

Universidad de las Ciencias Informáticas

Facultad 3



**Dashboard para el Sistema de Planificación de Actividades
(SIPAC)**

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

Autora:

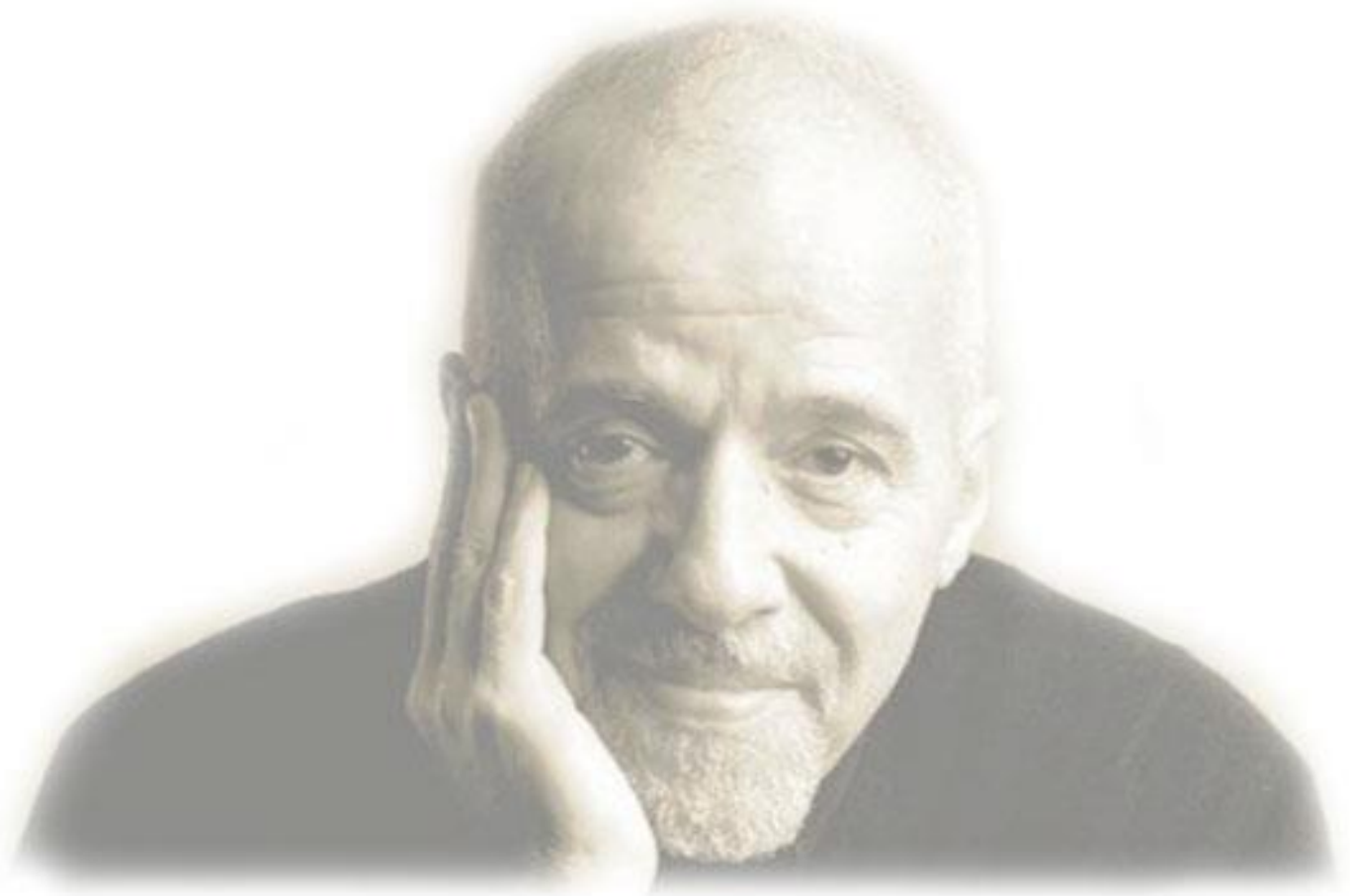
Veronica Martha Milanés Valero

Tutora:

Ing. Maria Teresa Rosales González

La Habana, junio del 2016

“Año 58 de la Revolución”



Sólo hay una cosa que hace que un sueño sea imposible de lograr: el miedo al fracaso.

Paulo Coelho.



DECLARACIÓN DE AUTORÍA

Declaro ser el único autor de la presente tesis **Dashboard para el Sistema de Planificación de Actividades (SIPAC)** y reconozco a la Universidad de las Ciencias Informáticas los derechos patrimoniales de la misma, con carácter exclusivo.

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

Firma del autor

Veronica Martha Milanés Valero

Firma del Tutor

Ing. Maria Teresa Rosales González

DATOS DE CONTACTOS

Nombre y Apellidos: Ing. Maria Teresa Rosales González

Correo electrónico: mtrosales@uci.cu

Año de graduado: 2012

Nombre y Apellidos: Veronica Martha Milanés Valero

Correo electrónico: vmmilanes@estudiantes.uci.cu

Mi primer agradecimiento y más especial va dirigido a mis dos tías Ismary y Martha, que a pesar de no estar hoy físicamente espero que donde quiera que estén se sientan orgullosas de mí.

A mi mamá Rosa por ser la principal impulsora de este sueño y por brindarme su apoyo incondicional.

A mi papá Juan Carlos por estar siempre presente y por ser el mejor padre del mundo.

A madrastra Dunia que ha sido como otra madre para mí y que siempre me ha apoyado en todo momento.

A mi abuela María del Carmen por ser esa persona especial que todos tenemos y a quien amamos con toda el alma.

A mis hermanos Ana María, Carlos, Carmen y Félix por ser lo más grande que tengo mi vida.

A mis sobrinas Yulsainer y Analía que han sido lo más hermoso que me ha podido regalar mi hermana.

A toda mi familia, especialmente a mis tías Damaris, Regla y Miriam por preocuparse siempre por mí.

A mis tres amigos Arianna, Leyriel y Yania que han tenido que soportarme y aguantarme durante todo este año, pero que no se les olvide que yo soy la primera ingeniera!!!

A mi novio Rosbel, a pesar de que tuvimos poco tiempo para conocernos has sido alguien importante para mí en esta etapa que tanto apoyo necesité.

A mi tutora por su apoyo incondicional y por tener paciencia con mi pérdida de memoria.

A mis nuevos amigos Yenisel, Ángel y Tellez que llegamos este año y le hemos poblado la casa.

Quiero agradecer a mis amigos de la Facultad de Ciego de Ávila que a pesar de la distancia les deseo lo mejor y que se gradúen con éxito.

A la UCI por permitirme realizar mi sueño de graduarme en esta universidad.

A todas las personas de las cuales estoy muy agradecida por la influencia positiva que tuvieron sobre mí para que este gran sueño se lograra. Mil gracias para todos.

Dedico los resultados de este trabajo especialmente a mis dos tías, las cuales han sido mi principal guía de inspiración.

A mis padres por su apoyo y amor incondicional, este logro también es de ustedes espero que se sientan muy orgullosos de mí.

Resumen

La Inteligencia de negocios permite aprovechar los datos operacionales y convertirlos en información valiosa en aras de establecer estrategias, generar reportes y obtener pronósticos que mejoren el proceso de toma de decisiones incrementando la efectividad de las instituciones (1). En la Universidad de las Ciencias Informáticas (UCI) se desarrolla el Sistema de Planificación de Actividades (SIPAC), el cual se basa en la Instrucción No. 1 del Presidente de los Consejos de Estado y de Ministros para la planificación de objetivos y actividades en Cuba. 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 como parte de la planificación a corto, mediano y largo plazo. Independientemente de las funcionalidades que el sistema brinda, no almacena la información histórica y unificada en forma de indicadores de los procesos de planificación, lo que provoca demora en el análisis de los datos.

En la presente investigación se realiza una propuesta de solución basada en la Inteligencia de negocios mediante la implementación de un Dashboard que permite mostrar la información de forma dinámica, apoyada en los indicadores claves de desempeño (KPI). Para guiar el desarrollo del producto se seleccionó la Metodología de desarrollo para proyectos de almacenes de datos, del Centro de tecnologías y gestión de datos (DATEC). Como resultado se obtuvo un Dashboard que permite visualizar la información mediante reportes candidatos y gráficos dinámicos contribuyendo a las decisiones que toman los directivos en las entidades.

Palabras Claves: Dashboard, indicadores claves de desempeño, Inteligencia de negocio, toma de decisiones.

Índice General

Introducción	1
Capítulo 1. Fundamentación teórica del Dashboard para SIPAC	4
1.1 Inteligencia de negocio(BI)	4
1.2 Modelo multidimensional	5
1.3 Proceso analítico en línea.....	5
1.4 Implementaciones OLAP	6
1.5 Antecedentes de los Dashboard.....	7
1.6 Dashboard.....	7
1.7 Indicadores clave de desempeño o KPIs	10
1.8 Integración de datos	10
1.9 Estudio de sistemas similares	11
1.10 Valoración de los sistemas	13
1.11 Metodología de desarrollo.....	14
1.12 Herramienta de modelado.....	15
1.13 Sistema gestor de bases de datos (SGBD)	16
1.14 Servidor web.....	16
1.15 Herramientas para la integración de datos	17
1.16 Herramientas para la Inteligencia de negocio	17
1.17 Conclusiones del capítulo	19
Capítulo 2. Análisis y Diseño del Dashboard para SIPAC	20
2.1 Definición del negocio	20

2.2	Tema de análisis	20
2.3	Reglas del negocio.....	21
2.4	Necesidades de los usuarios.....	21
2.5	Casos de uso del sistema	24
2.6	Diagrama de caso de uso del sistema	26
2.7	Descripción de los casos de usos del sistema.....	26
2.8	Arquitectura del sistema	28
2.9	Diseño del subsistema de integración de datos	29
2.10	Diseño del subsistema de almacenamiento.....	31
2.11	Diseño del subsistema de visualización	34
2.12	Seguridad del sistema.....	34
2.13	Conclusiones del capítulo	35
Capítulo 3: Implementación y Prueba del Dashboard para SIPAC.		36
3.1	Implementación del subsistema de almacenamiento	36
3.2	Implementación del subsistema de integración de datos.....	38
3.3	Implementación del subsistema de visualización	45
3.4	Pruebas de software.....	51
3.5	Modelo V	52
3.6	Herramientas para la aplicación de pruebas	53
3.7	Calidad de los datos	56
3.8	Resultado de las pruebas.....	57
3.9	Conclusiones del capítulo	58

Conclusiones Generales	59
Recomendaciones	60
Referencias	61
Anexos	66

Índice de Tablas

Tabla 1. Resumen de los sistemas similares.	13
Tabla 2. Descripción y permiso de los actores del sistema.....	25
Tabla 3. Descripción del caso de uso de información: Presentar información de Elaboración del Plan.	27
Tabla 4. Descripción del hecho.....	32
Tabla 5. Descripción de las dimensiones.....	32
Tabla 6. Descripción de la medida.	32
Tabla 7. Estándares de codificación.	36
Tabla 8. Modelo de datos físico.....	37
Tabla 9. Caso de prueba asociado al caso de uso de información: “Presentar información de Elaboración del Plan”.	53
Tabla 10. Aplicación de las listas de chequeo a los artefactos ETL.....	54
Tabla 11. Lista de chequeo aplicada al Perfilado de datos.....	55

Índice de Figuras

Figura 1. Ciclo de vida de la metodología (27).	14
Figura 2. Diagrama de caso de uso del sistema.....	26
Figura 3. Arquitectura del Sistema.....	29
Figura 4. Calidad y tipos de datos.....	30
Figura 5. Diseño del subsistema de integración.....	31
Figura 6. Modelo de datos dimensional.	33
Figura 7. Arquitectura de Información.	34
Figura 8. Estructura física de la BD.	38
Figura 9. Componente Insertar\Actualizar.	42
Figura 10. Transformación de la dimensión: dim_estado_aprobacion.....	43
Figura 11. Transformación hecho actividad.	43
Figura 12. Trabajo del hecho actividad.	44
Figura 13. Diseño del cubo y elementos que lo componen.	45
Figura 14. Cubo hecho actividad.....	46
Figura 15. Medida total de actividades del hecho actividad.	46
Figura 16. Dimensión categoría actividad.....	46
Figura 17. Jerarquía de la dimensión categoría de actividad.....	47
Figura 18. Nivel de la jerarquía categoría de actividad.....	47
Figura 19. Consulta MDX.	48
Figura 20. Vista del reporte: “Total de actividades por categoría”	48
Figura 21. Visualizar información por año.	49

Figura 22. Visualizar la información solicitada.....	50
Figura 23. Visualizar información por mes.	50
Figura 24. Modelo V (56).....	52
Figura 25. Calidad de los datos después de realizado el proceso ETL.....	56
Figura 26. Comportamiento de los indicadores medidos en las pruebas unitarias.....	57
Figura 27. Comportamiento de los indicadores medidos en las listas de chequeo.	57

Introducción

En la actualidad las Tecnologías de la Información y las Comunicaciones (TIC), han pasado a ocupar un lugar primordial en el desarrollo social y económico de cualquier país. Estas se han convertido para el área de los negocios en un elemento estratégico en cuanto a términos de competitividad para el entorno empresarial por los múltiples beneficios que brindan y el volumen de datos que son capaces de generar.

Desde hace varias décadas las empresas a nivel mundial han buscado en el almacenamiento de datos soluciones que ayuden a la toma de decisiones en el contexto empresarial, pues el nivel competitivo en esta área aumenta de forma gradual. No obstante, alcanzar el éxito en el ámbito de los negocios requiere de un análisis de datos y respuestas inmediatas, con el fin de cumplir con las demandas de los clientes. Por lo que en la década de los 90, surgen las herramientas de análisis y reportes, creándose la tendencia de generar reportes personalizados (2). Estas herramientas en un inicio resultaron ser muy complejas en su utilización, pues la información obtenida era dada en uno o varios reportes con un gran cúmulo de datos, lo que imposibilitaba el manejo eficiente de dicha información. Como solución a esta problemática, aparece la herramienta informática conocida como Dashboard.

Los Dashboard muestran la información más importante, presentada en una manera que permite hacer un seguimiento de lo que está ocurriendo en un instante de tiempo. Para servir a su propósito y maximizar sus prestaciones, deben mostrar abundante información en una pequeña cantidad de espacio de manera que se comunique con claridad e inmediatez. Esto requiere un diseño que se nutra y aproveche el poder de la percepción visual para lograr el procesamiento de grandes cúmulos de información (3).

En aras de aprovechar los beneficios aportados por la industria del software, y de la utilización de las nuevas tecnologías, surge la Universidad de las Ciencias Informáticas (UCI), que cuenta con diversos centros productivos dedicadas al desarrollo de soluciones informáticas. Uno de ellos es el Centro de Informatización de Entidades (CEIGE) perteneciente a la Facultad 3, en la cual se desarrolla el proyecto Sistema de Planificación de Actividades (SIPAC).

SIPAC facilita la implementación de la Instrucción No. 1¹ del Presidente de los Consejos de Estado y de Ministros para la planificación de los objetivos y actividades a todos los niveles organizacionales. Además,

¹ 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 del Estado (OACE), entidades nacionales y las administraciones locales del Poder Popular.

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 como parte de la planificación a corto, mediano y largo plazo. Cuenta con varios módulos y componentes (Configuración, Planeación, Notificaciones y Recuperaciones).

A pesar de todas las ventajas que brinda SIPAC, no almacena la información histórica y unificada en forma de indicadores de los procesos de planificación, lo que provoca demora en el análisis de los datos o el abandono de dicho análisis, incidiendo negativamente en la eficiencia del trabajo de los planificadores y directivos de las entidades. Por lo que las decisiones que toman los directivos, en su mayoría, no están respaldadas por un proceso previo de análisis y reflexión basado en indicadores reales de la ejecución de los procesos. Lo que trae como consecuencia la dirección del trabajo improvisada y operativa, no encaminada al cumplimiento de los objetivos de las organizaciones.

Teniendo en cuenta los elementos planteados anteriormente se identifica como **problema a resolver**: ¿Cómo contribuir al análisis de la información de los datos generados por SIPAC?

Con vista a la solución del problema se define como **objeto de estudio**: Soluciones de Inteligencia de negocios. La presente investigación se encuentra enmarcada en el **campo de acción**: Dashboard. Trazándose como **objetivo general**: Desarrollar un Dashboard que contribuya al análisis de la información de los indicadores de planificación generados por SIPAC. En correspondencia con el objetivo general se plantean los siguientes objetivos específicos:

- Elaborar el marco teórico de la investigación para sustentar conceptos, propuesta de desarrollo, herramientas y tecnologías a utilizar.
- Realizar el análisis y diseño del Dashboard para SIPAC teniendo en cuenta las necesidades del cliente.
- Implementar el Dashboard para SIPAC que responda a las necesidades del cliente.
- Validar el Dashboard para SIPAC mediante las pruebas de caja negra y la aplicación de las listas de chequeo.

Para encaminar la investigación en vista a resolver el problema planteado se propone la siguiente **idea a defender**: Si se desarrolla el Dashboard se contribuirá al análisis de la información de los indicadores de planificación generados por SIPAC.

Estructura Capítular

Capítulo 1: Fundamentos teóricos

Se definen los conceptos necesarios para entender el objetivo fundamental del trabajo, se detallan las tendencias y tecnologías actuales, metodologías y herramientas.

Capítulo 2: Análisis y Diseño del sistema.

Se realiza el análisis para comprender el negocio, del que se definen los requisitos de información, los funcionales y los no funcionales, así como las reglas del negocio. Se conforma el diagrama de casos de uso y se construye el modelo de datos donde se determinan: las dimensiones, los hechos y las medidas.

Capítulo 3: Implementación y Prueba.

Comprende la implementación de los subsistemas de almacenamiento, integración y visualización. Se exponen las pruebas utilizadas para la validación de la solución propuesta, así como los resultados arrojados.

Capítulo 1. Fundamentación teórica del Dashboard para SIPAC

En el presente capítulo se definen una serie de conceptos necesarios para entender el objetivo fundamental del trabajo, se detallan las tendencias y tecnologías actuales, metodologías y herramientas para el desarrollo de la solución propuesta.

1.1 Inteligencia de negocio²(BI)

La información es la clave para obtener una ventaja competitiva en el actual entorno empresarial. Las empresas necesitan poder acceder a sus datos de forma rápida y flexible. Una manera de lograr esto es por medio de la Inteligencia de negocios, que es la encargada de analizar los datos acumulados en una empresa y extraer el conocimiento de ellos, contribuyendo así con el proceso de toma de decisiones (4).

El Instituto de Almacenes de Datos³, define la Inteligencia de negocios (BI) como:

“...son los procesos, tecnologías, y herramientas que se necesitan para convertir los datos en información, la información en conocimiento, y el conocimiento en planes que impulsan acciones rentables para el negocio. La Inteligencia de negocios abarca el almacenamiento de datos, herramientas analíticas, y contenido y gestión del conocimiento...” (5).

Por lo que se puede decir que la Inteligencia de negocio actúa como un factor estratégico para una empresa u organización, generando una potencial ventaja competitiva, que no es otra que proporcionar información privilegiada para responder a los problemas de negocio (4).

Beneficios de la Inteligencia de negocio

En el siguiente apartado se mencionan los principales beneficios que presenta la BI para el entorno empresarial (6):

- Respuestas inmediatas a preguntas del negocio, que son básicas para la toma de decisiones.
- Integración de datos entre los diferentes sistemas de información existentes en la organización.
- Permite lograr una visión del futuro a través del análisis de datos históricos.

² Del inglés: Business Intelligence.

³ El Data Warehousing Institute proporciona educación, capacitación, certificación, noticias, e investigación para ejecutivos y tecnología de la información (TI) en todo el mundo. Fundada en 1995, es la institución educativa más importante para la inteligencia empresarial y almacenamiento de datos

- Logra medir el desempeño de la organización en función de sus metas y la industria en donde compite.
- Permite al cliente tener acceso en tiempo real a los datos posibilitando una mejor información y satisfacción para el mismo.
- Brinda a sus usuarios las herramientas para tomar mejores decisiones, proporcionando respuestas rápidas a las preguntas de los mismos.

1.2 Modelo multidimensional

La eficiente manipulación de la información se ha convertido en un hecho indispensable y fundamental para la toma de decisiones de cualquier organización. Para lograr la misma, es necesario unificar los datos que han sido tomados de distintas fuentes para posteriormente almacenarlos, por lo que surge el término modelo multidimensional. Los modelos multidimensionales proporcionan dos conceptos fundamentales hecho y dimensión, siendo el cubo el objeto central:

- **Tabla de Hechos:** es la tabla central en un esquema dimensional. Generalmente, almacena medidas numéricas, las que representan valores de las dimensiones. La tabla de hechos está constituida por medidas y por llaves foráneas. Se considera el evento específico que constituye la unidad fundamental de análisis de datos para la toma de decisiones (7).
- **Tabla de Dimensiones:** especifican la organización lógica de los datos y proporcionan el medio para analizar el contexto del negocio. Representan los ejes del cubo, y los elementos del análisis, proporcionándole al usuario el filtrado y manipulación de la información almacenada en la tabla de hechos (7). Existe una dimensión fundamental en todo almacén de datos: la dimensión tiempo, pues todo registro que se incluya constituye la ocurrencia de un fenómeno en un instante de tiempo definido (7).
- **Cubo:** es una estructura multidimensional, 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 (7).

1.3 Proceso analítico en línea

El Procesamiento analítico en línea (OLAP⁴) es una tecnología que se basa en el análisis multidimensional de los datos y que le permite al usuario tener una visión más rápida e interactiva de los mismos. Posee una

⁴ Del inglés: Online Analytical Processing, conocido por sus siglas OLAP.

gran capacidad para realizar cálculos de múltiples dimensiones, lo que permite gran variedad de informes y análisis de grandes volúmenes de datos, para ello utiliza estructuras multidimensionales o cubos OLAP que contienen datos resumidos de grandes bases de datos (8) .

Las principales características de OLAP son (8) :

- **Rápido:** el sistema debe ser capaz de responder de una forma rápida y ágil a la información que le sea solicitada por el usuario.
- **Análisis:** el sistema debe poder reflejar cualquier lógica del negocio para poder responder a las preguntas específicas y necesidades empresariales.
- **Compartido:** el sistema deberá proporcionar herramientas que garanticen la confidencialidad de los datos y la seguridad de acceso por perfiles de los usuarios.
- **Multidimensional:** la herramienta deberá proporcionar soporte a cada una de las múltiples jerarquías que puedan existir dentro de la organización de información.
- **Información:** son todos los datos e información derivada de este proceso de análisis, la cual permitirá la toma de decisiones.

1.4 Implementaciones OLAP

La tecnología OLAP posee tres tipos de implementaciones, las que se diferencian teniendo en cuenta sus características. Dichas implementaciones son:

- **Procesamiento analítico relacional en línea (ROLAP):** 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 Base de datos relacional (9).
- **Procesamiento analítico multidimensional en línea (MOLAP):** almacena físicamente los datos en estructuras multidimensionales de modo que las representaciones externas e internas concuerden. Entre sus principales ventajas se encuentran que ostenta una tecnología perfeccionada para las consultas y el análisis basado en el modelo multidimensional, aunque no permite trabajar con gran cantidad de tablas de dimensiones y también que las herramientas diseñadas para esta son restringidas y propietarias (9).
- **Procesamiento analítico híbrido en Línea (HOLAP):** sistema híbrido entre las arquitecturas ROLAP y MOLAP, que combina estas dos implementaciones para almacenar algunos datos, como los registros de

detalle, en un motor relacional y otros, como las agregaciones, en un almacén multidimensional. Brinda una solución con las mejores características de ambas: desempeño superior y gran escalabilidad (9) .

Una vez analizados los diferentes tipos de almacenamiento se selecciona ROLAP, pues se tuvo en cuenta el Sistema Gestor de Base de Datos (SGBD) seleccionado PostgreSQL que solo soporta el almacenamiento relacional. Además, ROLAP permite ahorrar espacio de almacenamiento, es escalable pues permite adicionar nuevas dimensiones a las ya existentes y es capaz de manejar conjuntos de datos muy grandes (10).

1.5 Antecedentes de los Dashboard

El antecedente de los Dashboard son los Cuadros de Mando Integral (CMI). El concepto de CMI fue presentado en el número de Enero/Febrero de 1992 de la revista Harvard Business Review en un artículo denominado “The Balance Scorecard”. Sus autores, Robert S. Kaplan⁵ y David P. Norton⁶, plantean que:

“El CMI es un sistema de administración o sistema administrativo que va más allá de la perspectiva financiera con la que los gerentes acostumbran evaluar la marcha de una empresa” (11) .

En la actualidad, no todos los Cuadros de Mando Integral están basados en los principios de Kaplan y Norton, aunque sí influenciados en alguna medida por ellos. Por este motivo, se suele emplear con cierta frecuencia el término Dashboard o Cuadro de Mando, que utiliza algunas características teóricas del Cuadro de Mando. De forma genérica, un Dashboard engloba a varias herramientas que muestran información relevante para la empresa a través de una serie de indicadores de rendimiento, también denominados KPIs⁷ (*key performance indicator*) (12).

1.6 Dashboard

Los Dashboard están diseñados para ayudar a interpretar el estado instantáneo de un elemento de control. Para controlar con rapidez el estado de cualquier sistema, los Dashboard deben ser diseñados de manera particular para aprovechar los puntos fuertes de la percepción visual y el conocimiento (13) . Según Stephen

⁵ Profesor de la Universidad de Harvard.

⁶ Consultor empresarial de Boston.

⁷ Del inglés: indicadores claves de desempeño.

Few⁸ un Dashboard se puede definir como “Una presentación visual de información importante, necesaria para lograr uno o más objetivos; consolidada y arreglada en una sola pantalla, de tal manera que la información pueda monitorearse con un vistazo.” (3).

Los Dashboard utilizan una metodología centrada en el usuario que integra datos de acuerdo con los problemas, funciones principales o procesos comerciales críticos de la empresa. Constituyen una guía para transformar los propósitos y la estrategia de una organización en objetivos e indicadores concretos. Además, facilitan monitorear el estado del negocio y no solo son capaces de medir el momento actual sino los períodos anteriores, lo que permite establecer una comparación de la evolución de la organización. Simultáneamente, posibilita mostrar resúmenes y mediciones detalladas, donde los datos se actualizan automáticamente en forma diaria, semanal o mensual y son utilizados por los empleados de toda la organización (2).

Características de los Dashboard

El uso de los Dashboard es amplio y llega a diferentes tipos de organizaciones y distintos temas de interés. Ellos tienen un conjunto de elementos que los caracterizan, dentro de los principales se encuentran los siguientes (2):

- **Número de KPIs:** se deben mostrar los KPI necesarios, los cuales deben ser lo más específicos posibles, posibilitando la toma de decisiones en el entorno empresarial.
- **Segmentación y contexto:** debe presentar los KPIs de forma que estas sean relevantes para el negocio.
- **Visualización:** la persona que toma las decisiones debe ser capaz de interpretar fácilmente la información que está viendo. Por lo que el Dashboard debe ser breve y su representación gráfica la adecuada para los datos que representa y lo suficientemente visual, para que resulte atractivo su estudio.
- **Análisis:** además de los KPIs, el Dashboard debe acompañarse de un análisis sobre lo ocurrido, las recomendaciones dadas y su potencial impacto sobre el negocio. El análisis debe recomendar acciones, no describirlas.

⁸ Presidente de Perceptual Edge y experto en visualización.

Beneficios de la utilización de los Dashboard (14)

- Son herramientas que despliegan la información más importante para el usuario o área, en el tiempo y formato adecuado a sus requerimientos.
- Cuentan con una interfaz de usuario amigable, principalmente vía Web.
- La flexibilidad en su arquitectura y desarrollo permite contar con una variedad de soluciones, dependiendo de si la información es para monitoreo, análisis o administración del desempeño.
- Ayudan a comunicar la estrategia de la organización mediante el despliegue de objetivos estratégicos, indicadores, metas, procesos y proyectos.
- Incentivan un mejor desempeño del personal mediante la premisa de “lo que se mide, se hace”. Los trabajadores pueden ser motivados a la excelencia en sus funciones, trayendo consigo premios por superar sus objetivos y metas.
- Facilitan el despliegue de una vista consistente del negocio, al consolidar e integrar información de diversas fuentes usando definiciones, reglas e indicadores comunes.

Ventajas de los Dashboard (15):

- Permiten compartir, agrupar y centralizar los datos clave de la organización para posibilitar que toda la información esté de forma accesible.
- Mediante gráficos dinámicos, selectores y diagramas interactivos, la presentación de indicadores clave del rendimiento (KPIs) facilita la toma de decisiones de una manera intuitiva y muy gráfica.
- El acceso a los indicadores permite una reacción rápida de cara a los cambios de la organización.
- Permiten el acceso de múltiples usuarios, posibilitando que toda la organización tenga acceso al sistema en cualquier instante de tiempo.

Desventajas de lo Dashboard (16):

- Si los indicadores (KPIs) no se escogen con cuidado, no se comunica con claridad el mensaje que se quiere transmitir, imposibilitando una correcta toma de decisiones.
- La información visualizada ocupe más de una página mostrando información innecesaria y creando ambigüedades.

- Son sistemas de control empresarial y no herramientas de aprendizaje.

1.7 Indicadores clave de desempeño o KPIs

Los indicadores clave de desempeño son medidas cuantificables que son de total importancia para el éxito de una entidad. Estos indicadores varían entre organizaciones e industrias, pero siempre deben, si se implementan y monitorean correctamente, ayudar a la organización a definir y medir el progreso hacia los objetivos de largo y corto plazo (17). En la mayoría de los casos, un KPI es el estado de un indicador, es decir, si está por encima o por debajo de una meta predeterminada. Los KPIs generalmente se muestran como una tasa o porcentaje y están diseñados para permitir que un usuario de negocios pueda saber instantáneamente si están dentro o fuera de su plan sin que tenga que buscar información adicional (18).

1.8 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 almacén de datos (19). Para llevar a cabo la integración se realiza el proceso de Extracción, Transformación y Carga (ETL⁹) el cual se explica a continuación.

Proceso de Extracción, Transformación y Carga

El ETL permite extraer los datos desde los sistemas fuentes, garantiza la calidad de los mismos y hace cumplir las reglas del negocio 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 (20).

Extracción: la extracción es el primer paso de la obtención de datos en el medio ambiente de almacenamiento de datos. Consiste en sustraer los datos brutos desde las fuentes de origen con el objetivo de integrarlos y estandarizarlos (21).

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 (21).

⁹ Del inglés: Extract, Transform and Load.

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 (21).

Fundamentalmente el ETL permite realizar funcionalidades como (20):

- Eliminar errores y corregir datos faltantes.
- Capturar el flujo de datos transaccionales para su protección.
- Concordar los datos de múltiples fuentes para ser usadas conjuntamente.
- Estructurar los datos para ser usables por las herramientas del usuario final.

1.9 Estudio de sistemas similares

En la actualidad son varias las organizaciones que hacen uso de los Dashboard. Estos son diseñados para lograr un análisis de la información basado en indicadores a nivel empresarial. A continuación, se hace un análisis detallado de estos sistemas.

Urban Dashboard

Permite explorar y comparar más de 150 indicadores cuantitativos, encuestas de opinión pública y mapas interactivos de ciudades intermedias de América Latina y el Caribe. Se trata de ciudades de crecimiento demográfico y económico sobresaliente, que reciben asistencia técnica de la Iniciativa de Ciudades Emergentes y Sostenibles (ICES) (22).

Beneficios (22):

- Permite elegir una ciudad para comparar su desempeño en relación con las demás y monitorear la evolución en el tiempo de sus indicadores.
- Los mapas interactivos presentan datos geospaciales generados a partir del análisis multi-temporal de imágenes satelitales.
- Los datos son actualizados con periodicidad a medida que se compila nueva información.
- Bajo la filosofía de los datos abiertos, el Urban Dashboard ofrece información a todos los ciudadanos para colaborar en la planificación de los gobiernos y promover la transparencia y la participación.

ScytI Control Dashboard

Su función principal es la de ofrecer un resumen gráfico de toda la configuración electoral. Permite que las autoridades electorales gestionen y supervisen fácilmente los procesos electorales, facilitándoles una visión

clara del estado y de la organización de las elecciones. Todos los datos se actualizan en tiempo real, de forma que se puede proporcionar asistencia y soporte según sea necesario (23).

Beneficios (23):

- **Panel diseñado para las elecciones:** es el único software del mercado de agrupación de datos y planificación de proyectos que es específico para elecciones.
- **Mayor control y eficiencia:** gracias a su visualización gráfica de la información crítica, la plataforma permite que todo el personal verifique al instante que todos los puntos de control estén listos antes de abrir una elección.
- **Personalizado para satisfacer sus necesidades:** las autoridades electorales varían considerablemente de un país a otro. ScytI Control Dashboard está personalizado para cada cliente, a fin de asegurar que la herramienta se adapte a sus necesidades.
- **Información en tiempo real:** proporciona a las autoridades electorales una imagen precisa del proceso electoral, la cual se actualiza en tiempo real según van sucediendo los acontecimientos durante el proceso electoral.

News & Dashboard Control

Es una aplicación diseñada para propósitos de monitoreo de noticias, publicidad, registro y manejo de crisis para Windows XP o superior. Captura la opinión de profesionales del medio, escritores y otros colaboradores sobre noticias de actualidad y publicidad. Permite monitoreo a medios como la televisión, radio y prensa escrita (24).

Beneficios (24):

- News Control registra las noticias de publicidad y News Dashboard permite analizar toda la información capturada mediante gráficos de tipo pastel y gráficos de barras, entre otras herramientas visuales.
- News & Dashboard Control permite a las empresas tomar las decisiones más adecuadas, así como definir sus estrategias comerciales y competitivas, a favor de una mejor comunicación corporativa con sus clientes, aun en un manejo de crisis.

Sixtina Dashboard

Es una efectiva herramienta de visualización, que permite que los tomadores de decisiones accedan a información clave y puedan analizarla en tiempo real. Es una herramienta preparada para obtener una efectiva y completa visualización de toda la información empresarial cuando el cliente la necesite. Cuenta con información rápida, de fácil comprensión y acceso, permitiendo que las decisiones empresariales se tornen más sólidas (25).

Beneficios (25):

- Permite visualizar información compleja en atractivos gráficos de datos que le darán una visión instantánea de la situación en que se hallan todas las áreas bajo su responsabilidad. Algunos gráficos que puede utilizar son: gráfico de torta, barras, barras apiladas, radares, histogramas, mapas estratégicos, mapas y gráficos de tiempo.

Tabla 1. Resumen de los sistemas similares.

Sistemas Estudiados	Indicadores	Tipo de Gráfico	Información en Tiempo Real
Urban Dashboard	Desempeño entre ciudades y evolución en el tiempo.	Gráficos y mapas interactivos.	No
Scytl Control Dashboard	Gestión de los procesos electorales y supervisión de los procesos electorales	Gráficos dinámicos.	Sí
News & Dashboard Control	Indicadores de publicidad, registro de noticias.	Gráficos de pastel, gráficos de barras.	No
Sixtina Dashboard	Indicadores empresariales.	Gráfico de torta, barras, radares, histogramas, y gráficos de tiempo.	Sí

1.10 Valoración de los sistemas

Después de un minucioso estudio de los Dashboard existentes en el mundo se llega a la conclusión de que en la actualidad es de vital importancia el empleo de dichos sistemas para el análisis de la información basada en indicadores dentro de una organización donde hay cantidad y diversidad de información. A pesar de las ventajas que presentan, ninguno se ajusta al modelo que persigue SIPAC, que es la planificación de los objetivos y actividades a todos los niveles organizacionales. Pues se especializan en áreas como el marketing y la supervisión de indicadores de diversos medios (radio, TV y prensa escrita). Además, son herramientas propietarias o forman parte de otro sistema.

No obstante, el estudio de los distintos Dashboards permitió saber cuáles son los gráficos que más se utilizan con el objetivo de diseñar una interfaz amigable, fácil de manejar y que cumpla con las tendencias actuales a las cuales los usuarios están adaptados. Por lo que se decide utilizar los gráficos de columna, de barras, los circulares y los de líneas. Otro de los resultados obtenidos fue la evaluación del comportamiento de la información en tiempo real, con el objetivo de determinar las características que se deben cumplir. Los resultados de la investigación manifiestan que el Dashboard para SIPAC no mostrará la información en tiempo real pues se necesitaría de tiempos de respuestas muy rápidos y complejos por parte del servidor, por lo que se decide actualizar la información de forma semanal con el objetivo que el sistema trabaje de forma estable.

1.11 Metodología de desarrollo

Las metodologías de desarrollo de software representan un marco de trabajo que tiene entre sus funciones estructurar, controlar, guiar, planificar, manipular y dirigir el proceso de desarrollo de sistemas de información. Surgen ante la necesidad de trabajar mediante el uso de procedimientos, técnicas, herramientas y documentos durante el desarrollo del software (26).

Para guiar el proceso de desarrollo de la solución se selecciona la Metodología para el Desarrollo de Almacenes de Datos que propone el Centro de Tecnologías de Gestión de Datos (DATEC). Se decidió aplicar este modelo debido a que el ciclo de vida está organizado por fases y algunas de ellas podrán ser implementadas de forma paralela, lo que proporciona que tenga un desarrollo ágil. Además, es una metodología mixta que reúne elementos de varias metodologías de desarrollo de proyectos de integración de datos, tomando como base la Metodología de Kimball e incluye los principios básicos que permiten una adecuada gestión del proyecto. La Figura 1 recrea el ciclo de vida de dicha metodología:

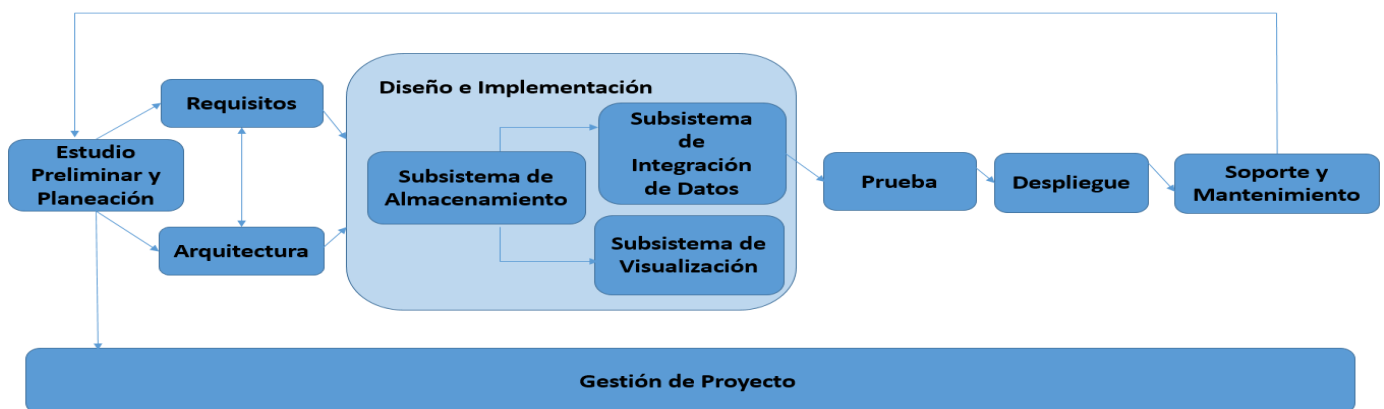


Figura 1. Ciclo de vida de la metodología (27).

Descripción de las fases de la metodología (27):

- **Gestión del proyecto:** constituye un flujo de trabajo que se ejecuta a lo largo de todo el ciclo de vida del proyecto. Está compuesto por un grupo de procesos que se encargan de mantener la adecuada gestión del proyecto a partir de la aplicación de conocimientos, habilidades, herramientas y técnicas.
- **Estudio preliminar y planeación:** se efectúa el diagnóstico de información, de datos y de infraestructura tecnológica. Además se elaboran las tareas de planeación del proyecto.
- **Levantamiento de requisitos:** se identifican las necesidades de información de los clientes, las reglas del negocio y los requisitos.
- **Arquitectura:** se define la arquitectura según los requisitos no funcionales (RNF). Esta fase puede desarrollarse paralelamente con la fase de levantamiento de requisitos.
- **Diseño e implementación:** se realiza el diseño de los subsistemas partiendo de las estructuras de almacenamiento. Conjuntamente se realizan las actividades de implementación de cada uno de los subsistemas.
- **Prueba:** se aplican diversas pruebas a la solución. Esta fase no es la única en la que se realizan pruebas durante el desarrollo del proyecto, pues en todas las fases hay actividades de aseguramiento de la calidad.
- **Despliegue:** se presenta la solución desplegada en el entorno real y en correcto funcionamiento.
- **Soporte y mantenimiento:** esta fase se inicia cuando la solución está implantada y en explotación, y se ejecuta según el contrato firmado y las condiciones de soporte establecidas.

En la solución se llevaron a cabo cinco de las ocho fases que presenta la metodología, las cuales son: estudio preliminar y planeación, requisitos, arquitectura, diseño e implementación de los subsistemas de almacenamiento e integración y por último la etapa de prueba.

1.12 Herramienta de modelado

Con la evolución de la industria del software se ha generado la necesidad de desarrollar herramientas para el modelado de sistemas, que faciliten de esta forma el trabajo de analistas y diseñadores de software. Como herramienta para el modelado de la solución se empleará Visual paradigm for UML¹⁰ 8.0 Enterprise Edition por las características que este presenta.

¹⁰ Lenguaje gráfico para visualizar, especificar, construir y documentar un sistema.

Visual paradigm for UML 8.0 enterprise edition

Visual paradigm es una herramienta CASE¹¹ profesional, que soporta el ciclo de vida completo de desarrollo de software: análisis y diseño orientados a objetos, construcción, pruebas y despliegue. Permite dibujar todos los tipos de diagramas de clases, código inverso, generar código desde diagramas y generar documentación (28). Es una herramienta colaborativa, es decir, soporta múltiples usuarios trabajando sobre el mismo proyecto y permite el control de versiones.

1.13 Sistema gestor de bases de datos (SGBD)

Un SGBD es un conjunto de programas que permiten el almacenamiento, modificación y extracción de la información en una base de datos, además de proporcionar herramientas para añadir, borrar, modificar y analizar los datos (29).

Postgres 9.4

Constituye el SGBD objeto-relacional de código abierto más desarrollado del mundo y opera bajo licencia BSD¹², sólo requiere que el código fuente licenciado mantenga la información de derechos de autor y licenciamiento. Es multiplataforma, ofrece una documentación bien organizada, pública y libre, presenta conectores de datos foráneos que permiten 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 (30).

PgAdminIII 1.20

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 (31).

1.14 Servidor web

Un servidor web es un programa que se ejecuta ininterrumpidamente en una computadora, en espera de peticiones que hagan los usuarios de la red. Estas peticiones son respondidas a través del navegador de forma adecuada en correspondencia con la solicitud (32). Siguiendo las características del proyecto SIPAC se utilizará como servidor web Apache tomcat 6.0, el cual se caracteriza por su

¹¹ Ingeniería de Sistemas Asistida por Computadora.

¹² Berkeley Software Distribution.

fácil uso y por ser muy ligero.

Apache tomcat 6.0

Apache tomcat fue implementado en Java, por lo que trabaja en cualquier sistema operativo que tenga instalado Java Development Kit (JDK). 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. Puede funcionar como servidor HTTP facilitando de esta forma la gestión de solicitudes y respuestas, o conectado a otro servidor HTTP como Apache HTTP Server (33). Esta herramienta fue escogida debido a que la plataforma de Pentaho BI utilizada la instala por defecto.

1.15 Herramientas para la integración de datos

Las herramientas de integración de datos brindan el soporte necesario para garantizar las actividades de extracción desde diversas fuentes, la transformación mediante un conjunto de procesos y por último la carga de los datos hacia el almacén (34).

DataCleaner 1.5.4

DataCleaner es una herramienta de código abierto para el perfilado, análisis, transformación y limpieza de datos. Estas actividades ayudan a administrar y controlar la calidad de los datos, pues generan reportes y gráficos sofisticados que permiten a los usuarios identificar y analizar la estructura del origen de datos, estableciendo vistas fáciles de interpretar (35).

Pentaho data integration 5.0.1

Pentaho data integration (PDI, también conocida como *Kettle*) es una de las soluciones más extendidas y mejor valoradas en el mercado, que reúne un conjunto de componentes que permiten modelar y ejecutar transformaciones sobre flujos de datos. Es una herramienta libre, con una interfaz de usuario sencilla que permite realizar el proceso de ETL. Permite la extracción, transformación y carga de datos para soportar las áreas de negocios. Proporciona la solución ideal para cualquier tipo de integración de datos, análisis de negocio o proyectos con grandes capacidades de datos (36).

1.16 Herramientas para la Inteligencia de negocio

El desarrollo de las herramientas de BI dentro de la organización permite el análisis de la información basada en indicadores a nivel empresarial. Los sistemas BI son un tipo de software en crecimiento en el actual mercado mundial, pues brindan una amplia gama funcionalidades (35). La herramienta seleccionada para la solución es Pentaho BI Suite.

Pentaho BI suite 5.2.0

Pentaho es un software de código abierto con licencias abiertas, es fácil de usar, fácil de personalizar y que requiere de actualizaciones sencillas. Posibilita que los usuarios de negocio puedan acceder, integrar, combinar, visualizar y analizar todos los datos que afectan al negocio. Permite la ejecución de varios trabajos y transformaciones como parte de la misma base de datos. Dentro de sus principales funcionalidades se encuentran la visualización de informes, tablas dinámicas OLAP y la minería de datos. Los módulos de la plataforma Pentaho BI Suite que se utilizarán en la solución son (36):

- **Pentaho BI server:** provee el soporte y la infraestructura necesarios para crear soluciones de inteligencia empresarial a problemas de negocios. El marco proporciona los servicios básicos, incluidos autenticación, registro, auditoría, servicios web y motor de reglas. La plataforma también incluye un motor de solución que integra reportes, análisis, tableros de comandos y componentes de minería de datos. Funciona como un sistema basado en administración web de informes y está diseñado para integrarse fácilmente en cualquier proceso de negocio (36).
- **Pentaho analysis services:** conocido como Mondrian, 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 Base de Datos es su característica más significativa. Mondrian OLAP Server permite la realización de consultas al almacén de datos y brinda la posibilidad de mostrar los resultados a través de un navegador de forma tal que el usuario pueda realizar actividades de navegación sobre cubos OLAP (36).
- **Pentaho schema workbench:** Es una herramienta para el desarrollo y prueba de cubos OLAP de forma visual. 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 (36).
- **Plugins CDE:** Es una herramienta muy poderosa y completa, que permite combinar fuentes de datos con componentes personalizados. Está diseñada para crear, editar y visualizar Dashboard. Dentro de sus principales funcionalidades están la creación de plantillas desde cero, facilitando adaptarla a las necesidades de los clientes y la previsualización de los tableros, posibilitando comprobar el aspecto y comportamiento de los Dashboard cuando aún se está trabajando en ellos (37).

1.17 Conclusiones del capítulo

Luego de finalizado el capítulo, se concluye que:

- Se demostró la necesidad de implementar un Dashboard para SIPAC, que permita el análisis de la información a través de indicadores claves de desempeño.
- Se realizó un análisis de los principales conceptos relacionados con los Dashboard y la Inteligencia de negocios, permitiendo una mejor comprensión del tema.
- Se seleccionó la Metodología para el desarrollo de almacenes de datos que permitirá estructurar, planificar y controlar el proceso de construcción de la solución.
- El estudio realizado de las tecnologías a utilizar, permitió determinar la base tecnológica necesaria para desarrollar el sistema propuesta.

Capítulo 2. Análisis y Diseño del Dashboard para SIPAC.

En el presente capítulo se realiza una descripción detallada del negocio para un mejor entendimiento del mismo. Se definen las reglas del negocio, los requisitos de información, los funcionales, no funcionales y los casos de uso del sistema. Se conforma el diagrama de casos de uso del sistema y el modelo de datos donde se determinan las dimensiones, los hechos y las medidas. Se diseña la arquitectura de sistema que regirá el desarrollo de la solución propuesta. Se realiza el proceso ETL permitiendo conocer el estado de los datos, para así poder formatearlos, limpiarlos y cargarlos.

2.1 Definición del negocio

El Sistema de Planificación de Actividades (SIPAC) 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; garantizando el seguimiento del desarrollo y cumplimiento de los objetivos y tareas principales en las entidades.

El sistema contará con un Dashboard funcional que permitirá compartir, agrupar y centralizar los datos clave de la planificación en las organizaciones, garantizando que la información solicitada sea mostrada mediante gráficos resúmenes, a través de indicadores claves de desempeño. Esto facilitará a los usuarios saber instantáneamente si están dentro o fuera de su plan y, a su vez, que exista una reacción rápida dentro de la organización.

2.2 Tema de análisis

Un tema de análisis es el fraccionamiento de la información que posee una organización según los objetivos que se persigan. El análisis comprende una serie de tareas como: estudio del negocio, investigación, identificación de las necesidades de información y especificación de los requisitos. Por lo que se definen para la presente investigación como temas de análisis los procesos de planificación (38):

- Elaboración del Plan.
- Aprobación Conciliación del Plan.
- Puntualización del Plan.
- Ejecución Control del Plan.

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, controlar e influenciar el comportamiento del mismo (38).

RN1: Todas las medidas identificadas deben ser valores positivos.

RN2: Las cifras deben estar representadas por valores numéricos.

RN3: Los valores que indiquen por ciento tienen que estar en un rango de cero a 100.

RN4: Si la cantidad de actividades internas más la cantidad de actividades de aseguramiento, es mayor que la cantidad de actividades externas, entonces el plan se elaboró correctamente.

RN5: Si la cantidad de actividades aprobadas del plan excede el 90%, entonces el plan está elaborado correctamente.

RN6: Si la cantidad de actividades rechazadas del plan excede el 10%, entonces el plan está elaborado incorrectamente.

RN7: Si la cantidad de actividades enviadas a aprobación excede el 50%, entonces el proceso de Aprobación-Conciliación no se realiza correctamente.

RN8: Si las puntualizaciones del plan exceden el 25 % del total de actividades, no se cumple el principio de Respeto al Plan.

RN9: Si las actividades extraplan exceden el 25 % del total de actividades, no se cumple el principio de Respeto al Plan.

RN10: Si la cantidad de actividades cumplidas es mayor que el 85% de las actividades planificadas, entonces el proceso de Ejecución y control del plan se realiza correctamente.

2.4 Necesidades de los usuarios

Para el análisis del proceso de negocio es preciso conocer las necesidades de los usuarios. La implicación de los mismos durante el ciclo de vida del producto es de vital importancia, de ahí se deriva la posibilidad de que los resultados sean satisfactorios o insatisfactorios. Para el caso de los Dashboard a los requisitos funcionales (RF) y no funcionales (RNF), se le agregan los requisitos de información (RI).

Requisitos de información

Los RI son el conjunto de funcionalidades que el sistema debe ser capaz de implementar para analizar los datos en cuanto a métricas previamente definidas. Dichos requisitos constituyen la principal entrada para las reglas del negocio y el proceso de Inteligencia de negocio (39). En la presente investigación fueron identificados 8 RI pertenecientes a los cuatro temas de análisis mencionados anteriormente.

- **RI1:** Obtener la cantidad y el porcentaje de actividades: internas, externas y de aseguramiento por capítulo del plan.
- **RI2:** Obtener la cantidad y el porcentaje de actividades: aprobadas, rechazadas o enviadas a aprobación.
- **RI3:** Obtener la cantidad y el porcentaje de actividades: pospuestas, suspendidas, incorporadas, afectadas por Factores que Influyen en el Plan (FIP), extraplanes y principales.
- **RI4:** Obtener la cantidad y el porcentaje de actividades: cumplidas, incumplidas y parcialmente cumplidas.
- **RI5:** Obtener la cantidad y el porcentaje de actividades por categorías en dependencia del estado de aprobación.
- **RI6:** Obtener la cantidad y el porcentaje de actividades por tipo en dependencia del estado de aprobación.
- **RI7:** Obtener la cantidad y el porcentaje de actividades: internas, externas con respecto al tiempo
- **RI8:** Obtener la cantidad y el porcentaje de actividades: extraplan y aseguramiento con respecto al tiempo.

Requisitos funcionales.

Para poder identificar qué debe hacer el sistema y entender su funcionamiento, es fundamental conocer los requisitos funcionales que debe cumplir (26). Los mismos deben estar orientados a las necesidades de los usuarios finales y garantizados por la plataforma de visualización BI server 5.2.0, la cual brinda algunas de las funcionalidades identificadas como RF.

- **RF_1:** Validar usuario y contraseña.
- **RF_2:** Adicionar usuario.
- **RF_3:** Eliminar usuario.
- **RF_4:** Visualizar usuario.
- **RF_5:** Modificar usuario.
- **RF_6:** Adicionar rol.
- **RF_7:** Eliminar rol.
- **RF_8:** Visualizar rol.

- **RF_9:** Modificar rol.
- **RF_10:** Adicionar reporte.
- **RF_11:** Eliminar reporte.
- **RF_12:** Visualizar reporte.
- **RF_13:** Modificar reporte.
- **RF_14:** Personalizar reporte.
- **RF_15:** Adicionar gráfico.
- **RF_16:** Eliminar gráfico.
- **RF_17:** Visualizar gráfico.
- **RF_18:** Modificar gráfico.
- **RF_19:** Exportar a otro formato.
- **RF_20:** Extraer datos de la fuente
- **RF_21:** Realizar transformación de los datos
- **RF_22:** Cargar los datos.

Requisitos no funcionales.

Los RNF detallan las propiedades o cualidades que el producto debe tener, aumentándole funcionalidad al sistema, pues hacen al producto atractivo, fácil de usar, rápido y confiable (26). 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:

Requisitos de usabilidad

- **RNF_1:** El sistema podrá ser usado por personas con conocimientos intermedios o que tengan experiencia en la utilización de computadoras.
- **RNF_2:** Mostrar los mensajes, títulos y demás textos que aparezcan en la interfaz del sistema en idioma español.
- **RNF_3:** Agilizar el acceso a los reportes del almacén de datos mediante la distribución de la información por áreas de análisis.

Requisitos de fiabilidad

- **RNF_4:** El acceso a la información debe estar disponible y se tendrán en cuenta los permisos establecidos.

Requisitos de seguridad:

- **RNF_5:** La información tiene que ser accedida solo por las personas autorizadas, por lo que serán establecidos niveles de acceso.

Requisitos de interfaz

- **RNF_6:** Las interfaces serán cargadas solo con la información necesaria.
- **RNF_7:** Los reportes mostrarán una interfaz sencilla que permita la interacción usuario-aplicación.
- **RNF_8:** Los gráficos usarán los colores definidos por el cliente siempre que permitan establecer un buen diseño.

Requisitos de restricciones de diseño

- **RNF_9:** Para la realización del Modelo de datos se utilizará como SGBD el PostgreSQL, la herramienta para la administración de este es el pgAdmin, el lenguaje utilizado por este es el PL/pgSQL.
- **RNF_10:** Para el proceso de integración de datos se usará la herramienta Pentaho data integration 5.0.1.
- **RNF_11:** Para la implementación de la capa de visualización se utilizarán las herramientas de Inteligencia de negocio definidas durante la investigación y el lenguaje utilizado serán las consultas MDX.
- **RNF_12:** Las estructuras del Almacén de datos se nombrarán de una manera estándar teniendo en cuenta el tipo de estructura que se maneje.

Requisitos de confiabilidad

- **RNF_13:** Para garantizar la persistencia de la información se realizará un respaldo total de los datos del almacén de datos.

Requisitos de Hardware

- **RNF_14:** El sistema deberá contar con las siguientes características de hardware: 8 GB RAM, Core i3 y 500 Gb de capacidad en disco duro.

Requisitos de Software

- **RNF_15:** El sistema deberá contar con las siguientes características de Software: máquina virtual de java jdk1.7 y sistema operativo Linux.

2.5 Casos de uso del sistema

Los casos de uso son un conjunto de escenarios que identifican una línea de utilización para el sistema que va a ser desarrollado, facilitando una descripción de cómo se usará el sistema (26). Para su elaboración se

tomaron como base los RI y RF agrupados e identificados anteriormente en 12 CU, los cuales se agruparon en 8 casos de uso funcionales y 4 casos de uso de información definidos por las áreas de análisis identificadas anteriormente.

Casos de uso funcionales

- CUF_1: Autenticar usuario.
- CUF_2: Gestionar usuario.
- CUF_3: Gestionar rol.
- CUF_4: Gestionar reporte.
- CUF_5: Gestionar gráficos.
- CUF_6: Realizar operaciones sobre los reportes.
- CUF_7: Realizar transformación y carga de los datos.
- CUF_8: Realizar extracción de los datos.

Caso de uso de información.

- CUI_9: Presentar información de Elaboración del Plan.
- CUI_10: Presentar información de Aprobación- Conciliación del Plan.
- CUI_11: Presentar información de Puntualización del Plan.
- CUI_12: Presentar información de Ejecución-Control del Plan.

Descripción y permiso de los actores del sistema

Los actores del sistema intercambian información con él, aunque no forman parte de este. Pueden representar el rol que juega una o varias personas (40). En la Tabla 2, se muestran los actores y los permisos que poseen en el sistema.

Tabla 2. Descripción y permiso de los actores del sistema.

Actores	Descripción	Permisos	
		Lectura	Escritura
Administrador	Administra el almacén y se encarga de todas las actividades referentes al proceso de ETL. Hereda las características del analista y presenta además permiso de escritura para administrar los roles, usuarios, permisos y reportes.	x	x
Analista	Es el encargado de analizar la información y visualizar los reportes. Tiene acceso de solo lectura.	x	

2.6 Diagrama de caso de uso del sistema

Un diagrama de caso de uso del sistema (DCUS) muestra la relación entre los actores del sistema y los casos de uso. Para la representación del diagrama (Figura 2) se utilizaron tres patrones: el CRUD, el de concordancia por adicción y el de generalización/especialización. El primer patrón se evidencia en los CU Gestionar usuario, Gestionar rol, Gestionar reporte y Gestionar gráfico, pues agrupa las funcionalidades de adicionar, modificar, visualizar y eliminar. El segundo se ve reflejado en la relación de extensión existente entre los CU de información y el CU Realizar operaciones sobre reportes. El tercer patrón se muestra en la herencia del comportamiento del administrador sobre el analista.

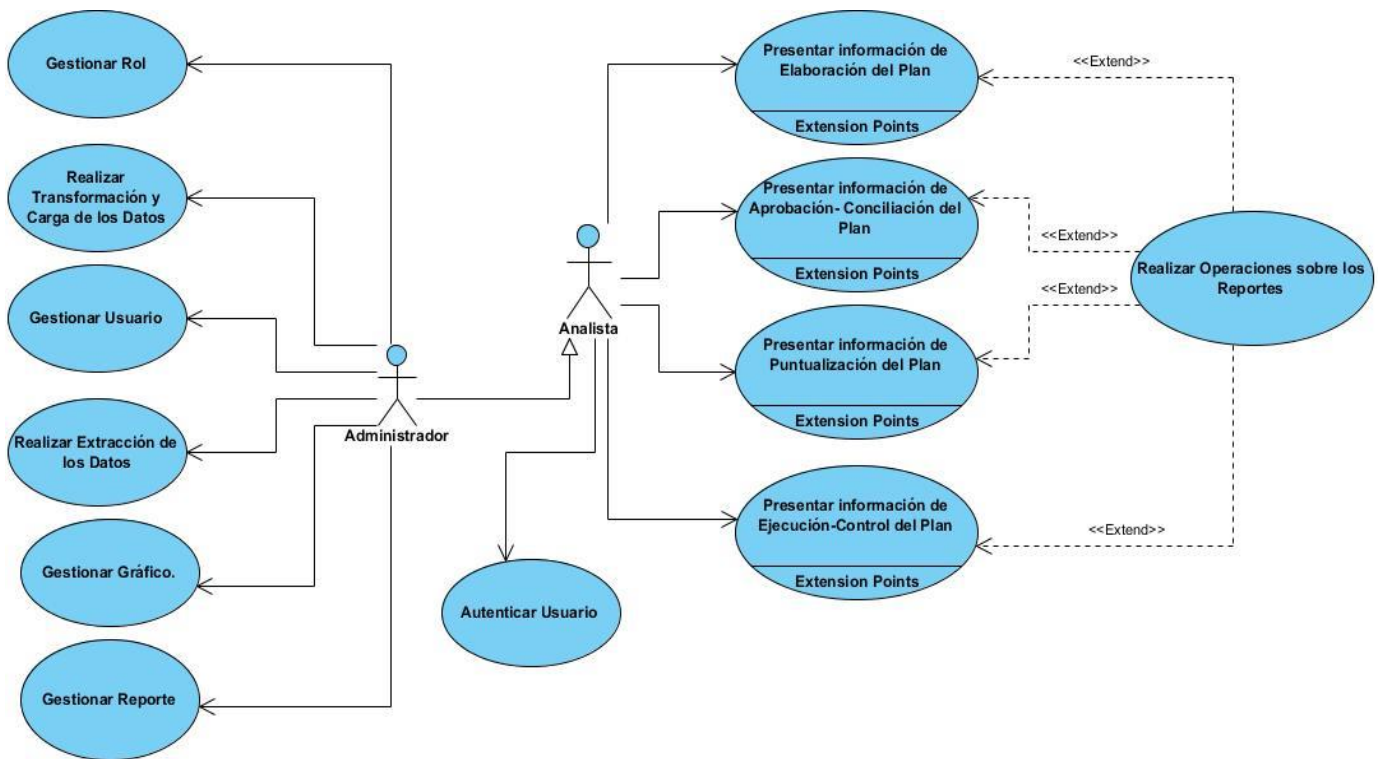


Figura 2. Diagrama de caso de uso del sistema

2.7 Descripción de los casos de usos del sistema

En el siguiente apartado se muestra como ejemplo la descripción del caso de uso de información “Presentar información de Actividades”. Se pueden consultar el resto de las especificaciones de casos de uso en el Anexo 1: Especificación de los casos de uso de información.

Tabla 3. Descripción del caso de uso de información: Presentar información de Elaboración del Plan.

Objetivo	Presentar información de la Elaboración del Plan.	
Actor	Analista	
Resumen	El caso de uso comienza cuando el analista solicita realizar un análisis de la información referente a la elaboración del plan. El sistema genera los reportes referentes con la información seleccionada y las opciones de los posibles cambios que se le puede realizar al mismo. El CU finaliza cuando el actor termina el análisis de la información.	
Complejidad	Media.	
Prioridad	Media.	
Precondiciones	Debe estar seleccionada el A.A.G_Dashboard_SIPAC.	
Postcondiciones	Se consultaron los reportes pertenecientes al CU Presentar información de Elaboración del Plan.	
Flujo de eventos		
Flujo básico: Presentar Información de Elaboración del Plan.		
	Actor	Sistema
1	Selecciona el área de análisis A.A_Elaboración_Plan.	
2		Muestra los libros de trabajo contenidos en el A.A_Elaboración_Plan.
3	Selecciona el libro de trabajo LT_total_actividades o LT_porcentaje_actividades.	
4		Muestra los reportes contenidos en el LT_total_actividades o LT_porcentaje_actividades.
5	Selecciona el reporte que desea analizar.	
6		Muestra la información contenida en el reporte seleccionado y ofrece opciones al actor para realizar cambios al reporte durante su análisis. En caso de que se desee realizar algún cambio se debe ir al CU Realizar operaciones sobre los reportes. Se

	finaliza el caso de uso.	
Opciones de reportes de Presentar información de Elaboración del Plan.		
Perspectivas de análisis	Posibles resultados	
	Medidas	Periodicidad
<ul style="list-style-type: none"> • Proceso • Relacionfip 	<ul style="list-style-type: none"> • Cantidad de actividades • Porciento de Actividades 	Diaria
Relaciones	CU Extendidos	CU Realizar operaciones sobre reportes

2.8 Arquitectura del sistema

La arquitectura de un sistema (Figura 3), es la organización de los componentes y las relaciones entre ellos. Para dicha arquitectura se deben tener en cuenta los requisitos del sistema, las restricciones a las que está sujeto y, a su vez, representa una guía para apoyar la integración de las distintas tecnologías (41). Para tener una perspectiva general del sistema se explica la arquitectura del mismo, el cual quedó conformado por tres subsistemas bases: integración de datos, almacenamiento y visualización (42).

- **Subsistema de integración:** realiza el proceso de extracción, transformación, integración, limpieza y estandarización de las distintas fuente de datos.
- **Subsistema de almacenamiento:** contiene las tablas de hechos definidas en el negocio, las dimensiones seleccionadas para la solución y las relaciones que existen entre ellas. Toma como elementos de partida los datos manipulados por el subsistema de integración.
- **Subsistema de visualización:** organiza los reportes por áreas de análisis de los datos almacenados en el subsistema de almacenamiento, además permitirá visualizar gráficos resúmenes con el objetivo de mostrarlos a los usuarios finales, contribuyendo al análisis de la información basada en los indicadores de planificación.

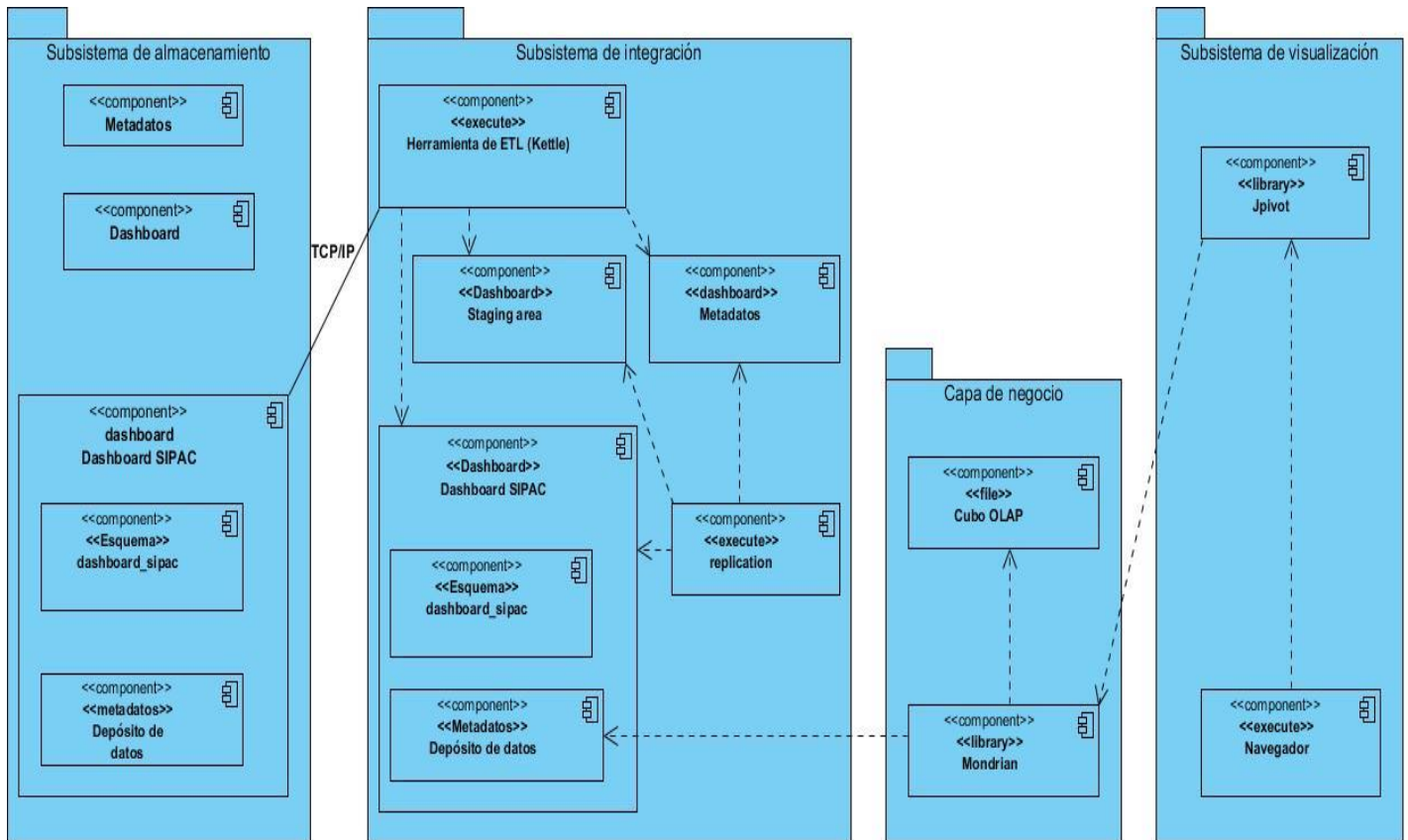


Figura 3. Arquitectura del Sistema.

2.9 Diseño del subsistema de integración de datos

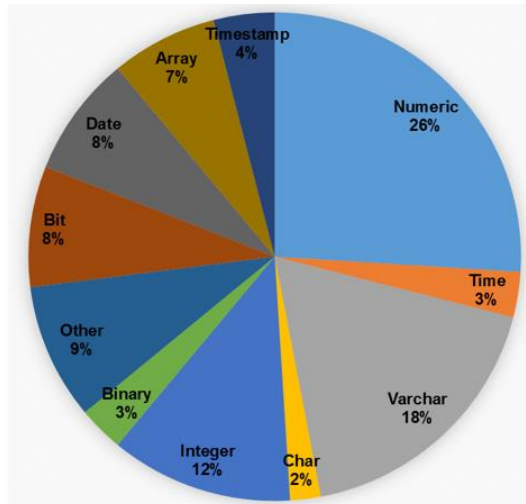
El subsistema de integración de datos comprende el perfilado de los datos y la extracción de los datos desde el sistema fuente. Para el diseño de dicho subsistema se realizaron dos importantes actividades: el perfilado de datos y el diseño del proceso de integración, las cuales serán descritas a continuación:

Perfilado de datos

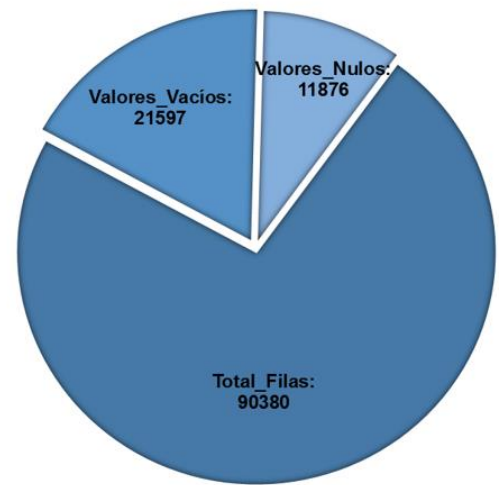
El perfilado de los datos es el proceso que se encarga de analizar las fuentes de datos con el objetivo de conocer el estado en que se encuentran, para así poder administrar y supervisar la calidad de los mismos. Además, verifica la existencia de valores nulos, distintos o duplicados, permitiendo definir nuevas reglas del negocio que posteriormente pasan a ser las reglas de transformación aplicadas durante la implementación del subsistema de integración.

Al realizar el perfilado a la fuente de datos de SIPAC se identificó que los tipos de datos que más predominaban eran *numeric* y *varchar*. Además hay presencia de otros tipos de datos como: *integer*, *other*,

bit, date, array, timestamp, bynary, time y char. No se hallaron valores negativos, pero sí hay presencia de valores nulos y vacíos, posibilitando definir 3 nuevas RN que serán aplicadas durante el proceso de transformación (Figura 4).



Tipos de datos



Calidad de los datos

Figura 4. Calidad y tipos de datos.

Proceso de integración

Una vez que se conoce la estructura, contenido y fiabilidad de los datos, se procede a realizar el diseño del proceso ETL (Figura 5). Siendo este el encargado de hacer transformaciones en los datos para lograr una integración exitosa y a su vez permite recopilar datos desde múltiples fuentes, para así poder darle el formato que se requiera, limpiarlos y cargarlos.

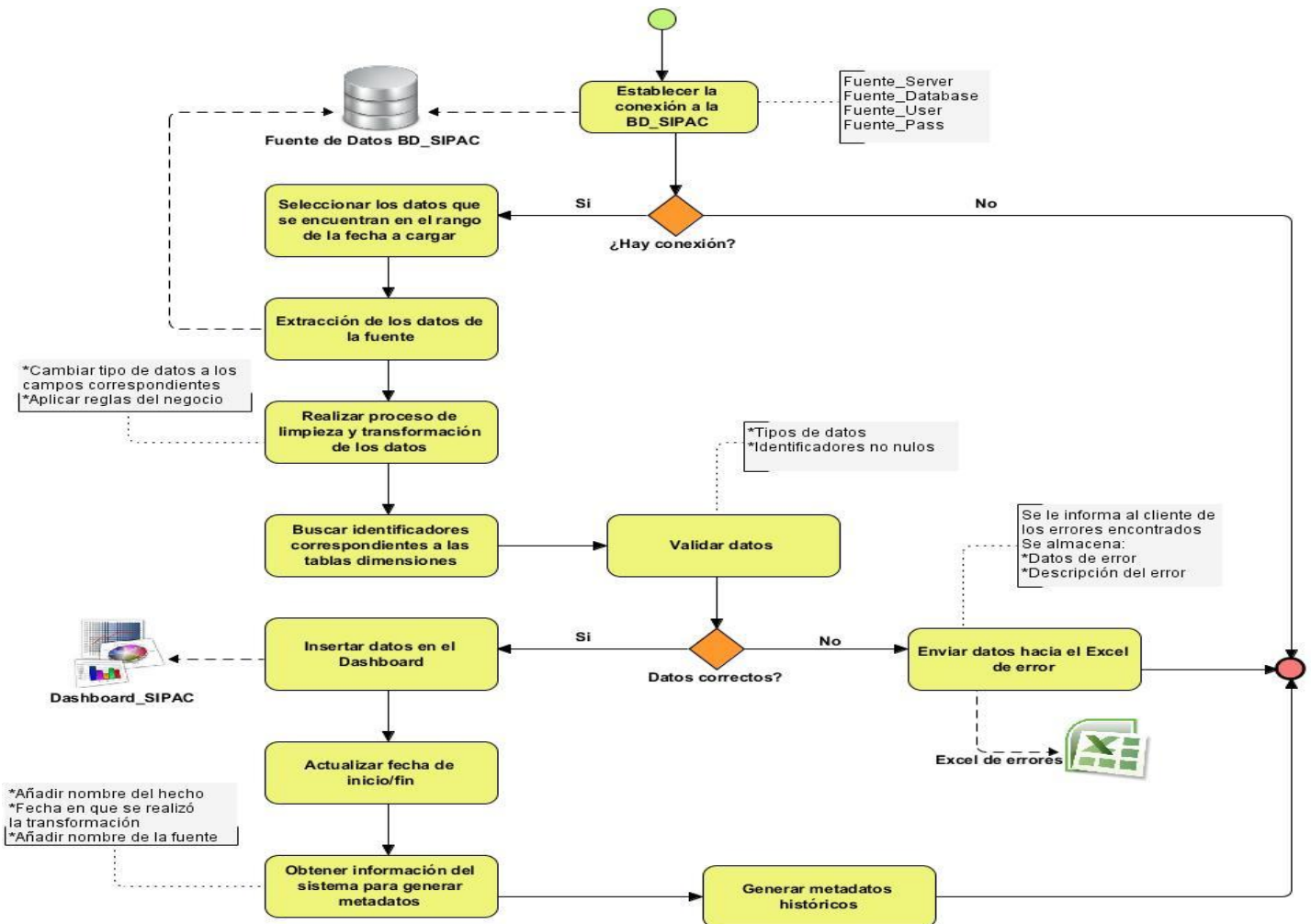


Figura 5. Diseño del subsistema de integración.

2.10 Diseño del subsistema de almacenamiento

El diseño del subsistema de almacenamiento se realiza a partir de los resultados obtenidos en la fase de requisitos y de los datos de interés del negocio, con el fin de transformarlos al modelo lógico representado por medidas, hechos, y dimensiones. Un hecho es la ocurrencia de un proceso específico, mientras que una dimensión establece una perspectiva a través de la cual se realiza el análisis sobre el hecho y las medidas establecen un valor sobre el hecho (7). Partiendo de lo descrito anteriormente se identificaron 8 dimensiones, 1 medida y 1 hecho.

Tabla 4. Descripción del hecho.

Hecho	Descripción
hech_actividades	Es objeto de análisis del negocio. Comprende todo lo relacionado con las actividades.

Después de descrito el hecho, se presenta en la siguiente tabla la descripción de las dimensiones identificadas.

Tabla 5. Descripción de las dimensiones.

Dimensión	Jerarquía	Descripción
dim_estado_aprobacion	estado_aprobacion	Determina el estado de aprobación de las actividades.
dim_estado_cumplimiento	estado_cumplimiento	Determina el grado de cumplimiento de los elementos.
dim_actividad	actividades	Indica la denominación de cada actividad y presenta su respectivo código.
dim_proceso	proceso	Determina el proceso en el cual se generan las actividades.
dim_capitulo	capitulo	Registra el nombre del capítulo. Presenta su respectivo código.
dim_tiempo_dia	tiempo_dia	Registra el año, el mes y el día. Presenta su respectivo código.
dim_categoria_actividad	categoria_actividad	Determina la categoría de las actividades.
dim_tipo_actividad	tipo_actividad	Determina el tipo de actividad.

Después de descrito el hecho y las dimensiones, se presenta en la siguiente tabla la descripción de la medida identificada.

Tabla 6. Descripción de la medida.

Medida	Descripción
cantidad_actividades	Muestra el total de actividades.

Modelo de datos

Para la realización del modelo de datos (Figura 6) se seleccionó el esquema estrella, pues existe una tabla de hecho central y varias tablas de dimensiones relacionadas a esta, permitiendo la mejora en los tiempos de respuesta ante las consultas a la base de datos.

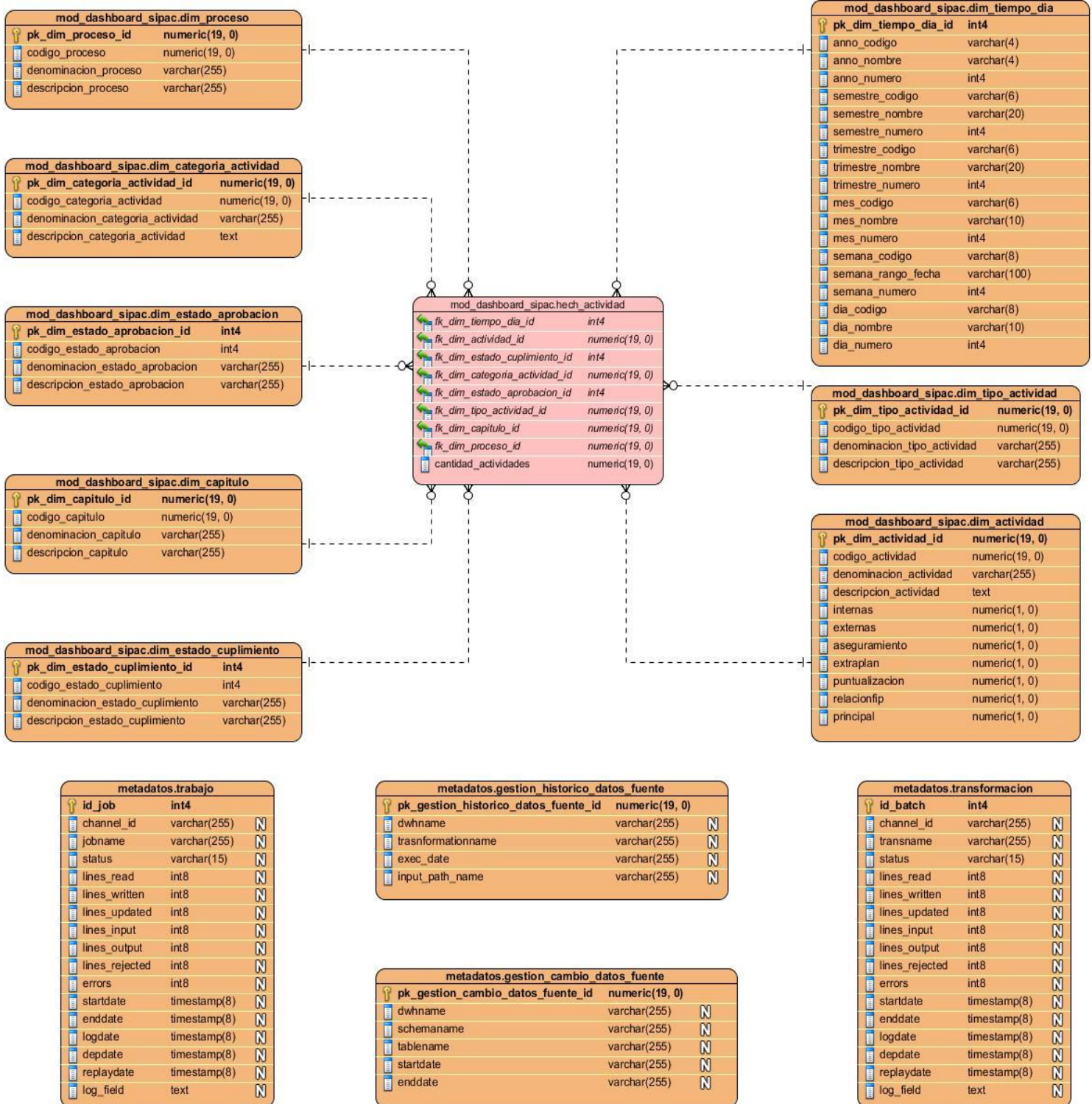


Figura 6. Modelo de datos dimensional.

2.11 Diseño del subsistema de visualización

El diseño del subsistema de visualización se realizó por áreas de análisis, facilitando a los usuarios finales que la búsqueda de información sea confiable y fácil.

Arquitectura de información

Para la arquitectura de información (Figura 7) se definió la estructura de navegación conformada por un área de análisis general (A.A.G), 4 áreas de análisis (A.A) y 2 libros de trabajo (LT) por cada área de análisis, para un total de 8 libros de trabajo. Cada LT contiene las vistas reportes. Para una mejor comprensión del tema, se muestra una representación del diseño establecido para la arquitectura de la información del sistema.

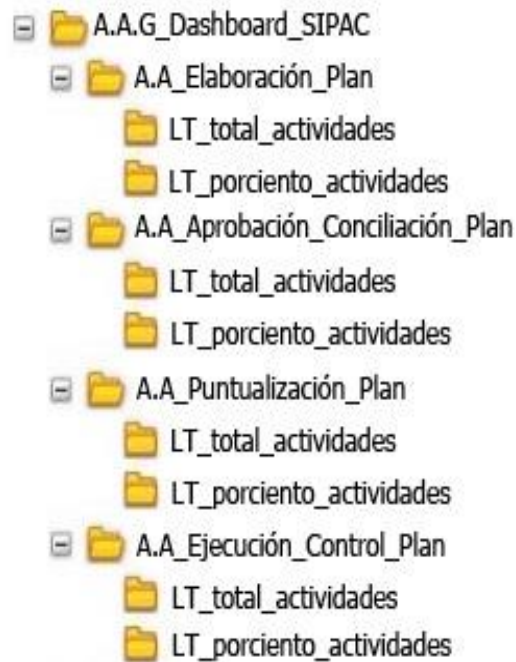


Figura 7. Arquitectura de Información.

2.12 Seguridad del sistema

Los sistemas informáticos deben estar provistos de mecanismos que aseguren la integridad de los datos y del sistema en general. Usualmente se establecen roles y permisos que se le designan a los distintos usuarios que posea la aplicación (43). La Plataforma Pentaho BI contiene su propia seguridad compuesta por cuatro áreas esenciales (44):

- **Seguridad de acceso a datos de objetos:** incluye usuarios, contraseñas, autorizaciones permitidas, recursos web y protección a datos.
- **Autenticación:** incluye el procesamiento de información interactiva de inicio de sesión, comparándola con la información del almacén de datos.
- **Autorización de recursos web (URL):** ofrece protección a las URL para responder a cada usuario si pueden o no acceder a una determinada página.
- **Autorización a objetos del dominio:** en el sistema los únicos objetos del dominio protegidos por la plataforma, son los objetos del repositorio otorgados al usuario autenticado. Es responsabilidad de los objetos del dominio autorizar las operaciones solicitadas por este.

2.13 Conclusiones del capítulo

Luego de realizarse el análisis y diseño del Dashboard para el Sistema de Planificación de Actividades se concluye que:

- Fueron identificadas 11 reglas del negocio, permitiendo la definición de las reglas de transformación que fueron utilizadas en el proceso de integración de datos.
- Se definieron 4 RI, 22 RF y 7 RNF, los cuales fueron agrupados en 4 CUI y 8 CUF permitiendo la realización del DCUS el cual contendrá las relaciones entre los actores y las funcionalidades del sistema.
- Con la definición de la arquitectura del sistema se identificaron los elementos y subsistemas que están implicados en el desarrollo de la solución.
- En el modelo de datos se identificaron 8 dimensiones, 1 medidas y 1 un hecho; posibilitando garantizar el correcto funcionamiento del sistema.
- El perfilado de datos realizado a la fuente arrojó el estado en que se encuentran los datos, permitiendo definir 3 nuevas reglas del negocio aplicables durante el proceso de transformación.

Capítulo 3: Implementación y Prueba del Dashboard para SIPAC.

El presente capítulo comprende el desarrollo de los tres subsistemas diseñados en el capítulo anterior. En subsistema de almacenamiento se definen los estándares de codificación y la construcción del modelo físico. Por su parte, en el subsistema de integración de datos se implementan las transformaciones y los trabajos, además se controla la gestión de cambio en las dimensiones, los metadatos y la captura de los cambios en los datos. En el subsistema de visualización se implementa los cubos OLAP y los reportes candidatos. Además se realiza la etapa de prueba que permite validar la calidad del producto.

3.1 Implementación del subsistema de almacenamiento

El objetivo de la implementación del subsistema de almacenamiento es otorgar al sistema una correcta estructura y organización de los datos. En él se realiza el desarrollo de la estructura física y se definen los estándares de codificación del almacén de datos.

Estándares de codificación

Los estándares de codificación tienen como objetivo organizar la forma en que se denominan las estructuras de datos, especificando un patrón que contribuya a la normalización de los términos utilizados. Esta codificación está dirigida a los desarrolladores, definiendo un vocabulario común en todo el almacén de datos que permita un entendimiento claro y conciso de cada una de las estructuras. En la siguiente tabla se muestra cómo quedan definidos estos estándares en el almacén Dashboard para SIPAC.

Tabla 7. Estándares de codificación.

Tipo de objeto	Función	Nomenclatura	Descripción
Esquema	Dashboard SIPAC	mod_dashboard_sipac	Esquema donde se almacenan las tablas de hechos y dimensiones.
	Metadatos	metadatos	Esquema donde se encuentran las tablas que almacenan los metadatos.
Tabla	Tabla de hecho	hech_ [nombre hecho]	La tabla hecho tendrán una cadena (hech_) y seguidamente el nombre que describe lo que realiza el hecho.
	Tablas de dimensiones	dim_ [nombre dimensión]	Las tablas dimensiones tendrán una cadena (dim_) y seguidamente el nombre que describe lo que realizan.









	Tablas de metadatos	[nombre metadatos]	Las tablas metadatos tendrán el nombre de lo realizan.
Constrains	Llaves primarias	pk_dim_[nombre tabla]_id	Todas las llaves primarias tendrán una cadena (pk_dim_) y el nombre de la tabla a la que pertenecen. Además poseerán el identificador (_id), que serán de tipo integer y autoincremental.
	Llaves foráneas	fk_dim_[nombre tabla]_id	Todas las llaves foráneas tendrán una cadena (fk_dim_) y el nombre de la tabla a la que pertenecen. Además poseerán el identificador (_id), que serán de tipo integer y autoincremental.
Índice	Índices	ind_[nombre atributo]	Los índices se nombrarán con una cadena que los identifique (ind_) y a continuación el nombre del atributo de la tabla al que hace referencia.








Implementación del modelo de datos físico

El modelo de datos físico constituye una colección compuesta de entidades que detallan las estructuras de los datos, las restricciones de integridad y las operaciones de manipulación de los mismos. Dicho modelo se genera a partir del modelo lógico dimensional, permitiendo describir las estructuras, las restricciones de integridad y las operaciones de manipulación de los datos.

Para garantizar un rápido acceso a las consultas realizadas a los datos se implementaron índices, mediante la estrategia btree (árboles B), debido a que proporcionan acceso rápido para consultas de gran complejidad (45). Además, se definieron dos esquemas “metadatos” y “Dashboard_SIPAC” para proveer una mejor organización del sistema. Por lo cual la solución de dicho modelo se compone por dos esquemas mencionados anteriormente, 9 secuencias y 13 tablas.

Tabla 8. Modelo de datos físico.

Esquemas	Tablas
 Dashboard_SIPAC	 hech_actividad
	 dim_capitulo
	 dim_proceso
	 dim_actividad
	 dim_estado_aprobacion
	 dim_estado_cumplimiento
	 dim_tiempo_dia

	 dim_categoria_actividad
	 dim_tipo_actividad
 Metadatos	 gestion_cambio_datos_fuente
	 gestion_historico_datos_fuente
	 trabajo
	 transformación

En la figura 9 se muestra cómo queda estructurada la BD en la herramienta pgAdmin:

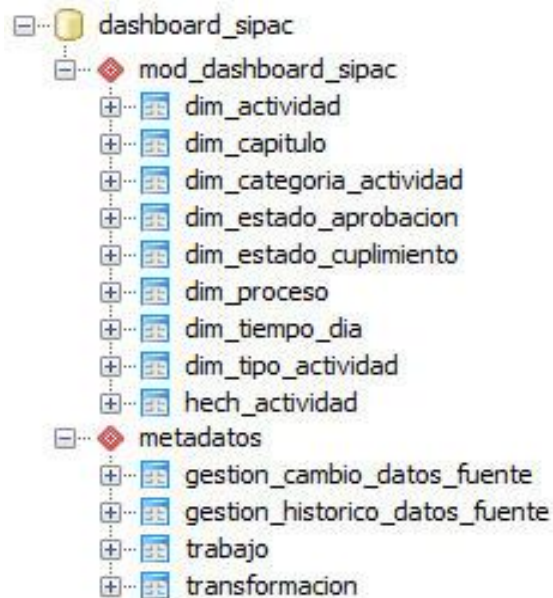


Figura 8. Estructura física de la BD.

3.2 Implementación del subsistema de integración de datos

La implementación del subsistema de integración de datos consta de tres etapas fundamentales: la extracción, transformación y carga de los datos. Los datos son extraídos y posteriormente se realiza el perfilado, el cual consiste en analizar la información para conocer el estado en que se encuentran los datos. Luego estos son transformados y cargados poblando las dimensiones y el hecho, que conforman la estructura del Dashboard para SIPAC.

Para realizar el proceso de integración de datos hay varios subsistemas que posibilitan una correcta implementación de cada uno de los subprocesos que lo componen. Kimball propuso 34 subsistemas ETL

(46). Para el desarrollo de la solución solamente fueron utilizados un conjunto de estos subsistemas, los cuales se describen seguidamente:

- **Subsistema de perfilado de datos:** permite explorar los datos para comprobar su calidad y el cumplimiento de los estándares acorde a los requerimientos.
- **Subsistema de extracción de datos:** permite la extracción de los datos desde la fuente origen hacia la fuente destino.
- **Subsistema de limpieza de datos:** permite realizar un conjunto de pasos correspondientes a la limpieza de los datos.
- **Subsistema de transformación:** permite realizar transformaciones como: el cambio de datos en algunos campos, el filtrado de valores, la búsqueda de información y el mapeo de valores.
- **Subsistema de carga:** permite realizar la carga de los datos a las tablas de hechos y dimensiones.
- **Tablas de hecho:** permite crear tablas de hecho.
- **Subsistema de control de versiones:** permite realiza la gestión histórica de los datos, registrando cuándo sucedió la última carga, empleándose mediante la tabla metadatos gestion_historico_datos_fuente.
- **Subsistema de repositorio de metadatos:** permite capturar los metadatos de los procesos ETL, los datos del negocio y los aspectos técnicos.
- **Subsistema de Dimensiones Lentamente Cambiantes (SCD):** permite implementar la lógica para crear atributos de variabilidad lenta a lo largo del tiempo.
- **Subsistema de rastreo de eventos de errores:** permite capturar los errores que suministran información precisa sobre la calidad de los datos y posibilita la mejora de los mismos.

Captura de cambio en los datos

Es de vital importancia conocer cómo se transmite la información dentro de las entidades que manipula SIPAC para así poder establecer como se efectuó la captura de los cambios dentro de estas. La captura de la información referente a estas entidades se realizó desde los departamentos donde se realizan las operaciones pertinentes referentes a los especialistas de cada departamento.

Teniendo en cuenta lo descrito anteriormente se definió la obtención de los cambios en los datos del Dashboard semanalmente, este cambio se gestiona mediante la tabla de metadatos gestion_cambio_datos_fuente. Dicha tabla almacena las fechas que corresponde a la carga del hecho, posibilitando que se actualicen las fechas de inicio y fin. Además permite conocer la fecha en que se realizó

la última carga al almacén de datos. El proceso de carga gestiona también el cambio histórico de los datos mediante la tabla `gestion_historico_datos` fuente.

Gestión de los metadatos del proceso de integración de datos

Los metadatos son datos que ayudan a identificar, describir y localizar recursos digitales, son información estructurada que describen o permiten encontrar, gestionar, controlar, entender y preservar otra información; o sea, que no son más que datos sobre los propios datos (47). Estos pueden ser agrupados en cinco categorías:

- **Metadatos de negocio:** permiten obtener los datos y la información relativa a los aspectos del negocio, usualmente son los datos resultantes de la fuente.
- **Metadatos de proceso:** permiten para capturar la información de los procesos en que se ejecutan, posibilitando que los administradores gestionen el sistema y aseguren que los procesos se realicen sin dificultades.
- **Metadatos administrativos:** permiten el manejo y administración de los recursos de información incluyendo el uso y control de acceso.
- **Metadatos técnicos o modelos:** permiten describir el funcionamiento de un sistema o el modo en que se interrelacionan sus componentes. Se usan por un personal más técnico como desarrolladores, analistas entre otros. Además incluyen temas como definiciones de tablas y tipos de datos.
- **Metadatos descriptivos y de descubrimiento:** permiten describir e identificar recursos de información, posibilitando a los usuarios la búsqueda y recuperación de la información.

En la presente investigación fueron utilizados los metadatos de proceso para obtener la información de las transformaciones y los trabajos referentes a los subprocesos ETL. Esta estrategia se aplicó con el objetivo de poder presentar los resultados de la ejecución del propio proceso de integración de datos.

Gestión de cambio en las dimensiones lentamente cambiables

Las dimensiones lentamente cambiantes (SCD)¹³ son dimensiones en las que sus datos tienden a modificarse a través del tiempo, ya sea de forma ocasional o constante, que implique a un solo registro o la tabla completa. Cuando ocurre este tipo de cambios se puede optar por registrar el historial de cambios o

¹³ Del inglés: Slowly Changing Dimensions.

reemplazar los valores que sean necesarios (48). Ralph Kimball propone seis estrategias a seguir para las SCD (49):

- **SCD Tipo 0 (no tiene en cuenta la gestión histórica):** presenta un enfoque pasivo, pues no se realiza ningún esfuerzo para lidiar con los problemas del cambio de la dimensión. Es decir, los valores permanecen intactos a cuando los registros fueron creados.
- **SCD Tipo 1 (sobrescribir):** es utilizado en casos en donde la información histórica no es importante, pues sobrescribe los datos antiguos por nuevos. Este tipo presenta como desventaja que no permanece ningún registro histórico en la dimensión.
- **SCD Tipo 2 (añadir fila):** permite que al suceder un cambio en la dimensión se cree una nueva entrada en la tabla, a este nuevo evento se le asigna una nueva llave subrogada y a partir de este momento será el valor usado. Este tipo permite guardar toda la información histórica en el almacén de datos.
- **SCD Tipo 3 (añadir columna):** controla el cambio agregando una nueva columna. Por lo que una de las columnas conservaría el dato anterior, mientras que la otra tendría el dato nuevo a añadir. Este tipo presenta como desventaja que solo guarda un historial limitado de datos, dependiendo del número de columnas que se cree.
- **SCD Tipo 4 (tabla de historia separada):** permite almacenar en una tabla adicional los detalles de cambios históricos realizados a la tabla de dimensión. La tabla con la información histórica indicará el tipo de operación que se ha realizado (insertar, modificar o eliminar), sobre qué campo y la fecha de modificación. Esta tabla tiene como objetivo de contar con detalle de todos los cambios, para luego analizarlos.
- **SCD Tipo 6 (híbrido):** este método combina los tipos anteriores y se le denomina tipo 6 debido a la suma de los anteriores que lo componen ($1+2+3=6$). Esta estrategia utiliza los 3 tipos anteriormente mencionados y añade una pareja adicional de columnas para indicar las fechas de validez.

Las dimensiones pertenecientes al Dashboard para SIPAC se clasifican como SCD tipo 2. Esto se evidencia a través del componente *Insertar/Actualizar*, marcando la opción de no realizar actualizaciones (Figura 9). De esta manera se asegura que no se ejecuten actualizaciones en el Dashboard, sino que se inserte una nueva fila en la tabla.

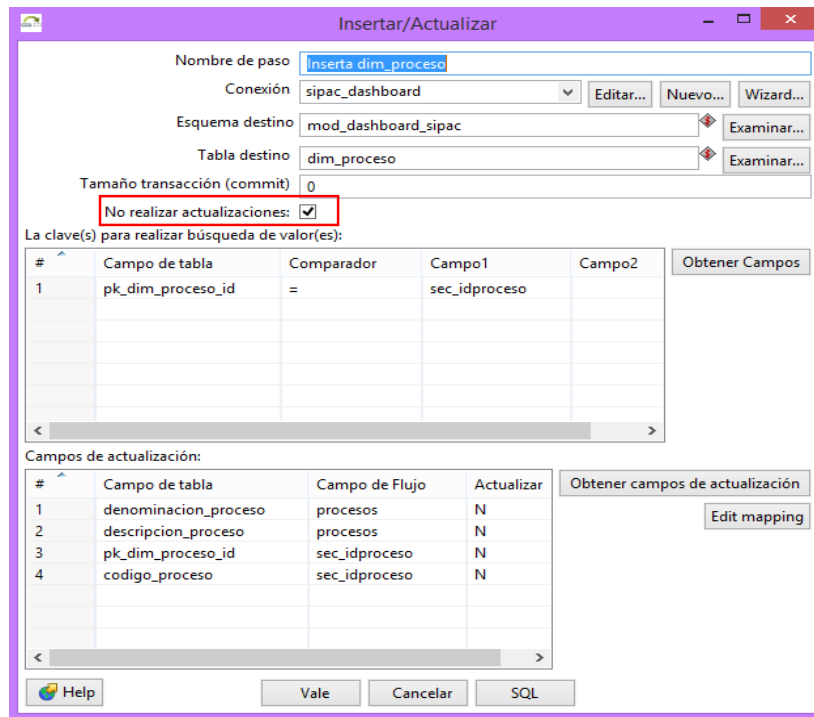


Figura 9. Componente Insertar/Actualizar.

Implementación de las transformaciones

Las transformaciones están compuestas por pasos vinculados entre sí. Los pasos son el elemento más pequeño entre las transformaciones y, a través de estos, la información fluye entre los diferentes pasos. En la presente investigación se realizó un flujo de transformación para la carga de cada una de las tablas pertenecientes al esquema Dashboard_SIPAC, por lo que se realizaron un total de 9 transformaciones, 8 para las dimensiones y 1 para el hecho. Además se ejecutó un trabajo general el cual define el orden lógico en que se realizan las transformaciones.

- **Implementación de las dimensiones**

Para la implementación de las dimensiones se ejecutaron 8 transformaciones. Las mismas se efectuaron a partir del acceso a la fuente de datos, de la cual se extraen los campos necesarios para poblar la BD y luego dar paso a la transformación de los hechos. Seguidamente se representa (Figura 10) y describe la transformación de la dim_estado_aprobacion:

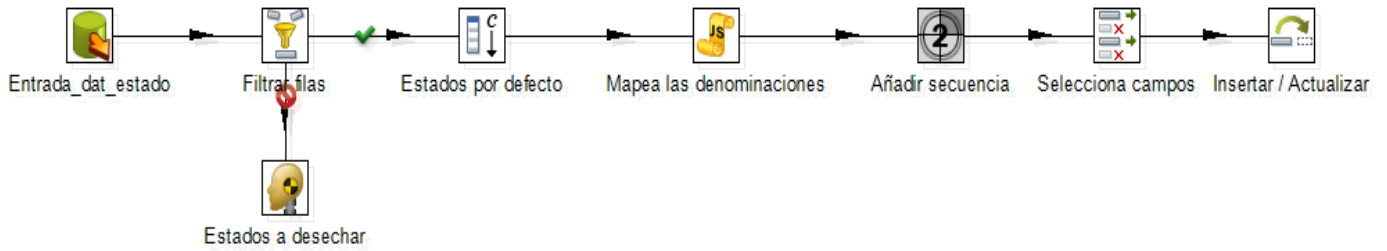


Figura 10. Transformación de la dimensión: dim_estado_aprobacion.

Primeramente, se accede a la fuente de datos y se extraen los campos. Luego se filtran las filas para ver cuáles cumplen con los parámetros requeridos con el objetivo de desechar las que no cumplan con estas características. Seguidamente se accede a la limpieza de los datos, para posteriormente realizar el proceso de estandarización en el cual se establecen parámetros por defecto dependiendo del tipo de estado que presenten los datos. Por último, se insertan los campos en la BD.

• **Implementación del hecho actividad**

Después de cargadas todas las dimensiones se realizó 1 transformación para la carga del hech_actividad. Utilizándose un conjunto de subsistemas propuestos por Kimball para el desarrollo del proceso ETL. Seguidamente se representa (Figura 11) y describe la transformación del hech_actividad:

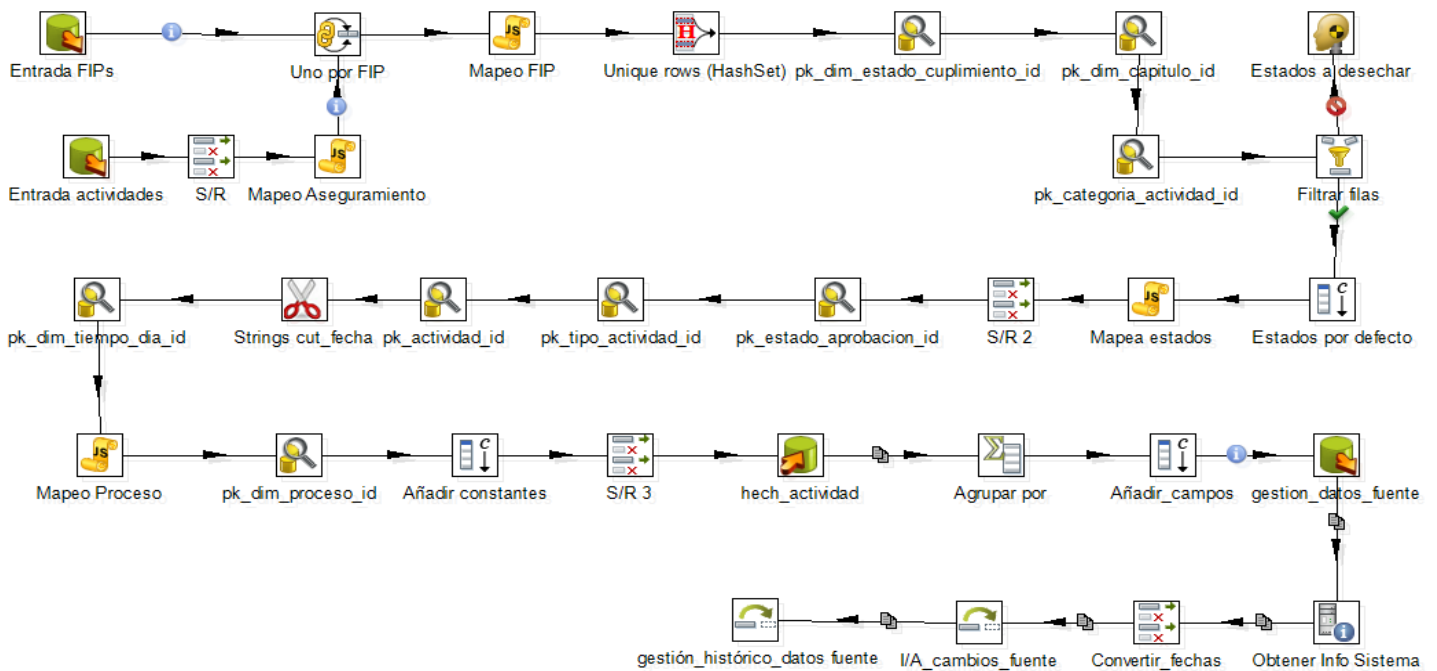


Figura 11. Transformación hecho actividad.

Para realizar la extracción de los datos correspondientes a la tabla de hecho se utilizó el subsistema de extracción, donde se obtuvieron los datos desde el componente *entrada tabla* el cual permite reconocer la información desde una BD. Luego de finalizado este proceso se realizó la limpieza y transformación de los datos logrando la estandarización de los mismos; conjuntamente fueron aplicadas las reglas de transformación identificadas con anterioridad. Mediante el subsistema rastreo de errores se capturaron los datos erróneos al realizar las transformaciones. Seguidamente se cargaron los datos ya transformados utilizándose el componente insertar/actualizar y por último a través del subsistema repositorio de metadatos se implementa el registro histórico de la fuente y las cargas.

• **Implementación de los trabajos**

Los trabajos o *job*, son un conjunto de tareas que se realizan con el objetivo de ejecutar una acción determinada. Estos definen una secuencia lógica para realizar las transformaciones o trabajos previamente diseñados, mediante el uso de pasos definidos que son diferentes a los disponibles para las transformaciones. Los trabajos se encuentran en un nivel superior a las transformaciones. En el caso del Dashboard para SIPAC se realizó un trabajo (Figura 12) para cargar el hecho actividad, en el cual se garantiza que si no se cargan las dimensiones el trabajo no se ejecute.

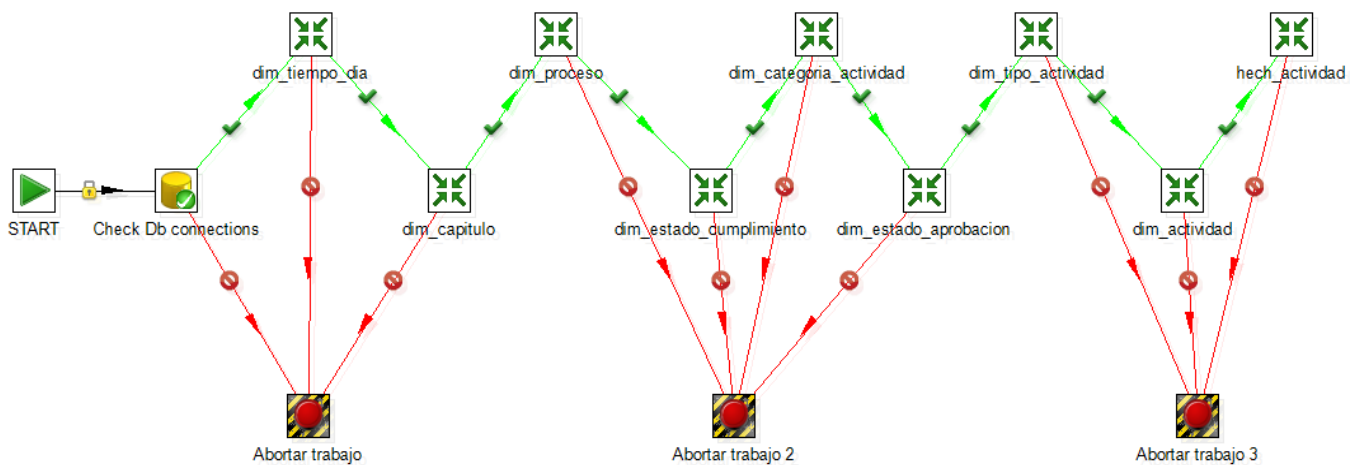


Figura 12. Trabajo del hecho actividad.

3.3 Implementación del subsistema de visualización

La implementación del subsistema de visualización es de vital importancia en el desarrollo de los Dashboard, pues define cómo se muestra la información para el análisis de los usuarios finales. En dicho subsistema se implementan los cubos OLAP y los reportes candidatos, además se definen los gráficos a utilizar.

Implementación del cubo multidimensional o cubo OLAP

Los cubos OLAP están orientados a bases de datos multidimensionales, permitiendo procesar grandes volúmenes de información en campos bien definidos y con un acceso inmediato a los datos para su consulta y posterior análisis. Los cubos OLAP representan una tecnología que mejora significativamente el proceso de análisis (50). Se modeló un cubo multidimensional el cual corresponde al hecho actividad, para ello se utilizó la herramienta Pentaho schema workbench. Dicho cubo contiene dimensiones que presentan jerarquías que van a estar compuestas por niveles que se corresponden con los campos definidos anteriormente en la base de datos. Seguidamente se muestra el cubo diseñado (Figura 13) y se explica detalladamente los elementos que lo componen.

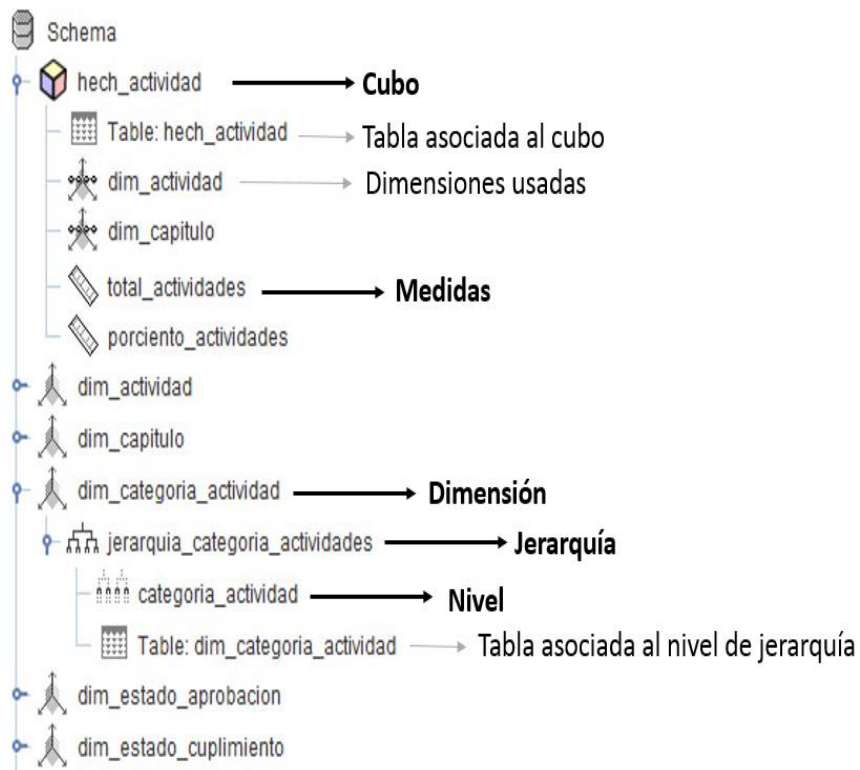


Figura 13. Diseño del cubo y elementos que lo componen.

Primeramente, se creó el cubo OLAP (Figura 14) para el hecho actividad en el cual se llenaron los parámetros *name* y *caption*, se seleccionaron las opciones *cache* y *enable* permitiendo la visibilidad del cubo. Para la implementación del mismo se especificó la tabla física con la que se asocia en la base de datos. Posteriormente se incluyeron las dimensiones correspondientes a cada cubo, heredando de esta forma las características que haya sido incluido en estas. Esto fue posible a través del elemento dimensiones usadas.

name	hech_actividad
description	
caption	hecho actividad
cache	<input checked="" type="checkbox"/>
enabled	<input checked="" type="checkbox"/>

Figura 14. Cubo hecho actividad.

Se definieron 2 medidas (Figura 15), todas calculadas a través de los campos de la BD, indicándose los parámetros: *name*, *aggregator*, *column*, *datatype* y *caption*. Además se selecciona la opción *visible* permitiendo la visibilidad de la medida.

Attribute	
name	total_actividades
description	
aggregator	sum
column	cantidad_actividades
formatString	
visible	<input checked="" type="checkbox"/>
datatype	Numeric
formatter	
caption	total de actividades

Figura 15. Medida total de actividades del hecho actividad.

Para el caso de las dimensiones se rellenaron algunos campos (Figura 16) como el *type*, seleccionando en todos los casos “*StandardDimension*”. Además se eligió la llave foránea en dependencia de la dimensión en la que se esté trabajando.

Attribute	
name	dim_categoria_actividad
description	
foreignKey	fk_dim_categoria_actividad_id
type	StandardDimension
usagePrefix	
caption	categoría actividad

Figura 16. Dimensión categoría actividad.

Después de elaboradas las dimensiones, se crearon 9 jerarquías (Figura 17). Los parámetros a seleccionar para cada una de estas serán *hasAll* y *allMemberName*, los cuales permitirán agrupar los valores. Otro de

los elementos a elegir será la primaryKey (llave primaria) la cual corresponde a la tabla dimensión en la cual se esté trabajando y por último se definirá el *caption* que será el nombre que representará a la jerarquía.

Attribute	
name	jerarquia_categoria_actividades
description	
hasAll	<input checked="" type="checkbox"/>
allMemberName	Todos
allMemberCaption	
allLevelName	
defaultMember	
memberReaderClass	
primaryKeyTable	
primaryKey	pk_dim_categoria_actividad_id
caption	categoria de actividades

Figura 17. Jerarquía de la dimensión categoría de actividad.

Para los niveles (Figura 18) que componen a las jerarquías se indicaron como atributos: *table* seleccionando la tabla a la cual pertenece y *column*, el cual se definirá la columna de la base de datos sobre la que se esté trabajando. Además se especificará los atributos como: *type*, *uniqueMembers*, *hideMemberIf* y *levelType*.

Attribute	
name	categoria_actividad
description	
table	dim_categoria_actividad
column	codigo_categoria_actividad
nameColumn	denominacion_categoria_actividad
parentColumn	
nullParentValue	
ordinalColumn	
type	String
uniqueMembers	<input checked="" type="checkbox"/>
levelType	Regular
hideMemberIf	Never
approxRowCount	
caption	categoria actividad
captionColumn	
formatter	

Figura 18. Nivel de la jerarquía categoría de actividad

Implementación de los reportes candidatos

Los reportes candidatos son la unidad básica que representa la información que el cliente desea ver como producto final. Estos se desarrollaron utilizando la herramienta Pentaho BI server 5.2.0, añadiendo algunas configuraciones para el uso propio del Dashboard. Una de estas configuraciones es la conexión única que se establece para el Dashboard nombrada “Actividades” en la cual se establecieron los parámetros necesarios para acceder a la BD. Además se crearon diversos roles como: administrador y analista garantizando de esta forma la seguridad del sistema.

Para la implementación de los reportes se emplearon las consultas MDX que son utilizadas por los sistemas OLAP. Seguidamente se muestra la consulta del reporte “Total de actividades por categorías”, la cual retorna como resultado la cantidad de actividades que hay por cada una de las categoría definidas.

```

MDX Query Editor
select NON EMPTY {[Measures].[cantidad_actividades]} ON COLUMNS,
  NON EMPTY Hierarchize(Union({[dim_categoria_actividad].[Todos]}, [dim_categoria_actividad].[Todos].Children)) ON ROWS
from [hech_actividad]
  
```

Figura 19. Consulta MDX.

Estos reportes serán utilizados como punto de partida para el proceso de toma de decisiones. A continuación, se muestra la imagen de la vista de análisis “Total de actividades por categorías” perteneciente al “LT Total de actividades” (Figura 20).

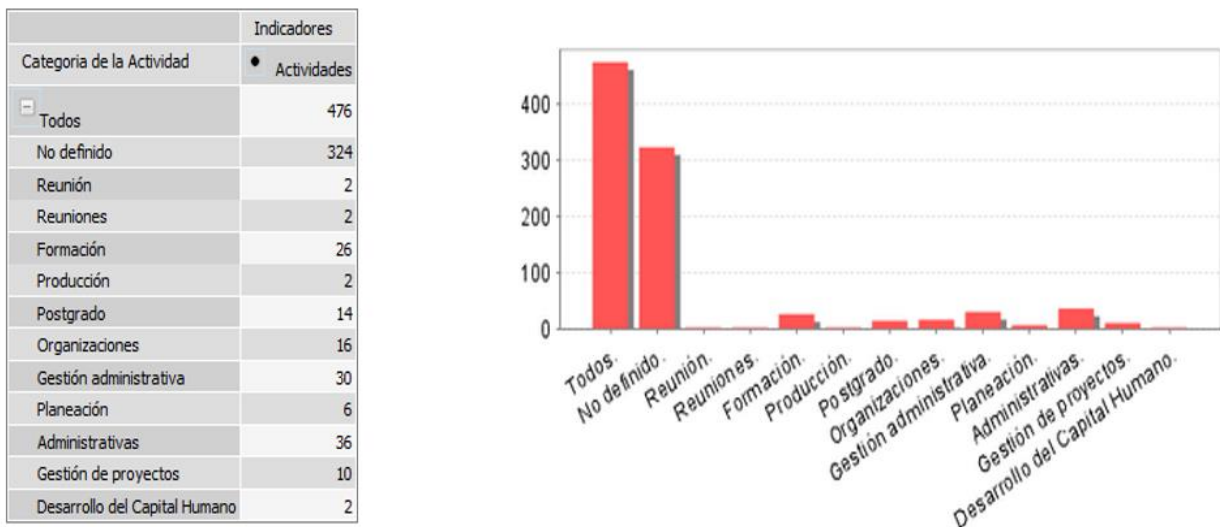


Figura 20. Vista del reporte: “Total de actividades por categoría”.

Gráficos dinámicos

Un gráfico dinámico permite mostrar los datos de forma dinámica y posibilita manejar la información según se desee. Seguidamente se describen los tipos de gráficos que se utilizaron para el desarrollo del Dashboard (51):

- **Gráficos de columnas:** este tipo de gráfico es útil para mostrar los cambios de los datos en un período de tiempo o para ilustrar comparaciones entre elementos. A menudo en estos gráficos las categorías se suelen organizar a lo largo del eje horizontal, mientras que los valores se representan a lo largo del eje vertical.
- **Gráficos de barras:** muestran comparaciones entre elementos individuales. A menudo en estos gráficos las categorías se suelen organizar a lo largo del eje vertical, mientras que los valores lo hacen a lo largo del eje horizontal.
- **Gráficos de líneas:** pueden mostrar datos continuos en el tiempo, establecidos frente a una escala común y, por tanto, son ideales para mostrar tendencias en datos a intervalos iguales. En un gráfico de líneas, los datos de categoría se distribuyen uniformemente en el eje horizontal y todos los datos de valor se distribuyen uniformemente en el eje vertical.
- **Gráficos circulares:** muestran el tamaño de los elementos de una serie de datos, en proporción a la suma de los elementos. Los puntos de datos de un gráfico circular se muestran como porcentajes del total del gráfico circular.

Después de descritos los gráficos utilizados en el Dashboard se hace necesario explicar cómo estos trabajan de forma dinámica, permitiendo a los usuarios interactuar con ellos. Un ejemplo práctico es el de los gráficos de líneas en los cuales se encuentra la opción “año” que permite visualizar la información referente al año seleccionado (Figura 21). Se pueden consultar el resto de las especificaciones de los gráficos dinámicos en el Anexo 2.

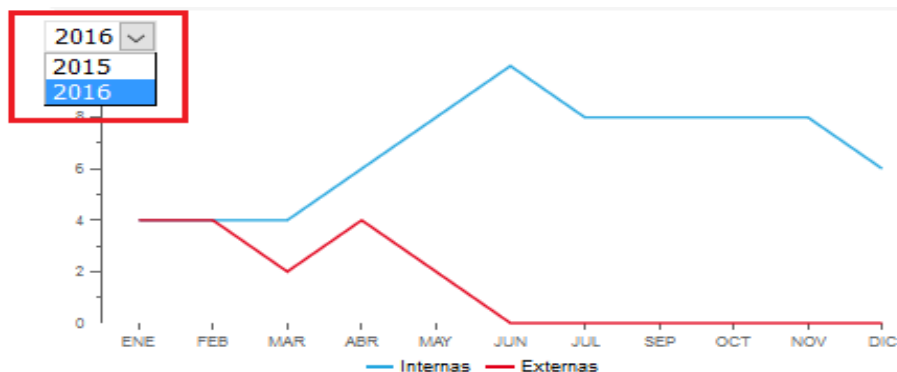


Figura 21. Visualizar información por año.

Otra de las opciones que permite el gráfico es poder visualizar todo o una parte de la información, lo cual se realiza marcando la opción que no se desee visualizar (Figura 22).

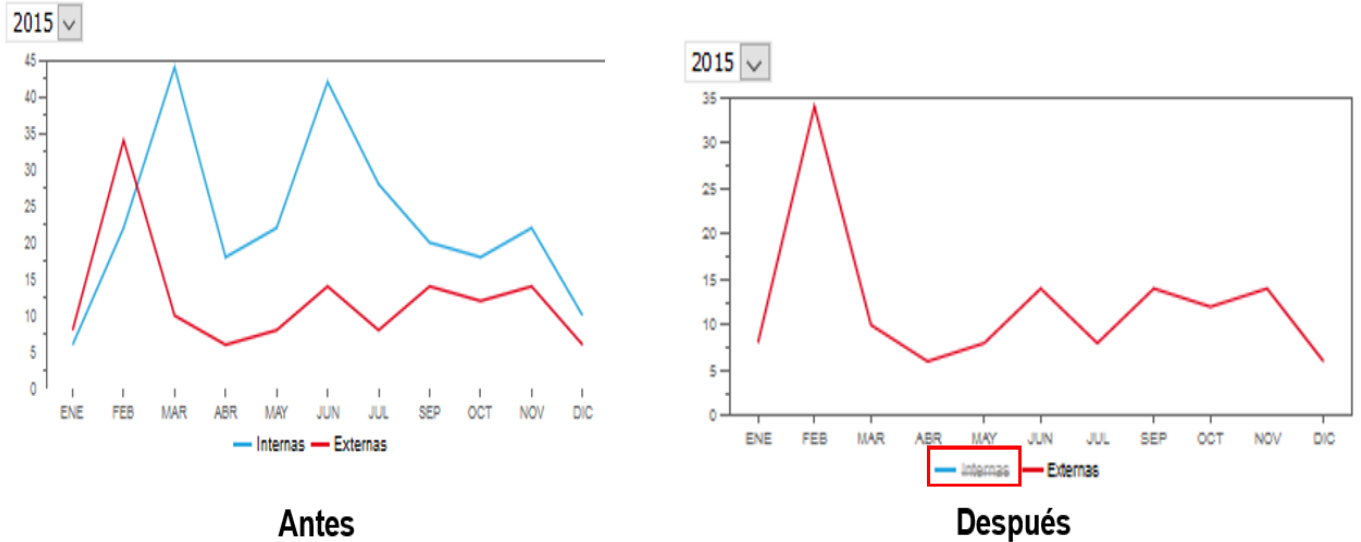


Figura 22. Visualizar la información solicitada.

Además estos gráficos permiten que al situar el mouse sobre ellos se muestre un mensaje con datos específicos sobre el mes que se desee visualizar (Figura 23).

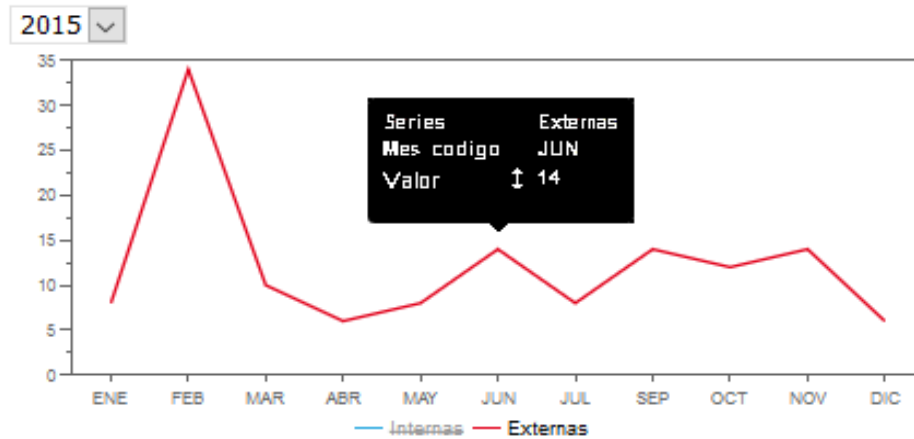


Figura 23. Visualizar información por mes.

3.4 Pruebas de software

Para determinar el nivel de calidad de un software se deben efectuar pruebas que permitan comprobar el grado de cumplimiento de las especificaciones del sistema. De acuerdo a la IEEE¹⁴ una prueba se define como: “*Actividad en la cual un sistema o componente es ejecutado bajo condiciones específicas, se observan o almacenan los resultados y se realiza una evaluación de algún aspecto del sistema o componente*” (52).

El proceso de pruebas comienza con la planificación, posteriormente la ejecución y control y finalmente la evaluación de las mismas, permitiendo comprobar que no existan defectos durante el proceso de implementación. Por lo que se hizo necesario aplicar diferentes pruebas con el propósito de verificar no solo que el producto cumpla con los requisitos del cliente, sino también para eliminar los posibles errores que pudiera tener.

Niveles de pruebas

El proceso de prueba se ejecuta en varios niveles y estos se encuentran estrechamente relacionados con los tipos de pruebas. Seguidamente se describen los niveles de pruebas a utilizar:

- **Pruebas unitarias:** consisten en probar diferentes componentes de la solución que tengan funcionalidades específicas del producto; su objetivo es asegurar el correcto funcionamiento de todas las partes del subsistema de manera independiente (26).
- **Pruebas de integración:** permiten verificar la correcta integración de los componentes y subsistemas que conforman la solución. Su objetivo principal es probar el sistema como un todo, pues los subsistemas por separados pueden funcionar correctamente, pero al ser integrados estos pueden ser no compatibles (53).
- **Pruebas de sistema:** verifican el funcionamiento del sistema como un todo. Comprueban las funcionales después de que el software y el hardware han sido integrados, permitiendo demostrar que el sistema satisface los requerimientos y es apto para su uso final (54).

14 Instituto de Ingenieros Eléctricos y Electrónicos.

Métodos de prueba

Para la validación de la propuesta de solución será aplicado el método de caja negra, el cual permitirá corregir problemas en la interfaz y se comprobará que la propuesta de solución realice las funciones requeridas por el usuario. Estas pruebas serán documentadas mediante los casos de pruebas.

3.5 Modelo V

Para validar el correcto funcionamiento del Dashboard se utilizó el Modelo V (Figura 24), el mismo fue definido por CALISOFT¹⁵. El principal objetivo de este modelo es crear un estándar de comprobación para que los productos cumplan con las especificaciones del negocio y así garantizar la calidad del producto final. Este modelo constituye una evolución del Modelo en Cascada, con la diferencia de que las actividades de prueba se ejecutan paralelamente con las actividades de análisis y diseño. Dicho modelo plantea un conjunto de pruebas, de las cuales solo se utilizarán las descritas anteriormente. Cada una de estas pruebas se encargará de prevenir futuros errores en los requerimientos, el diseño y la implementación del sistema (55).

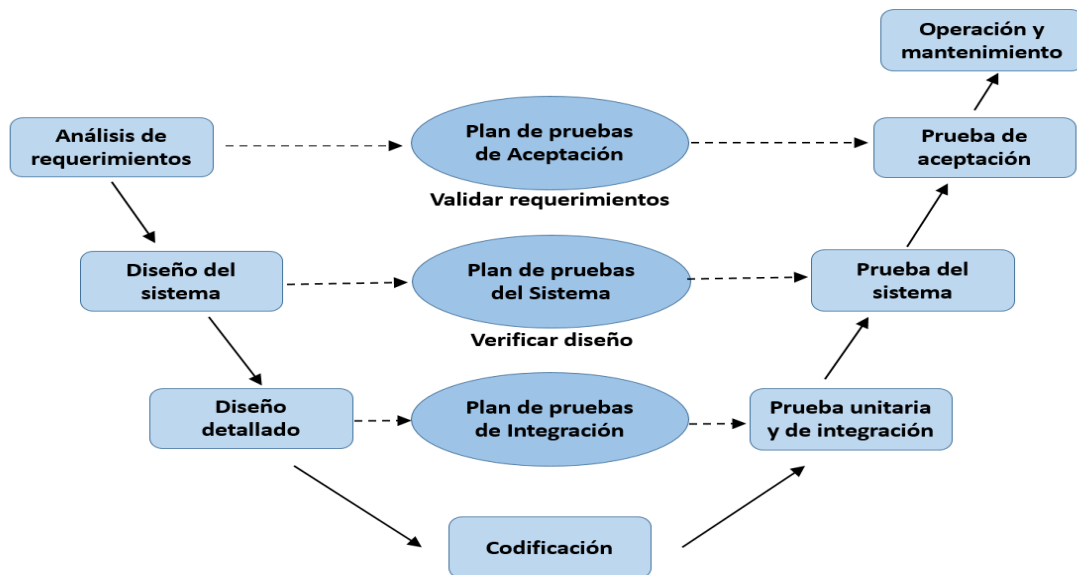


Figura 24. Modelo V (56).

¹⁵ Centro Nacional de Calidad del Software: Tiene como principales objetivos ofrecer servicios de evaluación de la calidad de los proceso de desarrollo de software. Además, tiene la tarea de establecer estándares técnicos y procedimientos que normalicen la industria informática, dirigiéndola hacia niveles de calidad superiores.

3.6 Herramientas para la aplicación de pruebas

Como herramientas para aplicar los tipos de pruebas descritos anteriormente se utilizaron las listas de chequeo, los casos de prueba (CP) para caso de uso de información (CUI) y el perfilado de datos de la BD destino.

Aplicación de los casos de prueba

Los casos de prueba permiten verificar la calidad de un software. Son utilizados para identificar posibles errores de implementación y comprobar el grado de desempeño de los requisitos iniciales del sistema. Si estos no son correctos, se pone en duda la calidad del sistema y las pruebas dejan de ser confiables (57).

Aplicación de los casos de prueba por caso de uso de información

Se desarrollaron 4 CP asociados cada uno a los CUI identificados en la etapa de análisis. Luego de realizado todos los CP, se cumplirá una de las dos condiciones posibles:

- Se detecta uno o varios errores en la funcionalidad y se genera una lista con las no conformidades encontradas.
- La funcionalidad cumple con la especificación y se acepta.

En la siguiente tabla se muestra el CP correspondiente al CUI: “Presentar información de Elaboración del Plan”.

Tabla 9. Caso de prueba asociado al caso de uso de información: “Presentar información de Elaboración del Plan”.

Escenario	Descripción	Variables de salida	Respuesta del sistema	Flujo central
1.1 Presentar información de Elaboración del Plan.	El CU inicia cuando el analista solicita realizar un análisis de la información referente a la elaboración del plan	Total de actividades. Porcentaje de actividades.	Se muestra la información referente al análisis de las actividades referentes a la elaboración del plan.	<ul style="list-style-type: none"> • Se selecciona el A.A.G_Dashboard_SIPAC. • Selecciona el AA_Elaboración_Plan. • Selecciona el LT_total_actividades o LT_porcentaje_actividades. • En el área de trabajo se visualiza la tabla correspondiente al reporte. • Muestra la información contenida en el reporte seleccionado.

Listas de chequeo

Las listas de chequeo son un mecanismo para controlar los riesgos. Su función básica consiste en poder identificar problemas comunes, encontrar errores y poder solucionarlos en un breve espacio de tiempo. Constituyen un listado o conjunto de preguntas sobre un aspecto determinado en forma de cuestionario, lo que posibilita verificar el grado de cumplimiento de determinadas reglas. Esta evaluación se desarrolla a través del análisis de un grupo de indicadores, distribuidos en tres secciones fundamentales (58):

- **Estructura del documento:** abarca todos los aspectos definidos por el expediente de proyecto o el formato establecido por el proyecto.
- **Indicadores definidos:** abarca todos los indicadores a evaluar durante la etapa de desarrollo del mercado.
- **Semántica del documento:** contempla todos los indicadores a evaluar respecto a la ortografía y redacción.

Las listas de chequeo se le aplicaron a los artefactos de ETL Registro del sistema fuente (RSF) y Perfilado de datos (PD). Seguidamente se muestra una tabla en la que se encuentran los principales aspectos que fueron evaluados y los resultados que arrojó la aplicación de las mismas.

Tabla 10. Aplicación de las listas de chequeo a los artefactos ETL.

Secciones	RSF	PD
Estructura	5	8
Indicadores	4	6
Semántica	5	5
TI	7	3
IC	5	8
NC	9	3
NC-IC	3	7

- **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.

Seguidamente se muestra un ejemplo de una lista de chequeo aplicada al artefacto Perfilado de datos.

Tabla 11. Lista de chequeo aplicada al Perfilado de datos.

Estructura del documento					
Peso	Indicadores a evaluar	Eval	(NP)	Cant 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	¿El objetivo expresa correctamente el propósito del documento?	0		0	
	¿En el entregable, la definición de las variables se hace correctamente?	0		0	
	¿Existe una adecuada correspondencia entre las variables definidas y las descripciones que tienen estas variables?	0		0	
	8. ¿En el entregable se crea una hoja por cada variable definida?	1		0	
	9. ¿Queda registrado en el entregable todos los posibles valores que van a tener las variables definidas?	0		0	
Indicadores definidos en el desarrollo					
Peso	Indicadores a evaluar	Eval	(NP)	Cant 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)	Cant 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	
	¿El número de página que aparece en el índice coincide con el contenido que se refleja realmente en dicha página?	0		0	

3.7 Calidad de los datos

Poder certificar que los datos cargados posean la calidad requerida es de vital importancia en el desarrollo de los Dashboard, pues de esta forma se puede afirmar que los datos no contienen errores. Este proceso de validación se ejecuta mediante el perfilado y auditoría de los datos luego de concluir el proceso ETL.

Perfilado de los datos

El perfilado de datos se realizó con el objetivo de verificar que los datos posean la calidad requerida después de culminado el proceso ETL. Para la realización del perfilado se empleó la herramienta DataCleaner 1.5.4 permitiendo obtener estadísticas e información sobre los datos. Los resultados arrojados por este proceso indican que después de poblada la base de datos no existe valores vacíos ni nulos, por lo que se puede determinar que son satisfactorios. (Figura 25)

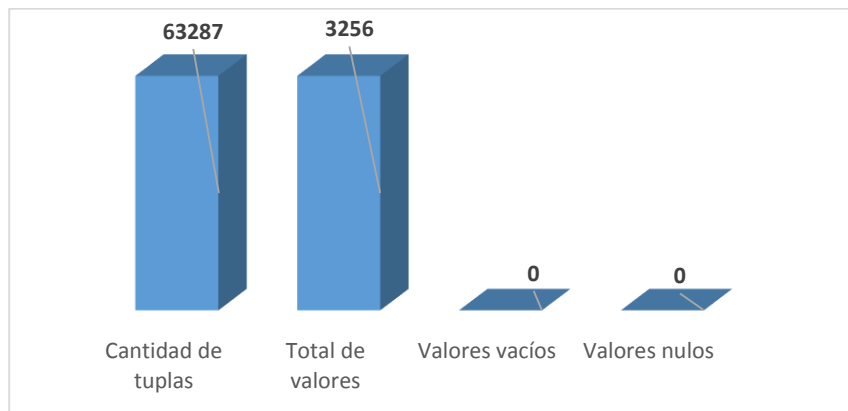


Figura 25. Calidad de los datos después de realizado el proceso ETL.

Auditoría de los datos

La auditoría de datos es el proceso de gestionar cómo los datos se ajustan a los propósitos definidos por la organización, permitiendo obtener el conocimiento referente a la confiabilidad de los mismos y la información asociada a la ejecución de las transformaciones (59). La estrategia definida para auditar los datos almacenados se basa principalmente en el uso de las tablas de metadatos, las cuales proporcionan la información necesaria para comprobar y garantizar que los datos cargados sean confiables.

3.8 Resultado de las pruebas

En el presente apartado se refleja el resultado de las pruebas realizadas a la aplicación.

Pruebas unitarias y de integración

Una vez culminada la implementación se realizaron pruebas unitarias siendo detectadas 7 no conformidades, todas de complejidad baja las cuales fueron resueltas satisfactoriamente (Figura 26).

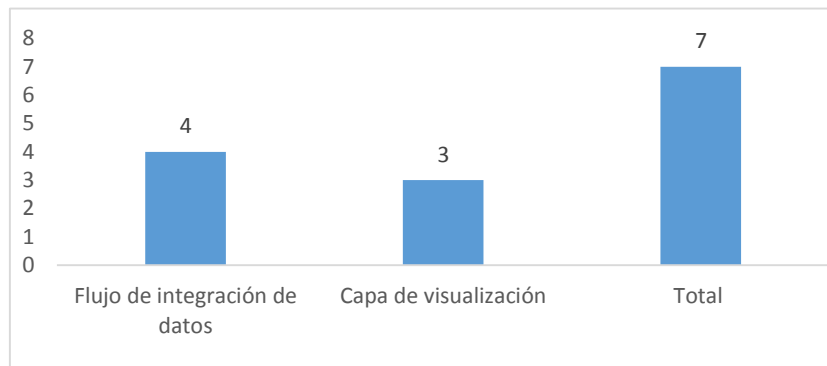


Figura 26. Comportamiento de los indicadores medidos en las pruebas unitarias.

Listas de chequeo

Se aplicaron dos iteraciones de las listas de chequeo a los artefactos RSF y PD, obteniendo en la primera 15 no conformidades las cuales se solucionaron satisfactoriamente y en la segunda iteración no se encontraron no conformidades. Seguidamente se muestra el comportamiento de los indicadores medidos en las listas de chequeos.

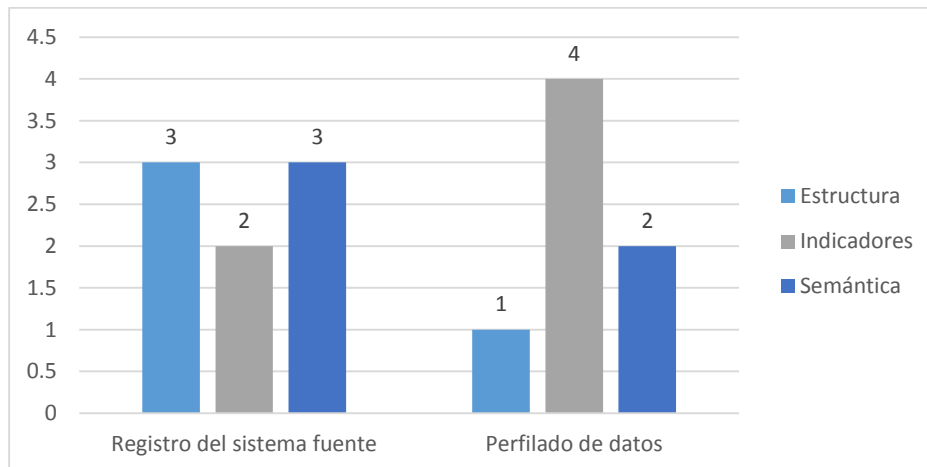


Figura 27. Comportamiento de los indicadores medidos en las listas de chequeo.

Pruebas de Caja Negra

A partir del diseño y ejecución de los casos de prueba de ir probando si todas las funcionalidades se cumplen, se detectaron errores como: funcionalidades incorrectas, errores en las consultas a la base de datos y al cubo, obteniéndose un total de 5 no conformidades, las cuales fueron resueltas satisfactoriamente.

Pruebas de aceptación

Se realizaron las pruebas de aceptación en conjunto con el cliente las cuales arrojaron resultados satisfactorios, quedando verificado que están satisfechos con el producto desarrollado (ver Anexo 3).

3.9 Conclusiones del capítulo

Luego de realizarse la implementación y prueba del Dashboard para el Sistema de Planificación de Actividades se concluye que:

- Con la implementación del subsistema de almacenamiento de datos se definen los estándares de codificación y la estructura física de la base de datos, obteniendo 2 esquemas, 10 tablas y 9 secuencias.
- Con la implementación del subsistema de integración de datos se ejecutó 1 trabajos y 9 transformaciones permitiendo poblar el almacén de datos.
- Con la implementación del subsistema de visualización se diseñó el cubo OLAP, se implementaron los reportes candidatos y se definieron los tipos de gráficos a utilizar.
- Se aplicaron pruebas unitarias, de integración y de aceptación utilizando como herramientas las listas de chequeo, los casos de prueba y el perfilado de los datos obteniendo como resultado un producto con calidad que cumple con los requisitos identificados.

Conclusiones Generales

Al finalizar la investigación se puede afirmar que se le ha dado cumplimiento de forma satisfactoria al objetivo general del presente trabajo de diploma, por lo que se concluye:

- 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.
- A través del análisis y diseño del Dashboard se obtuvo y describió la solución a implementar teniendo en cuenta las necesidades del cliente.
- La implementación del Dashboard para SIPAC permitió agrupar y centralizar los indicadores claves de planificación los cuales serán visualizados a través gráficos dinámicos, permitiendo obtener una solución que responde a las necesidades del cliente.
- Las pruebas efectuadas mediante la aplicación de los casos de pruebas, el perfilado de los datos y las listas de chequeo, permitieron comprobar la calidad del producto a partir de los requisitos establecidos, certificando que la solución tiene la calidad correcta para su utilización.

Recomendaciones

Con el propósito de mejorar la propuesta realizada en este trabajo, se sugiere:

- Incluir en el Dashboard para SIPAC gráficos y reportes candidatos que respondan a las perspectivas del análisis: objetivos y planes.
- Realizar pruebas de rendimiento a la BD del Dashboard, teniendo en cuenta que se realizan cargas semanales por lo que el número de tuplas debe aumentar considerablemente.

Referencias

1. **GARCÍA, J. González.** Perspectiva ambiental del cuadro de mando integral de una organización: aplicación de la inteligencia de negocio. *Ingeniería* . 2007. 448.
2. **Few, Stephen.** Information Dashboard Design. The effective Visual Communication of data.
3. —. *Common Pitfalls in Dashboard Design*. s.l. : Proclarity Corporation, 2010.
4. Sinexus. *Sinexus*. [En línea] [Citado el: 23 de enero de 2016.] www.sinnexus.com/business_intelligence.
5. Forrester Research. *Forrester Research*. [En línea] 21 de noviembre de 2012. [Citado el: 8 de diciembre de 2015.] <http://www.forrester.com/Topic+Overview+Business+Intelligence>.
6. Drigg, Fernando. Información Profesional. *Información Profesional*. [En línea] [Citado el: 31 de enero de 2016.] www.iprofesional.com/notas/51865-Los-beneficios-de-la-inteligencia-de-negocio.html.
7. Dapena Bosquet, Isabel , Muñoz San Roque, Antonio y Sánchez Miralles, Alvaro . *Sistemas de Información Orientados a la Toma de Decisiones: el enfoque multidimensional*. Madrid, España : s.n., 2005. ISSN 0003-2506.
8. Abril Frade, Diego Orlando y Pérez Castillo, José Nelson . *Estado actual de las tecnologías de bodega de datos y OLAP aplicadas a base de datos espaciales*. Bogotá : Univesidad de Bogotá, 2007.
9. Gupta, H. y Harinarayan, V. Index selection for OLAP. s.l. : IEEE. ISSN: 1063-6382.
10. Becker, Shirley. *Data Warehousing and Web Engineering*. Hershey, Estados Unidos : Idea Group Publishing, 2002. ISBN 1931777020.
11. Kaplan, Robert S. y Norton, David P. *The Balance Scorecard*. Cambridge, Estados Unidos : Harvard Business Review, 1992.
12. Rui, Luis Eduardo Ayala. *Philanthropic Projects Portal*.
13. Dashboard Zone. *Dashboard Zone*. [En línea] [Citado el: 9 de enero de 2016.] www.dashboardzone.com/what-is-a-dashboard.
14. Bimetrics. 2010.
15. López., Luis Fernando Montero. *Posgrados y Educación Ejecutiva*. 12 de octubre de 2012.

16. Dávila, Antonio. Nuevas herramientas de control: El cuadro de Mando Integral. s.l. : Revista de Antiguos Alumnos, 2006.
17. Warren, J. Key Performane Indicator. 2009.
18. Gestipolis. s.l. : Consulting Group Sixtina, 2008.
19. Pedrasa González, Yaniesis y Rojas Ricardo, Edgar. Almacén de datos Sala situacional. [En línea] 2012. [Citado el: 20 de diciembre de 2016.] <http://publicaciones.uci.cu/index.php/SC>.
20. Kimball, Ralph y Caserta, Joe. The Data Warehouse ETL Toolkit. Practical Techniques for Extracting, Cleaning Conforming and Delivering Data. Canada : Wiley Publishing, 2004. ISBN 0-764-57923-1.
21. Kimball, Ralph, y otros. The Data Warehouse Lifecycle Toolkit. Expert Methods for Designing, Developing, and Deploying Data Warehouses. Canadá : Wiley Publishing.
22. Urban Dashboard. *Urban Dashboard*. [En línea] [Citado el: 27 de enero de 2016.] http://www.urbandashboard.org/iadb/index_city.html?id=MVD&lang=ES.
23. Scytl Control Dashboard. *Scytl Control Dashboard*. [En línea] [Citado el: 28 de noviembre de 2015.] <https://www.scytl.com/es/products/fase-preelectoral/scytl-control-dashboard/>.
24. Nicasoft Tecnología de la Programación. *Nicasoft Tecnología de la Programación*. [En línea] 2014. [Citado el: 20 de diciembre de 2015.] <http://www.nicasoft.com.ni/NDC.aspx>.
25. French, Marcelo. Sixtina Consulting Group. *Sixtina Consulting Group*. [En línea] 30 de marzo de 2011. [Citado el: 29 de enero de 2016.] www.sixtinagroup.com/db-bsc-diferencia/ .
26. Pressman, Roger S. Software engineering: a practitioner's approach. 2010.
27. Hernández, Ing. Yanisbel González. Metodología de desarrollo para proyectos de almacenes de datos. Habana : s.n., 2013.
28. Software. *Software*. [En línea] 2 de enero de 2011. [Citado el: 28 de enero de 2016.] <http://www.software.com.ar/p/visual-paradigm-para-uml>.
29. PostgreSQL. *PostgreSQL*. [En línea] 2011. [Citado el: 7 de enero de 2016.] <http://postgresql.uci.cu/>.
30. Stonebraker, Michael . The Postgres Next Generation Database Management Systems. 2009.

31. Levin, Jonathan. Pgadmin. [En línea] [Citado el: 15 de enero de 2016.] <https://www.pgadmin.org/>.
32. Duplika. *Duplika*. [En línea] 2 de marzo de 2012. [Citado el: 15 de diciembre de 2016.] www.duplika.com/que-son-los-servidores-web-y-por-que-son-necesarios.
33. Moodie, Matthew. Pro Apache Tomcat 6. New York : Springer Verlag. ISBN-13: 978-1-59059-785-9.
34. Pulvirenti, Adrián Sergio y Roldán, María Carina. Pentaho Data Integration. s.l. : Packt Publishing, 2013. ISBN: 1783280670 .
35. Campazzo, Eduardo Nicolas y Santos, Virginia Ines. Business Intelligence: Negocios Inteligente para Empresas Inteligentes. Argentina : Universidad Nacional de la Rioja, 2007.
36. Pentaho Open Source Business Intelligence. *Pentaho Open Source Business Intelligence*. [En línea] [Citado el: 30 de noviembre de 2015.] www.pentaho.com.
37. Webdetails a Pentaho Company. *Webdetails a Pentaho Company*. [En línea] 2016. [Citado el: 7 de febrero de 2016.] <http://www.webdetails.pt/ctools/cde/>.
38. Amador, Mag Audrey. La ingeniería de requerimientos y su importancia en el desarrollo de proyectos de software. Costa Rica : Revista InterSedes, 2006. ISSN 1409-4746.
39. Schiefer, Josef, List, Beate y Bruckner, Robert. A Holistic Approach for managing requirements of data warehouse systems. . Vienna : Vienna University of Technology, 2002.
40. Dataprix. *Dataprix*. [En línea] [Citado el: 28 de febrero de 2016.] www.dataprix.com/category/integraciondatos/integracion-datos/calidad-datos/limpieza-datos/perfilado-datos.
41. Alegsa, Leandro. Diccionario de Informática y Tecnología. [En línea] 12 de mayo de 2010. [Citado el: 12 de marzo de 2016.] <http://www.alegsa.com.ar/Dic/arquitectura%20de%20sistemas.php>.
42. Sotolongo León, Anthony Rafael. Modelo de descripción de arquitectura de almacenes de datos para ensayos clínicos del Centro de Inmunología Molecular. Habana : s.n., 2010.
43. Bertolín, Javier Areitio. Seguridad de la Información. Redes, informática y sistemas de información. Madrid, España : Cengage Learning Paraninfo S.A., 2008. ISBN: 978-84-9732-502-8.
44. Apache Tomcat. *Apache Tomcat*. [En línea] [Citado el: 5 de marzo de 2016.] <http://tomcat.apache.org>.

45. PostgreSQL. *PostgreSQL*. [En línea] The PostgreSQL Global Development Group , 2016. [Citado el: 26 de abril de 2016.] <http://www.postgresql.org/docs/9.2/static/indexes-types.html>.
46. Kimball, Ralph y Ross, Margy . *The Definitive Guide to Dimensional Modeling*. Indianapolis : John Wiley & Song, 2013. 978-1-118-53080-1.
47. Baca, Murtha. *Introduction to Metadata*. Los Angeles : Getty Research Institute, 2008. ISBN: 978-0-89236-896.
48. Curto Díaz, Josep y Canesa, Jordi. *Introducción al Business Intelligence*. Barcelona : UOC, 2010. 978-84-9788-886-8.
49. Ross, Margy y Kimball, Ralph. *The Data Warehouse Toolkit: the Complete Guide to Dimensional Modelling*. New York : John Wiley & Sons, 2012. 0-471-200247.
50. Informática_Hoy. *Informática_Hoy*. [En línea] 2016. [Citado el: 29 de abril de 2016.] <http://www.informatica-hoy.com.ar/Cubo-OLAP-una-base-de-datos-multidimensional.php>.
51. Microsoft. *Microsoft*. [En línea] 2016. [Citado el: 2 de mayo de 2016.] <https://support.office.com/es-es/article/Tipos-de-gr%C3%A1ficos-disponibles-a6187218-807e-4103-9e0a-27cdb19afb90#bmcolumcharts>.
52. Institute of Electrical and Electronics Engineers. *IEEE Standard Glossary of Software Engineering Terminology*. s.l. : IEEE, 2000. 155937067X.
53. Mansilla, Ricardo. Slideshare. *Slideshare*. [En línea] 2016. [Citado el: 30 de abril de 2016.] <http://www.slideshare.net/cliceduca/pruebas-de-software-2420588>.
54. Bueno, Carlos Blanco. *Ingeniería del Software*. 2013. Creative Commons BY-NC-SA 3.0.
55. Forsberg, Kevin y Cotterman, Howard . *Visualizing Project Management m*. Nueva York : John Wiley and Sons, 1997. Vol. 18, 1.
56. Ling, Pang. *The Research of V Model in Testing Embedded Software*. *En Computer Science and Information Technolog*. 2008. Vol. vol. 7.
57. Estero Botaro, A. y Medina Bulo, I. *Una arquitectura para la generación de casos de pruebas*. Chile : Universidad de Cádiz, 2009.
58. Lovelle, Juan Manuel Cueva. *Calidad del Software*. España : Universidad de España, 1999.

59. Mustelier, Doris Medina. Data Integration Info. [En línea] 2016. [Citado el: 8 de mayo de 2016.]

<http://www.dataintegration.info/etl>.

60. Bentham, Jeremy. TCP/IP lean : web server for embedded systems. Lawrence, Kansas : s.n., 2009.

Anexos

Anexo 1: Especificación de casos de uso de información

CUI: Presentar información de Aprobación-Conciliación del Plan.

Objetivo	Presentar información de Aprobación-Conciliación del Plan.	
Actores	Analista	
Resumen	El caso de uso comienza cuando el analista solicita realizar un análisis de la información referente a la aprobación-conciliación del plan. El sistema genera los reportes referentes con la información seleccionada y las opciones de los posibles cambios que se le puede realizar al mismo. El CU finaliza cuando el actor termina el análisis de la información.	
Complejidad	Media.	
Prioridad	Media.	
Precondiciones	Debe estar seleccionada el A.A.G_Dashboard_SIPAC.	
Postcondiciones	Se consultaron los reportes pertenecientes al CU Presentar información de Aprobación Conciliación del Plan.	
Flujo de eventos		
Flujo básico: Presentar Información de Aprobación-Conciliación del Plan.		
	Actor	Sistema
1	Selecciona el área de análisis A.A_Aprobación_Conciliación_Plan.	
2		Muestra los libros de trabajo contenidos en el A.A_Aprobación_Conciliación_Plan.
3	Selecciona el libro de trabajo LT_total_actividades o LT_porcentaje_actividades.	
4		Muestra los reportes contenidos en el LT_total_actividades o LT_porcentaje_actividades.
5	Selecciona el reporte que desea analizar.	

6		Muestra la información contenida en el reporte seleccionado y ofrece opciones al actor para realizar cambios al reporte durante su análisis. Ir al CU Realizar operaciones sobre reportes. Se finaliza el caso de uso
Opciones de reportes de Presentar información de Aprobación-Conciliación del Plan.		
Perspectivas de análisis	Posibles resultados	
	Medidas	Periodicidad
<ul style="list-style-type: none"> • Estado aprobación • Estado de cumplimiento 	<ul style="list-style-type: none"> • Cantidad de actividades • Porciento de Actividades 	Diaria
Relaciones	CU Extendidos	CU Realizar operaciones sobre reportes

CUI: Presentar información de Puntualización del Plan.

Objetivo	Presentar información de Puntualización del Plan.	
Actores	Analista	
Resumen	El caso de uso comienza cuando el analista solicita realizar un análisis de la información referente a la puntualización del plan. El sistema genera los reportes referentes con la información seleccionada y las opciones de los posibles cambios que se le puede realizar al mismo. El CU finaliza cuando el actor termina el análisis de la información.	
Complejidad	Media.	
Prioridad	Media.	
Precondiciones	Debe estar seleccionada el A.A.G_Dashboard_SIPAC.	
Postcondiciones	Se consultaron los reportes pertenecientes al CU Presentar información de Puntualización del Plan.	
Flujo de eventos		
Flujo básico: Presentar Información de Elaboración del Plan.		
	Actor	Sistema
1	Selecciona el área de análisis	

	A.A_Puntualización_Plan.	
2		Muestra los libros de trabajo contenidos en el A.A_Puntualización_Plan.
3	Selecciona el libro de trabajo LT_total_actividades o LT_porcentaje_actividades.	
4		Muestra los reportes contenidos en el LT_total_actividades o LT_porcentaje_actividades.
5	Selecciona el reporte que desea analizar.	
6		Muestra la información contenida en el reporte seleccionado y ofrece opciones al actor para realizar cambios al reporte durante su análisis. Ir al CU Realizar operaciones sobre reportes. Se finaliza el caso de uso

Opciones de reportes de Presentar información Puntualización del Plan.

Perspectivas de análisis	Posibles resultados	
	Medidas	Periodicidad
<ul style="list-style-type: none"> • Puntualizaciones 	<ul style="list-style-type: none"> • Cantidad de actividades • Porcentaje de Actividades 	Diaria
Relaciones	CU Extendidos	CU Realizar operaciones sobre reportes

CUI: Presentar información de Ejecución-Control del Plan.

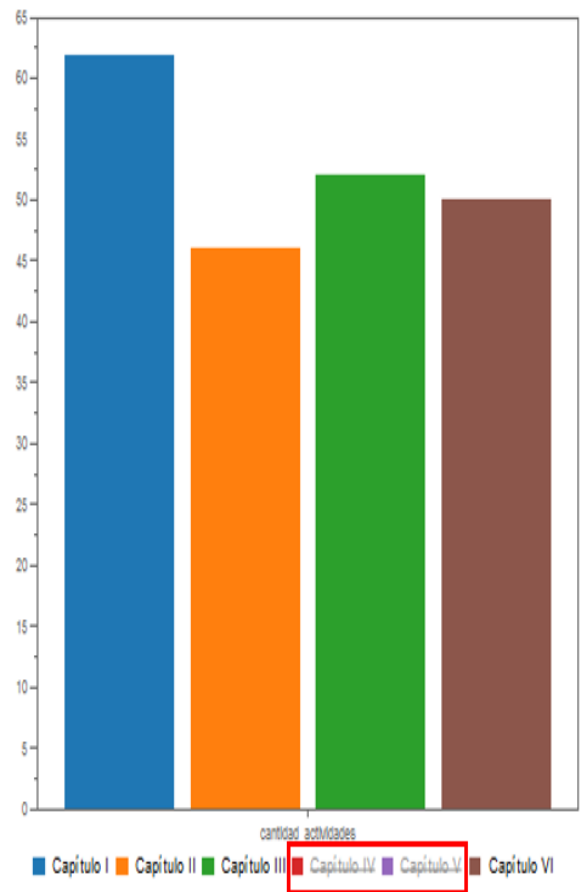
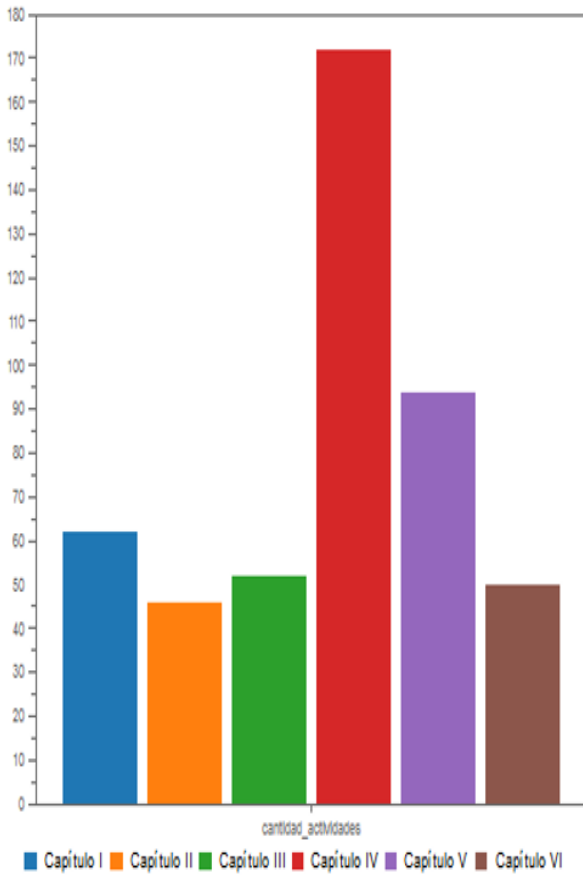
Objetivo	Presentar información de Ejecución-Control del Plan.
Actores	Analista
Resumen	El caso de uso comienza cuando el analista solicita realizar un análisis de la información referente a la ejecución-control del plan. El sistema genera los reportes referentes con la información seleccionada y las opciones de los posibles cambios que se le puede realizar al mismo. El CU finaliza cuando el actor termina el análisis de la información.
Complejidad	Media.

Prioridad	Media.	
Precondiciones	Debe estar seleccionada el A.A.G_Dashboard_SIPAC.	
Postcondiciones	Se consultaron los reportes pertenecientes al CU Presentar información de Ejecución-Control del Plan.	
Flujo de eventos		
Flujo básico: Presentar Información de Ejecución-Control del Plan.		
	Actor	Sistema
1	Selecciona el área de análisis A.A_Ejecución_Control_Plan.	
2		Muestra los libros de trabajo contenidos en el A.A_Ejecución_Control_Plan
3	Selecciona el libro de trabajo LT_total_actividades o LT_porcentaje_actividades.	
4		Muestra los reportes contenidos en el LT_total_actividades o LT_porcentaje_actividades.
5	Selecciona el reporte que desea analizar.	
6		Muestra la información contenida en el reporte seleccionado y ofrece opciones al actor para realizar cambios al reporte durante su análisis. Ir al CU Realizar operaciones sobre reportes. Se finaliza el caso de uso
Opciones de reportes de Presentar información Puntualización del Plan.		
Perspectivas de análisis	Posibles resultados	
	Medidas	Periodicidad
<ul style="list-style-type: none"> • Proceso • Capítulo 	<ul style="list-style-type: none"> • Cantidad de actividades • Porcentaje de Actividades 	Diaria
Relaciones	CU Extendidos	CU Realizar operaciones sobre reportes

Anexo 2: Gráficos dinámicos

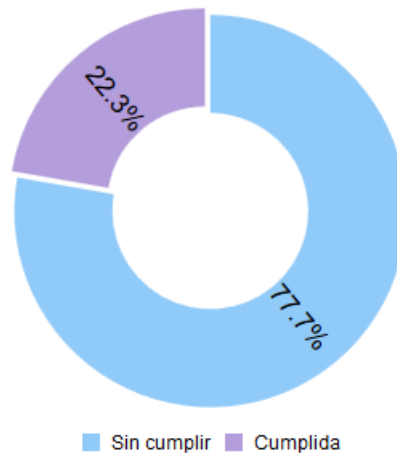
Gráfico de barras

Permite mostrar la cantidad de actividades por capítulos, además tiene la opción de seleccionar los capítulos que serán mostrados.

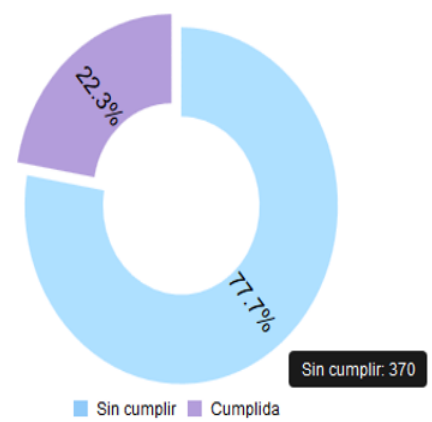
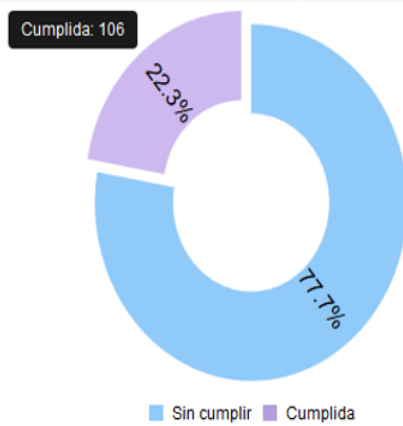


Gráficos Circulares

Este tipo de gráfico presenta los valores en forma de porcentos.



Además permiten que al ser seleccionada algunas de sus áreas muestre un mensaje con la cantidad de actividades.



Gráficos de columnas

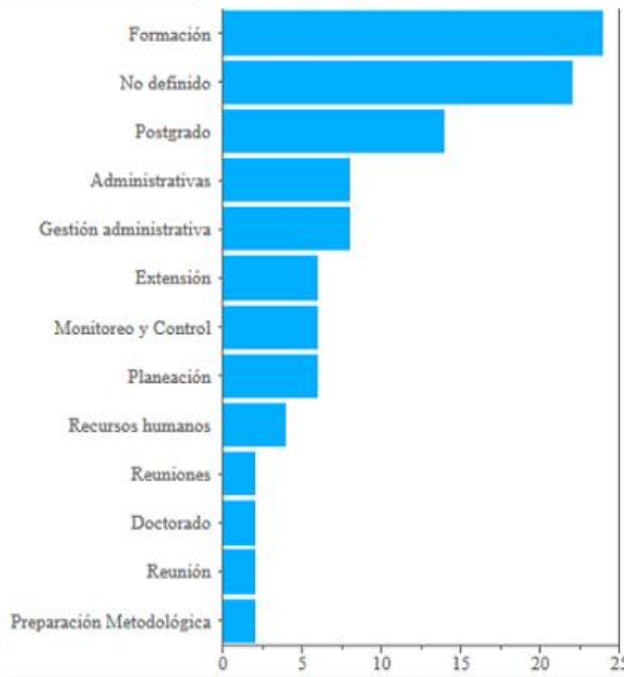
Este tipo de gráfico se actualiza en dependencia del gráfico circular “Estado de cumplimiento”,

1. Si se selecciona en el gráfico circular el área de actividades cumplidas los gráficos de columnas se actualizarían quedando de esta forma:

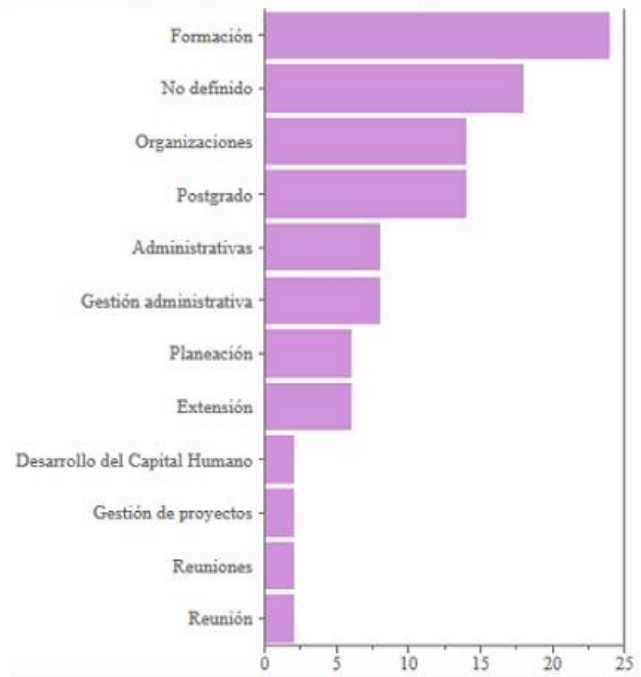


Estado de cumplimiento por:

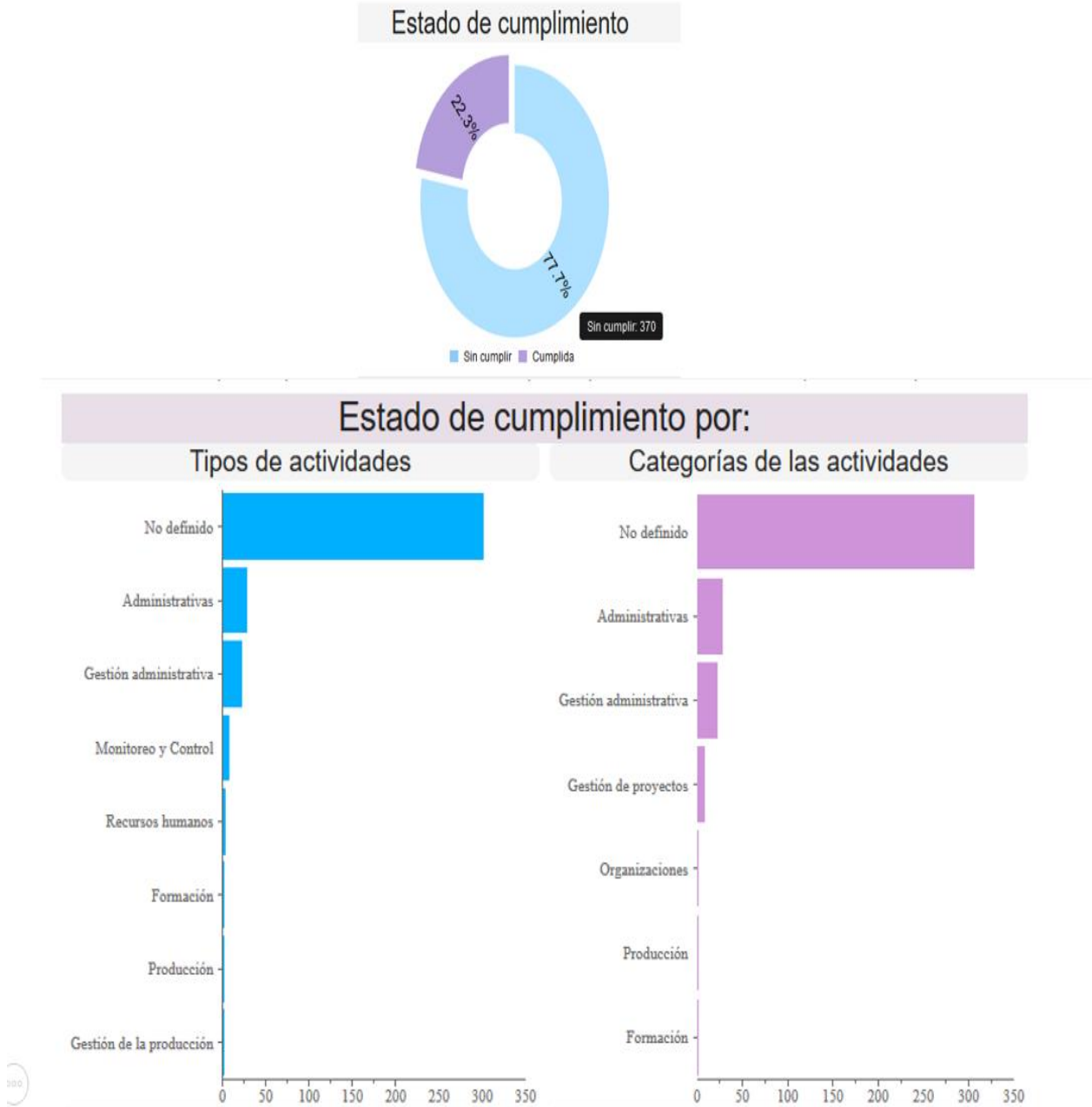
Tipos de actividades



Categorías de las actividades



2. Si se selecciona en el gráfico circular actividades incumplidas los gráficos de columnas se actualizarían quedando de esta forma:



Carta de Aceptación

