como parte de este subsistema se implementan los cubos OLAP y los reportes candidatos.

3.3.1 Cubos multidimensionales.

En los cubos OLAP la información es organizada de forma jerárquica, lo que posibilita llevar a cabo un análisis rápido, facilitando estructurar los datos para que concuerden con el modo que tienen los usuarios de analizarlos. En el DWH para SIPAC se desarrollaron un total de 5 cubos, uno por cada hecho, con sus dimensiones y medidas asociadas.

En este proceso inicialmente se definieron las dimensiones compartidas entre ellos, dentro de estas fueron creadas primeramente 22 jerarquías. Las dimensiones van a estar compuesta por una o más jerarquías y las jerarquías a su vez contienen un *Table* que hace referencia a la tabla que corresponde la dimensión (ver figura 15) y además contienen los niveles, que son los campos que se muestran.

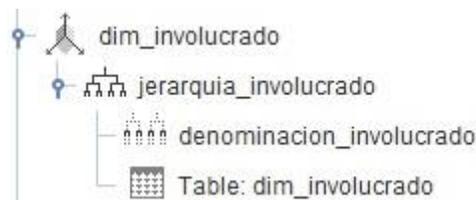


Figura 15. Dimensión involucrado.

Se indicaron los parámetros *hasAll* para agrupar todos los valores de la jerarquía y su descripción *allMemberName* que será el título de los valores que saldrán en los reportes; para ello se definió como *Todos* dependiendo del género de la dimensión. Se seleccionó también la

Capítulo 3: Implementación y prueba del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

llave primaria correspondiente a la tabla dimensión y por último el *caption* el cual es el nombre con el cual se representa la jerarquía en la columna. A continuación se muestra la jerarquía correspondiente a la dimensión involucrado:

Attribute	Value
name	jerarquia_involucrado
description	
hasAll	<input checked="" type="checkbox"/>
allMemberName	Todos
allMemberCaption	
allLevelName	
defaultMember	
memberReaderClass	
primaryKeyTable	
primaryKey	pk_dim_involucrado_id
caption	Involucrado

Figura 16. Jerarquía de la dimensión involucrado.

Para los niveles que componen las jerarquías se indicaron como atributos la tabla (*table*) a la que pertenecen, la columna de la base de dato que la distingue (*column*) y la columna de la cual se obtuvo dinámicamente el nombre (*nameColumn*), definiéndose además los atributos *tipo de dato*, *tipo de nivel*, el atributo *hideMemberIf* y el *caption* que define el nombre que va a representar los valores de este nivel. Esto se realizó para cada uno de los niveles de las jerarquías definidas.

Attribute	Value
name	denominacion_involucrado
description	
table	dim_involucrado
column	codigo_involucrado
nameColumn	denominacion_involucrado
parentColumn	
nullParentValue	
ordinalColumn	
type	String
uniqueMembers	<input checked="" type="checkbox"/>
levelType	Regular
hideMemberIf	Never
approxRowCount	
caption	Involucrado
captionColumn	
formatter	

Figura 17. Nivel de jerarquía de involucrado.

Seguidamente se crearon los cubos OLAP en el cual se definieron el *nombre*, el *caption*, la opción *cache* y la opción *enable* para que el cubo sea visible. Para la implementación de los mismos se definió además la tabla física de la base de datos con la que se asocia, luego se incluyeron las dimensiones correspondientes a cada cubo, el cual heredó todas las

Capítulo 3: Implementación y prueba del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

características que haya sido incluido en las mismas, incluyendo todas las jerarquías y sus correspondientes atributos, esto se logró mediante el elemento *dimensiones usadas*. Además el cubo contiene las *medidas* que van hacer los valores de análisis y miembros calculados que son fórmulas en las que se utilizaron otras medidas.

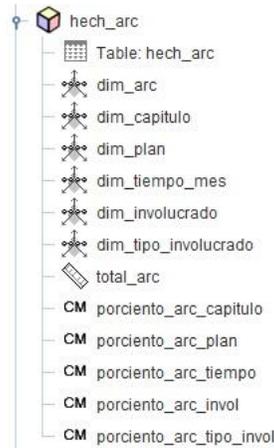


Figura 18. Cubo OLAP del hecho ARC.

Se definieron un total de 5 medidas, una por cada cubo, que se calculó directamente con campos de la BD, donde tomó el nombre total_[nombre del área de análisis]. De estas se identificaron la *función de agregación*, la *columna* que la generó, el *tipo de dato*, el *caption* y se marcó la opción *visible* en todos los casos.

Attribute	Value
name	total_arc
description	
aggregator	distinct count
column	pk_dim_arc_id
formatString	
visible	<input checked="" type="checkbox"/>
datatype	Integer
formatter	
caption	Total de ARC

Figura 19. Medida del hecho ARC.

Cuando se desea calcular una medida cuyo valor no proviene de una columna de la tabla de hechos, sino de una fórmula MDX, se utilizaron los *miembros calculados*. En el caso del DWH para SIPAC se crearon un total de 40 para el cálculo de porcentaje (_[área de análisis]_[elemento]). De ellos se indicaron el *nombre*, el *caption* que será el nombre que

Capítulo 3: Implementación y prueba del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

aparecerá en los análisis, la opción *visible*, el *formato* que contendrá el resultado y la *fórmula* que generará el valor de la medida calculada que es el atributo más importante.

Attribute	Value
name	porciento_arc_capitulo
description	
caption	% de ARC por capitulo
dimension	Measures
visible	<input checked="" type="checkbox"/>
formula formulaElement..	{[Measures].[total_arc]/[dim_capitulo].[Todos]}
formatString	#.00%

Figura 20. Miembros calculados del hecho ARC.

3.3.2 Reportes.

Se utilizó Pentaho BI server 4.8.0 para la realización de los reportes, añadiendo algunas configuraciones para el uso propio del DWH, definiéndose una conexión única para el almacén. Para la implementación de los reportes se utilizaron consultas MDX. A continuación se describe la consulta MDX del reporte *Total de ARC por capítulo*, la misma muestra los capítulos que tienen asociados los ARC y el total de ARC por capítulo.

```
Select NON EMPTY {[Measures].[total_arc]} ON COLUMNS,  
NON EMPTY Hierarchize (Union ({[dim_capitulo.jerarquia_capitulo].[Todos]},  
[dim_capitulojerarquia_capitulo] [Todos].Children)) ON ROWS  
from[hech_arc]
```

Estos reportes son el punto de partida para la toma de decisiones, facilitando el análisis de la información. A continuación se muestra la imagen de la vista de análisis “Total de ARC por capítulo” perteneciente al reporte “L.T_Total_de_ARC” (ver figura 21).

Capítulo 3: Implementación y prueba del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

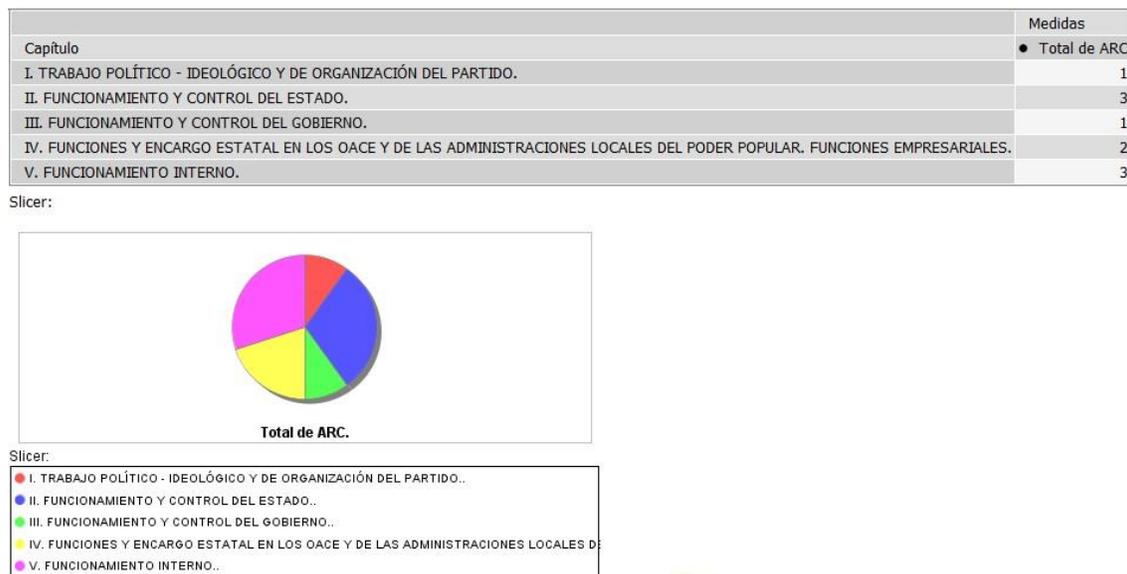


Figura 21. Vista del reporte “Total de ARC por capítulo”.

3.4 Resultados obtenidos.

Con el DWH para el Sistema de Planificación de Actividades se realizó el proceso de perfilado de los datos, a través del cual se obtuvo una visión del estado en que se encontraba la fuente que posteriormente fue sometida al proceso ETL. La implementación del subsistema de integración permitió que la información procedente de varias fuentes estuviera integrada y con buena calidad, centralizada en un sistema de soporte de almacenamiento de datos el cual fue diseñado para ser flexible a cambios. Además se implementaron todas las vistas de análisis que poseen información actualizada, la cual puede ser consultada por cualquier persona sin necesidad de ser un especialista en informática o posea altos conocimientos del negocio. Con estas vistas de análisis se solucionaron las principales necesidades de información de los clientes, contribuyendo a la mejora del proceso en la toma de decisiones de los directivos de las entidades que utilizan el SIPAC.

3.5 Validación de la solución.

El proceso de desarrollo de un software debe ir acompañado de una actividad que garantice la calidad del resultado final. Durante el desarrollo de un producto de software se puede evitar que los errores se produzcan, y sobre todo que se propaguen mediante pruebas y procedimientos de calidad que acompañen al producto a lo largo de su ciclo de vida. Por esta razón, se hizo necesario aplicar diferentes pruebas luego de culminada la etapa de implementación, con el

Capítulo 3: Implementación y prueba del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

propósito de verificar no solo que el producto cumpla con los requisitos del cliente, sino también para eliminar los posibles defectos que pudiera tener.

A continuación se muestran algunas de las pruebas que pudieron ser utilizadas para la validación de la solución: (46)

Pruebas unitarias: centra el proceso de verificación en la menor unidad del diseño del software: el componente de software o módulo.

Pruebas de interacción: es una técnica sistemática cuyo objetivo consiste en probar el sistema como un todo, para detectar errores asociados con la interacción entre los diferentes módulos que lo componen.

Pruebas de validación: proporciona una seguridad final que el software satisface todos los requisitos funcionales, de comportamiento y rendimiento.

Pruebas de regresión: consiste en volver a ejecutar un subconjunto de pruebas que se han llevado a cabo anteriormente, para asegurarse que los cambios que se hayan realizado no introduzcan un comportamiento no deseado a errores adicionales.

Pruebas de aceptación: se realizan para probar que el sistema cumpla con los requerimientos y expectativas del cliente.

SIPAC cuenta con una estrategia de prueba, pero por las características que posee la solución se tomó la decisión de regirse por una estrategia de prueba propia para un DWH. Para lograr obtener un producto con calidad que cumpla con los requerimientos establecidos y la satisfacción del cliente, se realizaron las pruebas de software de acuerdo al Modelo V definido por DATEC. Este modelo tiene mucha similitud con el Modelo en cascada pues constituye una evolución del mismo. En lugar de establecer una fase independiente al concluir el proyecto como define el Modelo en cascada, el modelo V toma las actividades de prueba como un proceso que corre en paralelo. El mismo propone utilizar un conjunto de pruebas, de ellas solo se utilizan las pruebas unitarias, de integración y pruebas de aceptación, evitando generar un plan de pruebas por cada una de ellas a través de los artefactos que se obtienen y teniendo en cuenta las características propias del DWH para SIPAC.

Capítulo 3: Implementación y prueba del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

En la siguiente figura, se ilustra una representación gráfica del ciclo de vida del software propuesta en el modelo V, donde a la izquierda se muestran las diferentes etapas de desarrollo, mientras que a la derecha se muestran las pruebas correspondientes a cada una de ellas.

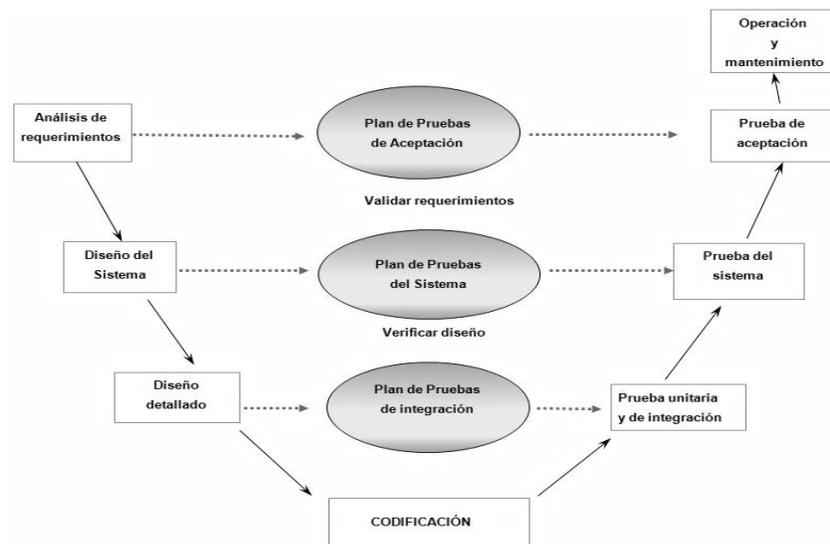


Figura 22. Modelo V

3.5.1 Herramientas para la aplicación de las pruebas.

Para lograr que un producto de software tenga calidad, es necesario realizar un conjunto de evaluaciones durante todo el proceso de desarrollo que implique al cliente y desarrollador. Con esto se garantiza que el sistema desarrollado cumpla con los requisitos, tanto funcionales como no funcionales. Se emplearon como herramientas los tipos de pruebas mencionados anteriormente, los CP para CU, las listas de chequeo a los artefactos ETL y el perfilado de los datos.

Pruebas unitarias

Una vez culminada la implementación se le aplicaron pruebas unitarias a los flujos de integración de datos y a los diferentes componentes relacionados con la capa de visualización. Durante la validación del subsistema de integración se detectaron elementos duplicados. Para solucionar este problema se agregaron a la transformación los componentes Ordenar y Unir filas, los cuales permitieron ordenar las tuplas por el nombre y luego agrupar en una sola fila todos los nombres repetidos. Una vez cargados los datos se verificó además que durante la ejecución de la transformación no se perdiera ninguno.

Capítulo 3: Implementación y prueba del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

Se validó además el subsistema de visualización detectando inicialmente que los totales que se mostraban no coincidían con los que estaban en la fuente. Al comprobar que los datos ausentes se encontraban en el directorio donde se almacenan los datos defectuosos, el problema quedó solucionado revisando que el rango de fecha de cada una de las tuplas se encuentre dentro del rango de la dimensión tiempo. Finalmente se comprobó que el total visualizado coincidía exactamente con la fuente, por lo que no había pérdida de la información.

Caso de prueba para CU

Los CP por CU se diseñan en correspondencia con las funcionalidades detalladas en los CU de información. Este diseño se confecciona antes de realizar las pruebas funcionales de la aplicación. Específicamente para el DWH fueron diseñados 12 CP.

Después de realizado cada CP de validación, se cumplió una de las dos condiciones posibles:

- La funcionalidad cumple con la especificación y se le acepta.
- Se descubre un error en la funcionalidad y se crea una lista con las no conformidades encontradas.

En la siguiente tabla se muestra el CP correspondiente al CU *Visualizar información de Actividades*, con una breve descripción, la respuesta que debe dar el sistema y el flujo central de ejecución del caso de prueba (ver Tabla 10):

Tabla 10. CP de CU *Visualizar información de Actividades*.

Escenario	Descripción	Variable de salida	Respuesta del sistema	Flujo central
1.1 Visualizar información de Actividades.	El CU inicia cuando el analista selecciona visualizar información en actividades, luego se realiza la selección de la	Actividades, FIP, ARC, Plan, Objetivo, Cumplimiento, Dirigentes, Participantes, Tipo de	Se muestran la información referente a las perspectivas de análisis que se relacionan con las actividades.	1. Crear una nueva vista de análisis a través de la opción Nueva vista de análisis 2. Seleccionar nueva vista de análisis de tipo Actividad 3. Dar

Capítulo 3: Implementación y prueba del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

	información a mostrar.	actividad, Categoría, Involucrado, Tipo de involucrado, Tiempo.		click sobre la lista desplegable Todos para detallar la información 4. El área de trabajo visualiza la información detallada referente al reporte.
1.2 No visualizar la información de Actividades.	El CU inicia cuando el analista selecciona visualizar información de Actividades, luego cancela la selección de la información a mostrar.		No se muestra la información referente a la perspectiva de análisis que se relacionan con las Actividades.	1. Crear una nueva vista de análisis a través de la opción Nueva vista de análisis 2. Seleccionar nueva vista de análisis de tipo Actividad 3. Dar click en la opción Cancelar para no visualizar la información.

Aplicación de las listas de chequeo

Las listas de chequeo son un mecanismo para controlar los riesgos, tienen como objetivo examinar cuidadosamente áreas importantes para lograr identificar problemas comunes, encontrar errores y solucionarlos en un espacio breve de tiempo. Constituyen un documento con un conjunto de parámetros a medir de un aspecto específico, sea documentación o aplicación.

Para elaborar las listas de chequeo se tuvieron en cuenta los elementos de evaluación que no deben faltar una vez realizado el proceso ETL y visualización de los datos del DWH para SIPAC. Las listas de chequeo contienen diferentes indicadores a evaluar, los cuales se encuentran distribuidos en las siguientes secciones:

Capítulo 3: Implementación y prueba del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

Estructura del documento: abarca todos los aspectos definidos por el expediente de proyecto o el formato establecido por el proyecto.

Indicadores definidos por la etapa: abarca todos los indicadores a evaluar durante la etapa ETL e Inteligencia de negocio.

Semántica del documento: contempla todos los indicadores a evaluar respecto a la ortografía y la redacción.

Las listas de chequeo se le aplicaron a los artefactos de ETL *Perfilado de los datos (PD)* y *Diccionario de datos (DD)*. Seguidamente se muestra una tabla en la que se encuentran los principales aspectos que fueron evaluados en las listas de chequeo y resultados que arrojó la aplicación de los mismos:

Tabla 11. Aplicación de las listas de chequeo a los artefactos ETL

Secciones	PD	DD
Estructura	4	7
Indicadores	1	1
Semántica	3	3
TI	7	11
IC	4	4
NC	6	5
NC-IC	2	2

TI: total de indicadores

IC: indicadores críticos

NC: no conformidades detectadas en la lista de chequeo

NC-IC: no conformidades detectadas que pertenecen a indicadores críticos

Capítulo 3: Implementación y prueba del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

A continuación se muestra un ejemplo de una lista de chequeo aplicada al artefacto Diccionario de Datos (DD):

Tabla 12. Lista de chequeo aplicada al Diccionario de Datos.

Estructura del documento					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
crítico	1. ¿El entregable contiene las secciones obligatorias de la plantilla estándar definida para el expediente de proyecto?	0		0	
crítico	2. ¿El alcance del proyecto describe correctamente los datos de las dimensiones y hechos del almacén de datos?	0		0	
crítico	3. ¿El objetivo expresa correctamente el propósito del documento?	0		0	
	4. ¿Se hace un uso adecuado del control del documento?	0		0	
	5. ¿En la sección de acrónimos se definen todos los acrónimos utilizados en el documento?	0		0	
	6. ¿En el entregable, la definición de las variables se hace correctamente?	0		0	
	7. ¿Existe una adecuada correspondencia entre las variables definidas y las descripciones que tienen estas	0		0	

Capítulo 3: Implementación y prueba del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

	variables?				
	8. ¿En el entregable se crea una hoja por cada variable definida?	1			
	9. ¿Queda registrado en el entregable todos los posibles valores que van a tener las variables definidas?	1			
Indicadores definidos en el desarrollo					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
	1. ¿Se utilizó un lenguaje cuyas sentencias son expresables mediante una sintaxis bien definida?	0		0	
Semántica del documento					
Peso	Indicadores a evaluar	Eval	(NP)	Cantidad de elementos afectados	Comentarios
Crítico	1. ¿Se han identificado errores ortográficos en los entregables?	1			
Crítico	2. ¿Se entiende claramente lo que se ha especificado en el documento?	0		0	
	3. ¿El número de página que aparece en el índice coincide con el contenido que se refleja realmente en dicha página?	0		0	

Calidad de los datos

Capítulo 3: Implementación y prueba del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

Tener la garantía de que los datos cargados tengan la calidad requerida es de vital importancia en el desarrollo del DWH para SIPAC, teniendo en cuenta que de esta forma se pudo afirmar que ellos no contienen errores. Este proceso de validación se llevó a cabo a través del perfilado de los datos, luego de haber concluido con el proceso ETL. Para ello se utilizó la herramienta *DataCleaner 1.5.4*, permitiendo valorar el nivel de calidad de los datos comprendidos en el almacén. De manera general se puede afirmar que la información contenida en el DWH es la que proviene de la fuente y que no existen valores nulos o vacíos, apoyándonos en los reportes generados por la herramienta los cuales sustentan con resultados positivos la calidad de lo cargado en el almacén. A continuación se presenta la gráfica que evidencia el resultado captado del perfilado de datos aplicado al hecho *Actividad*:

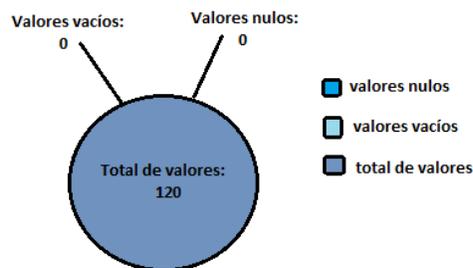


Figura 24. Calidad de los datos

3.6 Evaluación de los resultados.

Pruebas unitarias y de integración

Una vez culminada la implementación se realizaron pruebas unitarias a los flujos de integración de datos y a los diferentes componentes relacionados con la capa de visualización, siendo detectadas 10 no conformidades, todas de complejidad baja, las cuales fueron resueltas rápida y satisfactoriamente. A continuación se muestra una gráfica con los resultados de estas pruebas (ver figura 25):

Capítulo 3: Implementación y prueba del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

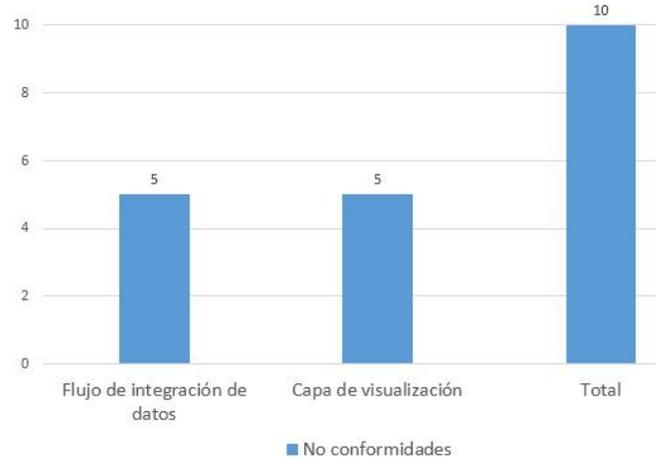


Figura 25. Comportamiento de los indicadores medidos en las pruebas unitarias

Listas de chequeo

Una vez realizadas las listas de chequeo a cada uno de los artefactos de ETL mencionados anteriormente, se detectaron 11 no conformidades de complejidad baja, las cuales fueron resueltas satisfactoriamente en un corto período de tiempo. La siguiente figura representa el comportamiento de los indicadores medios en las listas de chequeo (ver figura 26):

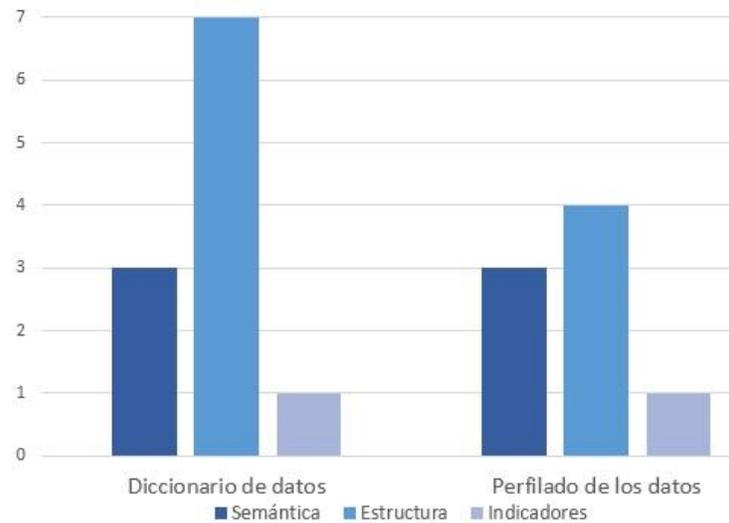


Figura 26. Comportamiento de los indicadores medidos en las listas de chequeo

Pruebas de aceptación

Capítulo 3: Implementación y prueba del almacén de datos para el Sistema de Planificación de Actividades (SIPAC)

En conjunto con el cliente se realizaron pruebas de aceptación a la solución, las cuales arrojaron resultados satisfactorios, quedando comprobado que el sistema cumple con sus necesidades y que están satisfechos con el producto elaborado (ver Anexos 3 y 4).

3.7 Conclusiones del capítulo.

Finalizada la etapa de implementación y prueba del DWH para SIPAC se concluyó que:

- Con la implementación del subsistema de almacenamiento quedaron definidas las estructuras físicas del DWH con 3 esquemas, 27 tablas: 18 de dimensiones, 5 hechos y 4 tablas de metadatos organizadas en 1 esquema.
- Con la implementación del subsistema de integración de datos, se logró ejecutar 10 trabajos y 31 transformaciones, pudiendo de esta forma poblar el DWH favorablemente.
- Se diseñaron 5 cubos OLAP, 5 AA y 10 libros de trabajo, así como 80 reportes candidatos para dejar implementado completamente el subsistema de visualización, divididos en 32 reportes candidatos por actividades, 14 por objetivos, 8 por planes, 10 por ARC y 16 por FIP, los cuales fueron implementados, permitiendo la visualización de la información.
- Se realizaron pruebas unitarias, de integración y de aceptación al DWH utilizando como herramienta las listas de chequeo, los casos de prueba y el perfilado de los datos, arrojando como resultados que todos los requerimientos de información, funcionales y no funcionales definidos fueron cumplidos satisfactoriamente.
- Quedó avalada la calidad de la solución mediante las pruebas aplicadas y la carta de aceptación firmada por el cliente, donde se confirmó que la solución cumple con las necesidades del cliente, quedando el almacén de datos listo para darle uso por parte de los especialistas del SIPAC.

CONCLUSIONES GENERALES

Con la realización del trabajo de diploma se logró cumplir con los objetivos planteados, desarrollándose una solución que ofrece soporte a la toma de decisiones mediante el almacén de datos para el Sistema de Planificación de Actividades de la Universidad de las Ciencias Informáticas.

Por lo que se concluye que:

- El estudio de los fundamentos teóricos de la investigación permitió seleccionar una metodología para organizar de manera estructurada el proceso de desarrollo de software, así como las herramientas que cuentan con las características necesarias para desarrollar la solución propuesta.
- El análisis y diseño del DWH posibilitó un mejor entendimiento de los procesos de negocio de SIPAC, permitiendo la implementación de una solución que responde a las necesidades del cliente.
- La implementación del Almacén de datos contribuyó al proceso de Ejecución y Control de la Planeación Estratégica u Operativa, a través del desarrollo de los diferentes subsistemas, garantizando la correcta organización, carga y visualización de los datos.
- La validación de la solución mediante la aplicación de los casos de pruebas, el perfilado de los datos y las listas de chequeo permitieron obtener un almacén funcional que cuenta con una correcta calidad de los datos y que cumple con los requisitos especificados por el cliente.

RECOMENDACIONES

Una vez concluida la implementación del DWH para SIPAC se recomienda para futuras investigaciones lo siguiente:

- Integrar el DWH al núcleo del Sistema de Planificación de Actividades.
- Utilizar otro producto de BI como los Cuadros de Mando Integral para lograr un mayor nivel de detalle.
- Realizar pruebas de rendimiento a la BD del DWH, teniendo en cuenta que se realizan cargas anuales por lo que el número de tuplas debe aumentar considerablemente.

REFERENCIAS BIBLIOGRÁFICAS

1. **Soto, R. y Cuervo Fernández, Ande.** *Nuevas Tendencias en Sistemas de Información: Procesos y Servicios.* Pecunia : Revista FCE, 2006. 1.
2. **Inmon, William H.** *Building the Data Warehouse. Fourth Edition.* Canada : Wiley Publishing, 2005. ISBN-13: 978-0-7645-9944-6.
3. **Kimball, Ralph y Ross, Margy.** *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modelling. Second Edition.* Canada : John Wiley and Sons, Inc., 2002. ISBN 0-471-20024-7.
4. **Kimball R., Reeves L., Ross M., Thornthwaite W.** *The Data Warehouse .* Inc 1998.
5. **Inmon W. H, John Wiley & Sons.** *Building the Data Warehouse.* Inc.1996.
6. Corret tu patrimonio seguro. [En línea] febrero de 2008. [Citado el: 9 de enero de 2014.] www.corett.gob.mx.
7. *Modelo de Data warehouse para los laboratorios Farmacéuticos.* San Salvador : s.n., 2010.
8. **System, Cisco.** *Configuraciones de referencia de Cisco para Fast Track Data Warehouse 4.0 de Micosoft SQL Server.* Estados Unidos : s.n., 2012.
9. **Noruega, Armando.** *Diseño de un Datawarehouse para los ensayos clínicos que se gestionan en el Centro de Inmunología Molecular.* Habana : s.n., 2008.
10. *Sistema DataWarehouse Comercial de la Corporación CIMEX.* s.l.: tomado de <http://www.redciencia.cu/empres/Intempres2002/Seminarios/CIMEXSAponencia.pdf>, 2006.
11. Décimo sexto Encuentro Nacional de Estadísticas. [En línea] 2004. [Citado el: 12 de noviembre de 2013.] www.one.gob.do/index.php?module=uploads&func=download .
12. **González, Yaneisi Pedrasa and Ricardo, Edgar Rojas.** Almacén de Datos Sala Situacional. [En línea] 2012. [Citado el: 11 de noviembre de 2013.] <http://publicaciones.uci.cu/indexPHP/SC>.

13. **Kimball, Ralph and Caserta, Joe.** *The Data Warehouse ETL Toolkit. Practical Techniques for Extracting, Cleaning Conforming and Delivering Data.* Canada : Wiley Publishing, Inc., 2004. ISBN 0-764-57923-1.
14. **Kimball, Ralph, Reeves, Laura, Ross, Margy and Thornthwaite, Warren.** s.l. : Wiley Publishing, Inc., *The Data Warehouse Lifecycle Toolkit. Expert Methods for Designing, Developing, and Deploying Data Warehouses.*
15. **Gerolami, Nicolás, Venzal, Germain and Revello, Esteban.** *Implantación de Data Warehouse Open Free.* Montevideo-Uruguay : Universidad de la República, Facultad de Ingeniería Instituto de Computación : s.n., 2011.
16. **Valdés Rodríguez, Yisel y Gómez, Jandy Miguel.** *Mercado de datos para la dirección de cuadros de la Administración provincial de Artemisa.* La Habana : Revista Cubana de Ciencias Informáticas, 2013.
17. *Procesamiento Analítico en Línea.* La Habana : Universidad de las Ciencias Informáticas, 2010.
18. Mis respuestas.com. *Qué es una metodología ?* [En línea] [Citado el: 12 de octubre de 2013.] <http://www.misrespuestas.com/que-es-una-metodologia.html>.
19. **Ramírez, MSc. Martha Denia Hernández.** Procedimiento para el desarrollo de un sistema de inteligencia de negocios en la gestión de ensayos clínicos en el Centro de Inmunología Molecular. [En línea] <http://scielo.sld.cu/pdf/aci/v22n4/aci06411.pdf>.
20. **Gustavo R. Rivadera.** La metodología de Kimball para el diseño de almacenes de datos (Data warehouses). [En línea] <http://www.ucasal.net/html/ingenieria/cuadernos/archivos/5-p56-rivadera-formateado.pdf>.
21. **Hernández, Yanisbel González.** *Propuesta de metodología para el desarrollo de almacenes de datos en DATEC.* Centro de Tecnología de Gestión Datos. La Habana, Cuba : s.n., 2012.
22. **Somerville, Ian.** *Ingeniería de Software.* s.l. : Prentice Hall, 2005. ISBN 8478290745.
23. Postgres SQL. [En línea] 2011. [Citado el: 20 de diciembre de 2013.] <http://postgresql.uci.cu>. Online.

24. PostgreSQL. [En línea] 25 de mayo de 2012. [Citado el: 21 de diciembre de 2013.] <http://www.postgresql.org/es/node/655>.
25. pgAdmin. [En línea] Online, enero de 2012. [http:// www.pgadmin.org/](http://www.pgadmin.org/).
26. **Levin, Jonathan**. [En línea] 20 de marzo de 2008. <http://mysqlbarbeque.blogspot.com>.
27. **Pulvirente, Adrian Sergio y Roldan, Maria Carina**. Pentaho Data Integration. [En línea]
28. [Online]. [En línea] [http://www.summan.com/pentaho/ pentaho-bi-plataform-server](http://www.summan.com/pentaho/pentaho-bi-plataform-server).
29. FreeCode. [En línea] 2012. <http://ep.freecode.com/projects/datacleaner/releases/306957>.
30. **Campazzo, Eduardo Nicolas y Santos, Virginia Ines**. *Business Intelligence:Negocios Inteligente para Empresas Inteligentes*. Argentina : Universidad nacional de la Rioja. Rene Favalaro. esq Laprida. CP5300.
31. Pentaho Open Source Business Intelligence. [En línea] \Online\. <http://www.pentaho.com>.
32. Ecured. [En línea] [http://ecured.cu/index.php/control de versiones](http://ecured.cu/index.php/control%20de%20versiones).
33. **Rendón Artola, Ariadna**. *Programa Técnico SIPAC*. [Documento] Habana : Universidad de la Ciencia Informáticas, 2013.
34. **Corporation, IBM**. *Ayuda de Rational Unified Process*. [Online]. 1987-2006. 7.0.1,.
35. **Sommerville, Ian**. *Ingeniería de Software. Séptima Edición*. Madrid : Person Educación S.A, 2005. 84 7829 074 5.
36. **Audrey Amador, Mag**. Ingeniería de Requerimientos - Reglas del Negocio. [En línea] <https://code.google.com/p/maestria/>.
37. **Elliot, J**. *Manual para la construcción de un Datawarehouse*. Lima : Publicado por el Instituto Nacional de Estadísticas e Informática (INEI), Septiembre de 1997.
38. **Basurto, Cristhian Kirs Herrera**. Apuntes Datawarehouse. *Tutorial. Quito-Ecuador*. [En línea] Autentia: Real business solutions, 2007. <http://www.adictosaltrabajo.com/tutoriales>.
39. Cubos en modelos multidimensionales. [En línea] 4 de Diciembre de 2013. <http://msdn.microsoft.com/es-es/library/ms175641.aspx>.

40. **Isabel Dapena Bosquet, Antonio Muñoz Roque, Álvaro Sánchez Miralles.** *Sistemas de Información Orientados a la Toma de Decisiones: el enfoque multidimensional.* Junio 2005.
41. **ETL-TOOLS.INFO.** Business Intelligence - Data warehousing - ETL. [En línea] 2013. [Citado el: 5 de Diciembre de 2013.] <http://etl-tools.info/>.
42. **Juan Trujillo, Emilio Soler, Eduardo Fernández Medina y Mario Piattini.** Un conjunto de transformaciones QVT para el modelado de almacenes de datos seguros. [En línea] <http://ceur-ws.org/Vol-227/paper05.pdf>.
43. Apache Tomcat. [En línea] 2013. <http://tomcat.apache.org/>.
44. SQL Eficiente. *Indices B-tree.* [En línea] 2011. [Citado el: 4 de mayo de 2014.] <http://sqleficiente.wordpress.com/2008/07/19/indices-b-tree/>.
45. DataPrix_Dimensiones lentamente cambiantes. *DataPrix.* [En línea] 19 de marzo de 2010. [Citado el: 29 de marzo de 2014.] <http://www.dataprix.com/blogs/bernabeu-dario/dimensiones-lentamente-cambiantes>.
46. **Scalone, Fernanda.** *Estudio comparativo de los modelos y estándares de calidad del software.* Buenos Aires: Universidad tecnológica nacional. : Facultad regional de Buenos Aires, 2006.

BIBLIOGRAFÍA

1. **Wiley, John.** *DATA WAREHOUSING FUNDAMENTALS. A Comprehensive Guide for IT Professionals.* s.l. : A Wiley-Interscience Publication, 2001. 0-471-41254-6.
2. **Kimball, Christopher Adamson with a foreword by Ralph.** *Mastering Data Warehouse Aggregates. Solutions for Star Schema Performance.* Canada : Wiley Publishing, Inc., 2006. ISBN-13: 978-0-471-77709-0, ISBN-10: 0-471-77709-9.
3. **Gerolami, Nicolás, Revello, Esteban y Venzal, Germain.** *Implantación de Data Warehouse Open Free.* Montevideo-Uruguay : Universidad de la República, Facultad de Ingeniería Instituto de Computación, 13 de diciembre del 2011.
4. **EVA.** *Conferencia 7: Extracción, Transformación y Carga. Sistemas de Bases de Datos II.* La Habana : Universidad de las Ciencias Informáticas, 2013-2014.
5. **Ranjan, Jayanthi.** *Business Intelligence: Concepts, components, techniques and benefits.* Ghaziabad, Uttar Pradesh, India : Institute of Management Technology, 2005-2009.
6. **EVA.** *Conferencia 8: Inteligencia de Negocio. Sistemas de Bases de Datos II.* La Habana : Universidad de las Ciencias Informáticas, 2013-2014.
7. **Gomez, Alveiro Alonso Rosado y Bautista, Dewar Willmer Rico.** *Inteligencia de Negocios: Estado del Arte.* Pereira, Colombia : Scientia Et Technica. Universidad Tecnológica de Pereira, 2014. Vol. XVI, 44. 0122-1701.
8. **EVA.** *Conferencia 6: Modelo Multidimensional.* La Habana : Universidad de las Ciencias Informáticas, 2013-2014.
9. **Cárdenas, Felipe de Jesús Núñez.** *Arquitectura de un sistema de almacén de datos.* [En línea] Agosto-Noviembre 2011. http://www.uaeh.edu.mx/docencia/P_Presentaciones/huejutla/sistemas/minería_datos/almacen_datos.pdf.
10. **Basurto, Cristhian Kirs Herrera.** *Datawarehouse.* [En línea] 30 de Octubre de 2007. <http://www.adictosaltrabajo.com/tutoriales/tutoriales.php?pagina=datawarehouse>.

GLOSARIO DE TÉRMINOS

SIPAC: Sistema de Planificación de Actividades.

Almacén de datos: gran repositorio lógico de datos que permite el acceso y la manipulación flexible de grandes volúmenes de información procedente de diferentes fuentes.

Mercado de datos: implementación de un almacén con alcance limitado a un departamento.

BI (Inteligencia de negocio): habilidad corporativa para tomar decisiones mediante el uso de metodologías, herramientas y tecnologías que permiten aplicar en ellos técnicas analíticas de extracción de información.

OLAP: Procesamiento Analítico en Línea.

SGBD (Sistema Gestor de Bases de Datos): colección de programas cuyo principal objetivo es servir de interfaz entre la base de datos, el usuario y las aplicaciones.

Jerarquía: organiza los niveles dentro de una dimensión, donde cada uno de ellos representa el total agregado de los datos del nivel inferior.

Granularidad: nivel de detalle de almacenamiento de los datos, definido particularmente para el negocio.

TCP/IP: siglas de Protocolo de Control de Transmisión/Protocolo de Internet (del inglés *Transmission Control Protocol/Internet Protocol*), sistema de protocolos que posibilita diversos servicios de red.

Cubo: conjunto de dimensiones y medidas asociadas a un tema en específico.

Hecho: representa la ocurrencia de un proceso específico en el interior de la organización.

Dimensión: perspectiva mediante la cual se puede llevar a cabo el análisis sobre el hecho.

Medidas: constituye un valor o indicador de análisis del hecho.

CMMI: Capability Maturity Model Integration.