

# **Modelo Dimensional para Proyectos y Grupos de Investigación de la VRI de la Universidad del Cauca**

**Wilfran Camilo Loaiza Mera  
William Ortega García**

**TRABAJO DE GRADO PARA OPTAR AL TÍTULO DE INGENIERO DE SISTEMAS**

**DIRECTOR: PhD. MARTHA ELIANA MENDOZA BECERRA**

*Universidad del Cauca*

**Facultad de Ingeniería Electrónica y Telecomunicaciones**

**Departamento de Sistemas**

**Línea de Investigación: Gestión de la Información,**

**Bodega de Datos**

**Popayán, marzo de 2022**

# **Modelo Dimensional para Proyectos y Grupos de Investigación de la VRI de la Universidad del Cauca**



**Wilfran Camilo Loaiza Mera**  
**William Ortega García**

Director: PhD. Martha Eliana Mendoza Becerra

*Universidad del Cauca*

**Facultad de Ingeniería Electrónica y Telecomunicaciones**

**Departamento de Sistemas**

**Línea de Investigación: Gestión de la Información,**

**Bodega de Datos**

Popayán, marzo de 2022

## **TABLA DE CONTENIDO**

<i>Capítulo 1</i> .....	1
1 INTRODUCCIÓN.....	1
1.1 PLANTEAMIENTO DEL PROBLEMA .....	1
1.2 OBJETIVOS .....	2
1.2.1 Objetivo general .....	2
1.2.2 Objetivos específicos.....	2
1.3 PRODUCTOS OBTENIDOS.....	3
1.4 APORTE.....	3
1.5 ESTRUCTURA DEL DOCUMENTO .....	4
<i>Capítulo 2</i> .....	5
2 ESTADO DEL ARTE Y MARCO TEÓRICO .....	5
2.1 ESTADO DEL ARTE .....	5
2.2 MARCO TEÓRICO .....	7
2.2.1 Bodegas de datos.....	7
2.2.2 Modelado dimensional.....	8
2.2.3 Procesamiento analítico en línea .....	10
2.2.4 Procesamiento de transacciones en línea .....	11
2.2.5 Técnicas de diseño.....	11
2.2.6 Proceso de extracción, transformación y carga (ETL).....	13
2.2.7 Visualización de datos .....	14
2.2.8 Norma ISO/IEC 25000.....	14
<i>Capítulo 3</i> .....	16
3 METODOLOGÍA.....	16
3.1 FASE DE INICIACIÓN.....	16
3.2 FASE DE PLANEACIÓN .....	17
3.3 FASE DE ANÁLISIS Y DISEÑO .....	18
3.3.1 Recolección de requerimientos.....	18
3.3.2 Diseño .....	22
3.4 FASE DE DESARROLLO .....	22
3.4.1 Definición de arquitectura .....	22
3.4.2 Desarrollo del Back Room .....	26
3.4.3 Desarrollo del Front Room.....	26

3.4.4	Integración.....	27
3.4.5	Despliegue .....	27
3.5	FASE DE MANTENIMIENTO Y CRECIMIENTO.....	27
3.6	FASE DE GESTIÓN DEL PROYECTO.....	27
<i>Capítulo 4</i> .....		28
4	MODELADO DIMENSIONAL.....	28
4.1	MATRIZ BUS.....	28
4.2	LISTA DE DIMENSIONES.....	28
4.3	MODELOS DIMENSIONALES.....	31
4.3.1	Modelo de Cumplimiento de los proyectos de investigación .....	31
4.3.2	Modelo de Proyectos de investigación.....	35
4.3.3	Modelo de Financiación de proyectos de investigación .....	39
4.4	ESPECIFICACIÓN DE LAS TÉCNICAS DE DISEÑO.....	41
4.4.1	Minidimensión.....	41
4.4.2	Dimensión degenerada.....	42
4.4.3	Dimensión multivaluada y tabla puente.....	42
4.4.4	Juego de roles .....	43
<i>Capítulo 5</i> .....		44
5	PROTOTIPO BODEGAS DE DATOS PARA PROYECTOS DE INVESTIGACIÓN ...	44
5.1	DESARROLLO DE BACK ROOM.....	44
5.1.1	Desarrollo del proceso ETL .....	44
5.1.2	Población y validación de datos.....	50
5.2	DESARROLLO DEL FRONT ROOM .....	50
5.2.1	Identificación y priorización de reportes candidatos .....	50
5.2.2	Diseño de la estrategia de navegación .....	54
5.2.3	Formato y nombrado de los reportes .....	59
5.2.4	Selección de un enfoque de implementación.....	60
5.2.5	Desarrollo de la aplicación de usuario final .....	60
5.3	INTEGRACIÓN.....	61
5.3.1	Procesar el cubo.....	61
5.3.2	Conexión a la DW con las herramientas de reportes .....	64
5.3.3	Pruebas sobre el prototipo.....	64
5.3.4	Presentación del prototipo a los usuarios .....	64
5.3.5	Proceso de aceptación de usuario.....	65
5.4	DESPLIEGUE.....	65

Capítulo 6 .....	66
6 EVALUACIÓN DEL PROTOTIPO .....	66
6.1 OBJETIVO DE LA EVALUACIÓN .....	66
6.2 NORMATIVA PARA EL NIVEL DE SATISFACCIÓN DE USUARIO .....	66
6.3 IMPLEMENTACIÓN DEL TEST DE NIVEL DE SATISFACCIÓN DE USUARIO .....	67
6.4 APLICACIÓN DEL TEST DE NIVEL DE SATISFACCIÓN .....	68
6.5 PROCESO DE EVALUACIÓN .....	68
6.6 RESULTADOS OBTENIDOS .....	69
Capítulo 7 .....	71
7 CONCLUSIONES, RECOMENDACIONES Y TRABAJO FUTURO .....	71
7.1 CONCLUSIONES .....	71
7.2 RECOMENDACIONES .....	72
7.3 TRABAJO FUTURO .....	72

## LISTA DE TABLAS

Tabla 1. Cadena de búsqueda básica en inglés. Fuente propia.....	5
Tabla 2. Cadena de búsqueda básica en español. Fuente propia.....	5
Tabla 3. Comparación de trabajos relacionados.....	7
Tabla 4. Priorización de requerimientos.....	20
Tabla 5. Matriz de comparación de herramientas para motor relacional. Fuente propia...	24
Tabla 6. Matriz de comparación de herramientas OLAP. Fuente propia.....	24
Tabla 7. Matriz de comparación herramientas para ETL. Fuente propia.....	25
Tabla 8. Matriz de comparación herramientas para reportes. Fuente propia.....	25
Tabla 9. Matriz bus. Fuente propia.....	30
Tabla 10. Medidas Cumplimiento proyectos de investigación. Fuente propia.....	32
Tabla 11. Dimensión Proyecto. Fuente propia.....	32
Tabla 12. Dimensión grupo. Fuente propia.....	33
Tabla 13. Dimensión Fecha. Fuente propia.....	33
Tabla 14. Dimensión Persona. Fuente propia.....	34
Tabla 15. Dimensión Estado Compromiso. Fuente propia.....	34
Tabla 16. Dimensión Estado Proyecto. Fuente propia.....	34
Tabla 17. Dimensión Tipo Compromiso. Fuente propia.....	35
Tabla 18. Medidas Proyectos de investigación. Fuente propia.....	36
Tabla 19. Dimensión convocatoria. Fuente propia.....	36
Tabla 20. Dimensión Localización. Fuente propia.....	37
Tabla 21. Dimensión Indicadores Documentos Proyecto. Fuente propia.....	37
<b>Tabla 22. Dimensión Línea de Investigación. Fuente propia.....</b>	<b>38</b>
Tabla 23. Dimensión Palabra Clave. Fuente propia.....	38
Tabla 24. Dimensión Datos Socio Demográficos. Fuente propia.....	38
Tabla 25. Dimensión Rol Integrante Proyecto. Fuente propia.....	39
Tabla 26. Medidas Financiación de proyectos de investigación. Fuente propia.....	40
Tabla 27. Dimensión Rubro. Fuente propia.....	40
Tabla 28. Dimensión Entidad Financiadora. Fuente propia.....	40
Tabla 29. Dimensión Tipo de Egreso. Fuente propia.....	40
Tabla 30. Dimensión Tipo de Financiación. Fuente propia.....	41
Tabla 31. Reportes candidatos. Fuente propia.....	52
Tabla 32. Métricas para la característica de calidad Satisfacción. Tomado de [28].....	67
Tabla 33. Calificación de los cuestionarios según la escala de Likert. Tomado de [29]....	67
Tabla 34. Grado de satisfacción según niveles de puntuación. Tomada de [28]. .....	69
Tabla 35. Cantidad de preguntas evaluadas en los tests. Fuente propia.....	69
Tabla 36. Matriz de calidad. Fuente propia.....	70

## **LISTA DE ILUSTRACIONES**

Ilustración 1. Esquema estrella. Tomado de [14] .....	9
Ilustración 2. Esquema de copo de nieve. Tomada de [14]. .....	10
Ilustración 3. Cubo dimensional con 3 dimensiones. Basado en [16]. .....	11
Ilustración 4. Métricas subcaracterística Utilidad. Tomada de [23] .....	15
Ilustración 5. Métricas subcaracterística Utilidad. Tomada de [24] .....	16
Ilustración 6. Matriz Impacto - Viabilidad .....	17
Ilustración 7. Arquitectura prototipo de DW. Fuente propia. ....	26
Ilustración 8. Esquema Estrella Cumplimiento de proyectos. Fuente propia. ....	31
Ilustración 9. Esquema Estrella de Proyectos. Fuente propia. ....	35
Ilustración 10. Esquema Estrella Financiación proyectos. Fuente propia. ....	39
Ilustración 11. Minidimensión. Fuente propia .....	42
Ilustración 12. Dimensión degenerada. Fuente propia. ....	42
Ilustración 13. Dimensión multivaluada y tabla puente. Fuente propia. ....	43
Ilustración 14. Juego de roles. Fuente propia. ....	43
Ilustración 15. Paso y salto en una transformación. Tomado de [27] .....	45
Ilustración 16. Cargue dimensión Proyecto. Fuente propia. ....	46
Ilustración 17. Cargue dimensión Localización. Fuente propia.....	47
Ilustración 18. Cargue dimensión Fecha. Fuente propia. ....	48
Ilustración 19. Cargue tabla de hechos TH_ProyectoCumplimiento.....	49
Ilustración 20. Cargue tabla puente Palabras Clave .....	50
Ilustración 21. Trabajo para cargue de dimensiones - Modelo Proyecto .....	51
Ilustración 22. Trabajo cargue tabla de hecho y tablas puente - Modelo Proyecto .....	51
Ilustración 23. Carpetas de reportes listadas para el encargado del SI .....	54
Ilustración 24. Reportes sobre Cumplimiento para el encargado del SI .....	54
Ilustración 25. Reportes sobre Financiación para el encargado del SI .....	55
Ilustración 26. Carpetas de reportes listadas para la jefe .....	55
Ilustración 27. Reportes sobre Detalle de Proyectos para la jefe .....	55
Ilustración 28. Reportes sobre Estadística de Proyectos para la jefe .....	55
Ilustración 29. Carpetas de reportes listadas para la secretaria .....	56
Ilustración 30. Reportes sobre Detalle de Proyectos para la secretaria .....	56
Ilustración 31. Reportes sobre Estadística de Proyectos para la secretaria .....	56
Ilustración 32. Reportes sobre Financiación para la secretaria .....	56
Ilustración 33. Área de trabajo de usuarios. Fuente propia .....	57
Ilustración 34. Consultas Encargado SI. Fuente propia.....	57
Ilustración 35. Consultas Secretaria. Fuente propia .....	58
Ilustración 36. Consultas Jefe. Fuente propia .....	58
Ilustración 37. Tabla de consulta Ad-hoc en Power BI .....	59
Ilustración 38. Gráfico de consulta Ad-hoc en Power BI.....	59
Ilustración 39. Formato general reportes estándar. Fuente propia .....	60
Ilustración 40. Área de trabajo consultas Ad-Hoc.....	61
Ilustración 41. Esquema en Schema Workbench.....	62
Ilustración 42. Conexión del esquema con el motor relacional.....	62
Ilustración 43. Procesamiento del cubo en visor analítico La Azada .....	63
Ilustración 44. Conexión base de datos relacional y esquema de datos .....	63

Ilustración 45. Conexión a la DW con POWER BI..... 64  
Ilustración 46. Conexión a la DW con Pentaho Report Designer ..... 65

# Capítulo 1

---

## 1 INTRODUCCIÓN

En este capítulo, inicialmente, se presenta el problema identificado para este trabajo de grado, posteriormente, los objetivos que se especificaron para abordarlo, luego se detallan resultados y aportes logrados con la realización del trabajo. Finalmente, se presenta la organización de este documento.

### 1.1 PLANTEAMIENTO DEL PROBLEMA

La Vicerrectoría de Investigaciones (VRI) es el ente rector responsable de diseñar, promover, ejecutar y controlar las políticas de investigación de la Universidad del Cauca. Dentro de las dependencias de esta Vicerrectoría se encuentra la División de Gestión de la Investigación, la cual tiene como función administrar los proyectos formulados y presentados, las convocatorias, además, definir estrategias de financiamiento y suministrar la documentación relacionada con los procedimientos de investigación [1]. Esta división cuenta con un sistema de información (SIVRI), que permite gestionar los datos relacionados con los grupos de investigación y sus semilleros, los proyectos de investigación, los integrantes, áreas de conocimiento, así como también, las convocatorias internas o externas que se presentan en la Universidad.

Pese a que SIVRI brinda diferentes funcionalidades, no permite la generación de informes gráficos sobre la información, pues al consultar sobre los datos, se obtiene en gran mayoría resultados de forma textual que no permiten una interpretación rápida y eficiente. Por ejemplo, si se quiere conocer la cantidad de proyectos asociados a un grupo de investigación, con sus respectivos estados, el sistema presenta un listado de todos los proyectos asociados al grupo, donde el conteo de estos se debe realizar de forma manual por parte de los funcionarios.

Además, SIVRI carece de una funcionalidad que permita obtener reportes consolidados sobre información relacionada con los grupos, semilleros, convocatorias y proyectos de investigación, en los cuales se puede observar aspectos más importantes sobre estos procesos, tales como la cantidad, el rol y nivel de estudio de los integrantes en un grupo de investigación, las líneas de investigación y/o palabras claves abordadas en la mayoría de proyectos presentados en las convocatorias de los últimos años. Por lo tanto, obtener este tipo consultas y reportes es de gran ayuda para la planeación estratégica de las convocatorias internas que financia la VRI, y los temas más relevantes y que son tendencia en la academia y campos en los cuales se aplican. Por ejemplo, si se cuenta con un reporte que muestre las líneas de investigación con más proyectos asociados en las últimas convocatorias, la VRI puede elaborar un plan que sirva de apoyo para promover el desarrollo de las actividades de investigación, formación y enseñanza, de dichas líneas de investigación, en los diferentes programas académicos de la Universidad [1].

Por esto, es útil contar con herramientas que permitan dar respuesta a este tipo de consultas, para que sean analizadas y estudiadas, y sirvan de apoyo a la toma de decisiones por parte de la Vicerrectoría.

Una alternativa de solución a este problema, se obtiene mediante la implementación de una bodega de datos, la cual es una base de datos que extrae, transforma, organiza y almacena grandes volúmenes de información proveniente de diversas fuentes; al mismo tiempo que permite realizar consultas dinámicas (Ad Hoc) y creación de reportes que apoyan el proceso de toma de decisiones de una organización [2]. Las bodegas de datos incluyen el modelamiento dimensional para permitir el procesamiento, almacenamiento y recuperación de gran cantidad de datos. Kimball plantea técnicas de diseño que permiten modelar situaciones comunes para obtener modelados dimensionales más efectivos [3].

En la literatura se encuentran trabajos relacionados con el modelamiento de procesos de Publicaciones de artículos y de Gestión de datos financieros de una universidad [4][5][6][7]. No obstante, estos modelos dimensionales presentan solo un esquema general, sin demostrar el detalle de sus componentes (dimensiones, medidas) ni de los casos o técnicas de diseño empleadas, lo cual dificulta su entendimiento y apropiación para organizaciones que estén interesadas en modelar e implementar este tipo de procesos.

Teniendo en cuenta lo anterior surge la siguiente pregunta ¿Cuáles técnicas de diseño<sup>1</sup> deben considerarse en el modelado dimensional para convocatorias, grupos, semilleros y proyectos de investigación de la División de Gestión de la Investigación de la VRI de la Universidad del Cauca?

Por lo tanto, en este trabajo de grado se define un modelo dimensional para la División de Gestión de la Investigación, que contempla y aplica las técnicas de diseño identificadas, para modelar los procesos de: convocatorias, grupos, semilleros y proyectos de Investigación. Además, se crea un prototipo de bodega de datos para los proyectos de Investigación, abordando temas financieros y de cumplimiento de compromisos, que permita apoyar el proceso de análisis de información y toma de decisiones, para mejorar así la planificación de procesos y actividades relacionadas a los proyectos de Investigación de la Universidad.

## **1.2 OBJETIVOS**

### **1.2.1 Objetivo general**

Proponer un modelo dimensional para convocatorias, grupos, semilleros y proyectos de investigación de la División de Gestión de la Investigación de la VRI de la Universidad del Cauca, que permita la construcción de un prototipo de bodega de datos para el proceso de proyectos de investigación.

### **1.2.2 Objetivos específicos**

- Identificar las técnicas de diseño que permitan la creación de modelos dimensionales para los procesos de convocatorias, grupos, semilleros y proyectos de investigación.
- Construir un prototipo de una bodega de datos basado en el modelo dimensional propuesto para el proceso de proyectos de investigación, que permita la visualización de los datos por medio de consultas analíticas Ad Hoc y de reportes estándar.
- Evaluar la utilidad de los reportes estándar y las consultas analíticas que genere el prototipo de la bodega de datos, por medio de un test de nivel de satisfacción de los funcionarios encargados del análisis de los proyectos en la VRI, con base en la métrica nivel de satisfacción de la ISO/IEC 25022: Medidas de Calidad en Uso, definida para la sub característica utilidad.

### 1.3 PRODUCTOS OBTENIDOS

A continuación, se presentan los productos obtenidos con la realización de este trabajo de grado:

- **Monografía del trabajo de grado:** este documento detalla el trabajo realizado para alcanzar los objetivos del proyecto. En él se especifican las bases teóricas para el proyecto, la revisión de la literatura, trabajos relacionados, la metodología empleada como base para la creación del prototipo de DW, el modelado dimensional con detalles de tablas de hechos, dimensiones y medidas, la elaboración del prototipo planteado, la evaluación de satisfacción aplicada a los usuarios finales, conclusiones y trabajos futuros.
- **Modelo dimensional:** contempla la matriz bus del proceso de proyectos de investigación, los esquemas estrella para los modelos de información general de proyectos, manejo financiero y cumplimiento de compromisos, describiendo su granularidad, hechos, dimensiones y jerarquías, así como también técnicas de diseño identificadas.
- **Prototipo de DW:** sistema software compuesto por: la bodega de datos relacional de los proyectos de investigación, la cual incluye un estándar de nombrado, diseño físico, script de creación, plan de indexación y de particionamiento; trabajos y transformaciones para el cargue de datos que incluyen mapa origen-destino, detalle del flujo de los elementos de cada uno de los paquete de carga; cubo multidimensional que incluye descripción de la conexión; la herramienta Ad-hoc configurada y el módulo de reportes estándar; También se incluye la matriz de comparación de productos, con base a la investigación hecha para la selección de herramientas necesarias para implementar el prototipo de DW.
- **Artículo:** contiene la descripción del trabajo realizado para alcanzar los objetivos de este proyecto, el documento está realizado en formato APA.

### 1.4 APORTE

Con la realización de este proyecto, se aportan conocimientos acerca de las técnicas de diseño identificadas para realizar el modelado dimensional de convocatorias, grupos, semilleros y proyectos de investigación, incluyendo datos personales socio demográficos. Este trabajo estará puesto a disposición de las demás universidades públicas nacionales, como punto de partida para realizar el diseño de sus propias bodegas de datos enfocadas a la gestión de la investigación. Los trabajos relacionados que se encontraron abordan los temas de publicaciones de producción intelectual y montos financieros, pero no detallan ni presentan las técnicas para apoyar el modelado dimensional. Además, en la literatura no se encuentran trabajos referentes a bodegas de datos que aborden las temáticas de Convocatorias, Semilleros y Grupos.

El personal de la División de Gestión de la Investigación de la Universidad del Cauca tendrá a disposición el prototipo de bodega de datos para la gestión de los proyectos de investigación, de tal modo que se tenga acceso a reportes consolidados y consultas analíticas para apoyar el proceso de toma de decisiones.

## **1.5 ESTRUCTURA DEL DOCUMENTO**

Este trabajo de investigación está dividido en siete capítulos, como se muestra a continuación:

**Capítulo I.** Introducción: se explica el problema encontrado en la División de Gestión de la investigación con referencia a las falencias del sistema que se tiene para el manejo de información; se plantean los objetivos para abordar esta problemática, el primero busca identificar las técnicas de diseño utilizadas en la realización de los modelos dimensionales propuestos, el segundo implementar un prototipo de DW para el modelo de proyectos de investigación y el tercero evaluar la utilidad del prototipo; finalmente se describe artefactos y aporte obtenidos con el desarrollo del proyecto.

**Capítulo II.** Estado del arte y marco teórico: se presenta la revisión sistemática efectuada para encontrar trabajos con modelos dimensionales que abarcan temas de investigación, publicaciones y montos financieros. Además, los conceptos teóricos necesarios para la realización del modelado dimensional y los componentes de bodegas de datos utilizados en este proyecto.

**Capítulo III.** Metodología: describe la metodología denominada *Metodología de desarrollo de bodegas de datos para MiPymes*, empleada en la elaboración de los modelos dimensionales y el prototipo de DW, presentando las fases, subfases, actividades y artefactos que se tomaron como guía para el desarrollo de este proyecto.

**Capítulo IV.** Modelado dimensional: se presenta la matriz bus de los procesos de convocatorias, grupos, semilleros y proyectos de investigación y los modelos dimensionales realizados a partir de esta matriz, los cuales corresponden a: información de convocatorias, grupos, semilleros y proyectos; cumplimiento de compromisos y aspectos financieros de los proyectos. Para estos modelos se incluye una descripción detallada de las dimensiones, hechos y técnicas de diseño identificadas.

**Capítulo V.** Elaboración del prototipo propuesto: se especifican las fases y subfases de desarrollo abordadas para la implementación del prototipo de DW de proyectos de investigación, describiendo algunas de las actividades con sus artefactos.

**Capítulo VI.** Evaluación del prototipo: se exponen los resultados de la evaluación realizada por parte de los usuarios finales al prototipo de DW de proyectos de investigación, donde se mide el grado de satisfacción respecto a los reportes y consultas analíticas propuestos, haciendo uso de la medida *Nivel de satisfacción* que se encuentra en la subcaracterística de *Utilidad* planteada en la norma ISO/IEC 25022; obteniendo un nivel de *Muy satisfactorio* por parte de los usuarios.

**Capítulo VII.** Conclusiones, recomendaciones y trabajos futuros: por último, se describen las conclusiones obtenidas con la realización de este trabajo de investigación, recomendaciones y posibles trabajos futuros.

## Capítulo 2

### 2 ESTADO DEL ARTE Y MARCO TEÓRICO

Esta sección presenta los resultados obtenidos de la revisión de la literatura de los trabajos relacionados y describe los conceptos necesarios para el proceso de construcción de un proyecto de bodega de datos.

#### 2.1 ESTADO DEL ARTE

Con base en los lineamientos planteados por B- Kitchenham y S. Charters [8], se realizó una revisión sistemática en diferentes fuentes de información para investigación científica (Scopus, Science Direct, SpringerLink, IEEE), definiendo dos cadenas de búsqueda una en inglés (ver Tabla 1) y otra en español (ver Tabla 2) con palabras claves referentes al tema de investigación:

**Tabla 1. Cadena de búsqueda básica en inglés. Fuente propia.**

(Data warehouse OR Data warehousing OR Business Intelligence) AND (Multidimensional model OR Multidimensional modeling OR Dimensional model OR Dimensional modeling OR multidimensional design OR dimensional design) AND (Academic OR University OR Higher Education OR Research OR Projects)
---

**Tabla 2. Cadena de búsqueda básica en español. Fuente propia.**

(Bodegas de datos OR Inteligencia de negocios) AND (Modelo dimensional OR Modelo multidimensional OR Diseño dimensional OR Diseño multidimensional) AND (Universidad OR Sistema financiero OR Grupos OR Semilleros OR Proyectos OR Publicación)
--

Además, se definió como criterio de inclusión que fueran artículos publicados desde el año 2012 y como criterios de exclusión artículos que: no eran una bodega de datos, no existía relación entre los modelos y descripción presentada, no presentan un modelado dimensional.

Luego se aplicó la cadena de búsqueda en cada fuente de información, obteniendo como resultado dieciocho trabajos relacionados a los procesos de negocio, a los cuales se les aplicaron los criterios de inclusión y exclusión; quedando cuatro trabajos relacionados definidos como relevantes para esta investigación, los cuales se describen a continuación:

En Indonesia, en el año 2014 se realizó una investigación [4] que integró minería de datos, para modelar una bodega de datos que permite manipular información sobre el registro de las investigaciones. El modelo presenta un esquema de tipo estrella, con una tabla de hechos denominada *Investigacion\_Hecho*, la cual gestiona la información sobre las investigaciones realizadas, por medio de las dimensiones: *tipo*, *autor*, *campo académico*, *director* y *fecha de realización*; y como medida la cantidad de investigaciones. Las dimensiones y las medidas no son explicadas en la investigación, sólo se muestran gráficos y explicaciones cortas sobre la implementación del modelo propuesto, pero no se tiene una evaluación de éste.

En Italia, en el año 2015 se desarrolló una bodega de datos [5] que integra fuentes de datos actuales e históricas de un ambiente académico para almacenar información sobre el Estado de la investigación, que tiene como objetivo evaluar la producción científica en los departamentos de la universidad. Se presenta el esquema de la tabla de hechos que se llama *Publicación*, la cual no tiene medidas. Tanto la tabla de hechos como las dimensiones son explicadas, donde la dimensión *autor* tiene como subdimensiones *facultad*, *departamento* y *rol institucional*. La bodega de datos fue implementada en la web, pero no se tiene una evaluación de la aplicación por parte de los usuarios.

En Letonia, en el año 2018 se crea un artículo [6] en el cual se propone un sistema de información de las investigaciones de la Universidad de Lituania. Los objetivos principales son asegurar la integración de los datos del sistema interno de la Universidad con fuentes externas y garantizar la calidad de éstos. En este artículo se presenta un esquema de tipo estrella donde se tiene una tabla de hechos llamada *Publicación*, donde se identifican tres medidas denominadas: páginas, scopus cantidad de visitas, WoS cantidad de visitas, de las cuales no se tiene explicación; diez dimensiones relacionadas, las cuales son: *autor SCOPUS*, *autor*, *persona LU*, *afiliación SCOPUS*, *departamento LU*, *categoría FOS*, *asunto SCOPUS*, *título de serie*, *rango de revista*, *métricas de citas de revistas*, que hacen referencia a los autores e información de la publicación, de lo cual no se tiene ninguna explicación. El artículo no presenta la implementación del sistema propuesto, por lo cual tampoco se ofrece una evaluación de este.

En Filipinas, en el año 2018 se realizó una bodega de datos financieros y su herramienta de visualización [7] para una institución estatal de educación superior, la cual permite almacenar y agregar datos financieros. Este trabajo presenta un esquema estrella con una tabla de hechos llamada *monto financiero* con las dimensiones *fondo*, *unidad financiera*, *cuenta* y *tiempo*. Las medidas usadas son: asignación inicial, asignaciones adicionales, fondos transferidos a otros centros de costos, fondos transferidos de otros centros de costos, asignación total, obligaciones pagadas, obligaciones impagas, obligaciones totales y saldo. Las dimensiones presentan una breve descripción, sin embargo, los atributos y medidas no son explicados. El sistema fue implementado y se sometió a una prueba de rendimiento de consulta, para evaluar la diferencia entre el sistema corporativo usado actualmente y la bodega de datos. También se realizó una prueba de usabilidad aplicada a cinco (5) usuarios, para evaluar la calidad de visualización de los datos del sistema.

En la Tabla 3 se presenta una comparación entre los trabajos relacionados, mostrando los componentes más importantes del modelo dimensional de cada uno de los artículos. El campo *Ref*, representa la referencia bibliográfica del trabajo citado y el año de publicación; el campo *Tabla de Hechos*, presenta la tabla de hechos que se presenta en el modelo dimensional del trabajo; en el campo *Exp. Dimens*, se menciona cuántas dimensiones son explicadas (EXP), si no se cuenta con explicación (NO EXP), cuántas son entendidas (EN) y el total (T) que se presentan; en el campo *Exp. Medida*, se indica para las medidas del mismo modo que para las dimensiones; en el campo *Datos Socio Dem*, se indica si el modelo presenta datos sociodemográficos y de qué tipo son; el campo *Imp*, indica si en el trabajo se realizó la implementación de la bodega de datos; por último, en el campo *Evaluado*, se indica si la bodega de datos propuesta en el trabajo fue evaluada.

**Tabla 3. Comparación de trabajos relacionados.**

Ref	Tabla de hechos	Exp. Dimens.	Exp. Medida	Datos. Socio Dem	Imp.	Evaluated
[4] 2014	Investigación	NO EXP (12 EN de 18 T)	NO EXP (1 EN de 1 T)	NO	SI	NO
[5] 2015	Publicación	10 EXP (17 EN de 20 T)	No tiene medidas	SI (Ubicación, Fecha de Nacimiento)	SI	NO
[6] 2018	Publicación	NO EXP (6 EN de 10 T)	NO EXP (1 EN de 3 T)	SI (Ubicación)	NO	NO
[7] 2018	Monto financiero	4 EXP (4 EN de 4 T)	NO EXP (9 EN de 9 T)	NO	SI	SI

Al analizar los resultados de la Tabla 3, se observa que los artículos [4], [5] y [6] tienen relación con el tema de publicaciones realizadas en una organización, pero en general no presentan una explicación de las dimensiones y medidas, además, los datos socio demográficos son mínimos. Por otra parte, el artículo [7] presenta una tabla de hechos relacionada con los montos financieros, pero tampoco explican todas las dimensiones y medidas ni presentan datos socio demográficos.

Aunque los artículos encontrados sirven como referencia para este proyecto, al tratar los temas de *Publicaciones* (con información del autor y su rol institucional, fecha de publicación, categoría y editores) y *Montos financieros* (con información de la gestión de sus transacciones financieras generales), que son similares a los procesos que se llevan a cabo en la VRI, éstos no ofrecen información detallada de todos sus componentes, principalmente dimensiones y medidas; además, no describen técnicas de modelado que se presentan en el modelo.

Por otra parte, sobre los temas de *Convocatorias*, *Semilleros* y *Grupos* de investigación, no se encontraron artículos relacionados.

## 2.2 MARCO TEÓRICO

En esta sección se presentan los conceptos para el diseño, implementación y evaluación de un proyecto de bodegas de datos, revisando conceptos como: definición, modelado dimensional, procesamiento analítico en línea, procesamiento de transacciones en línea, técnicas de diseño, proceso de extracción, transformación y carga, visualización de datos y la norma ISO/IEC 25000.

### 2.2.1 Bodegas de datos

Para Inmon [9], una bodega de datos (DW) es “una colección de datos temáticos, integrados, no volátiles y variables en el tiempo, lo cual va a permitir el apoyo a la toma de decisiones de una organización”. Esta definición involucra cuatro características que son [9]:

- **Temáticas:** los datos dentro de una DW se organizan por temas, de esta manera los usuarios finales podrán acceder y comprender más fácilmente la información almacenada, ya que ésta se encontrará categorizada.

- **Integradas:** la DW debe unir de manera consistente la información que se obtiene de diferentes fuentes, permitiendo resolver las inconsistencias que se presentan con respecto a los datos y nombres de campos, unidades y medidas, antes de que sean almacenados.
- **No volátiles:** dentro de sistemas operacionales convencionales se permiten operaciones para consultar, insertar, eliminar y modificar la información almacenada, pero en una DW se permite principalmente consultar e insertar datos. Esto con el fin de que los datos se almacenen de forma permanente, y así poder realizar análisis que permitan el apoyo a la toma de decisiones.
- **Variantes en el tiempo:** las DW recogen la información de todos los cambios que se realizan sobre los datos en el tiempo. Esto permite que una organización pueda analizar datos actuales e históricos sobre el negocio, estudiar tendencias, tomar decisiones y definir nuevas estrategias.

Dentro de las DW se pueden encontrar diferentes divisiones, las cuales se denominan Datos departamentales (Data mart). Estos están orientados al análisis, almacenamiento e integración de datos dentro de un área específica, lo cual implica que poseen la misma funcionalidad y complejidad que las DW, aunque cuentan con necesidades distintas dependiendo del tema de negocio que trabaja [10].

### **2.2.2 Modelado dimensional**

Es una técnica que brinda una estructura que permite resolver consultas analíticas presentando los datos de forma sencilla e intuitiva. Esto hace referencia a que se tienen modelos sencillos que aseguran buenos tiempos de respuesta ofreciendo a los usuarios accesos más rápidos a la información [11].

Para este modelado, se dividen los procesos de negocio entre medidas y dimensiones. Las medidas en su mayoría son de tipo numérico y se conocen como Hechos, para los cuales se tiene un contexto que va a describir cómo se registraron y en qué momento. Pese a que el contexto se ve como un todo, se tienen registros que tienen diferentes características que describen un mismo hecho [12]. Por ejemplo, si el hecho a tratar es un proyecto de investigación, se puede describir el contexto en una serie de divisiones que rodean al hecho, el autor del proyecto, su presupuesto asignado, su fecha de publicación, entre otros. Estas divisiones son conocidas como dimensiones, las cuales son textos descriptivos a diferencia de los hechos que son numéricos.

Kimball recomienda que los modelados dimensionales presenten entre cinco y quince dimensiones. En el primer caso, cuando se tienen menos de cinco, se reduce el poder de consulta y análisis sobre el modelo, por esto se sugiere revisar si algunas dimensiones se han quedado por fuera del modelado. En el segundo caso, más de quince dimensiones, cuando el usuario realiza las consultas se puede presentar mayor complejidad debido a la cantidad de dimensiones, por esto es importante estudiar la posibilidad de combinar varias dimensiones en una sola dimensión [12].

Los principales componentes del modelado dimensional son las tablas de hechos y las tablas de dimensiones, que se definen de la siguiente manera [10]:

- **Tablas de hechos:** es la tabla principal de un modelo dimensional, contiene campos claves que unen a las tablas de dimensión, así como también métricas o medidas para aquello que se va a medir y/o analizar.
- **Tablas de dimensión:** son tablas simples que se unen a las tablas de hechos a través de un atributo específico. Los atributos de estas tablas describen información característica de las tablas de hechos.

En la metodología de Kimball se encuentran diferentes técnicas de diseño para la realización del modelado dimensional, las cuales son importantes porque permiten realizar una clasificación de las dimensiones, dependiendo de su contenido, su estructura y los tipos de datos que almacena. Al utilizar estas técnicas, los tipos de dimensiones identificadas de uso más frecuente son: dimensiones degeneradas, dimensión de múltiples jerarquías, dimensiones de juego de roles, dimensiones basura y mini dimensiones [10].

## ESQUEMAS DE REPRESENTACIÓN DIMENSIONAL

Todo proceso de una organización está representado por un modelo dimensional donde se muestran los hechos y las dimensiones que lo conforman, entre estos esquemas de representación podemos encontrar:

- **Esquema en estrella** [13]: refleja la manera en cómo los usuarios, en términos de métricas de negocio, piensan y necesitan los datos para consulta y análisis. El esquema estrella se compone de una tabla de hechos central y una serie de dimensiones. La tabla de hechos consta de métricas y las dimensiones contienen los atributos que los usuarios normalmente consultan y analizan. En la Ilustración 1, se muestra la estructura general de un esquema en estrella.

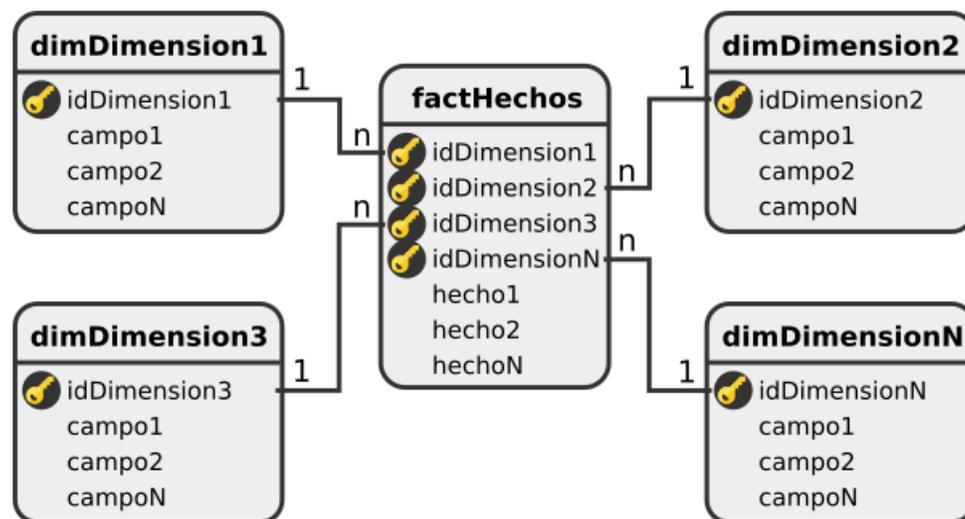


Ilustración 1. Esquema estrella. Tomado de [14]

- **Esquema de copo de nieve:** al igual que un esquema en estrella, el esquema de copo de nieve tiene una tabla de hechos. Sin embargo, los esquemas copo de nieve

contienen diversas tablas de dimensión, donde cada una representa un nivel en una jerarquía. En este esquema se normaliza cada dimensión en un esquema estrella, donde la estructura resultante se parece a un copo de nieve con su tabla de hechos en el centro [15]. En la Ilustración 2, se muestra la estructura general de un esquema de copo de nieve.

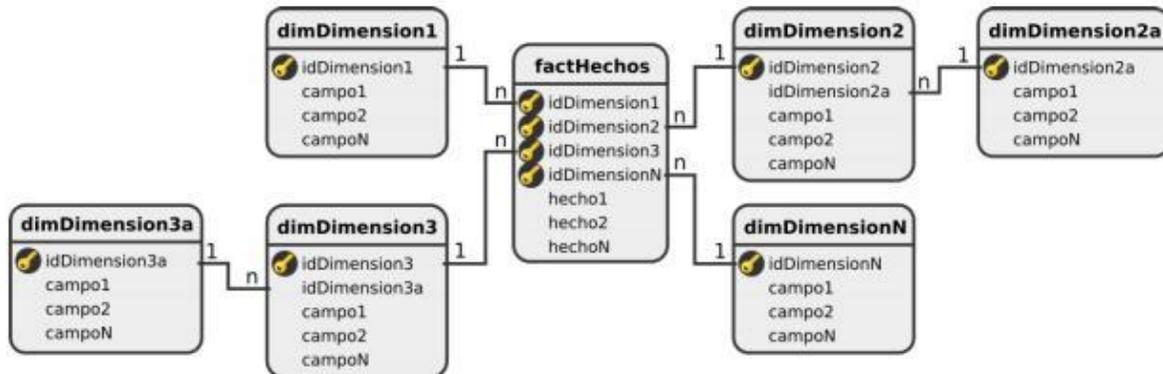


Ilustración 2. Esquema de copo de nieve. Tomada de [14].

### 2.2.3 Procesamiento analítico en línea

El procesamiento analítico en línea (OLAP) presenta datos dimensionales de forma intuitiva, permitiendo a usuarios analíticos segmentar datos, brindando un rendimiento óptimo en consultas y análisis más interesantes desde una amplia gama de herramientas de consulta [3].

En la Ilustración 3, se muestra esquemáticamente un cubo OLAP, cuyos lados comprenden las dimensiones *Producto*, *Canal de Ventas* y *Periodo*, donde su tamaño se ve determinado por el número de celdas.

La información alojada en una DW puede ser explotada por el OLAP, el cual es un enfoque para responder de forma rápida a consultas analíticas multidimensionales, de acuerdo con tres tipos de implementación [16]:

- **ROLAP** (relational online analytical processing): los datos agregados y detallados del cubo se almacenan en un motor de bases de datos relacionales (RDBMS).
- **MOLAP** (multidimensional online analytical processing): los datos agregados y detallados se almacenan en el cubo multidimensional.
- **HOLAP** (hybrid online analytical processing): los datos agregados se almacenan en el cubo y los datos detallados en un RDBMS.

Las consultas ad hoc son los pilares del acceso a la información en una DW, estas admiten la exploración directa e interactiva de los datos, para generar y devolver un resultado como un conjunto de filas y columnas [3].

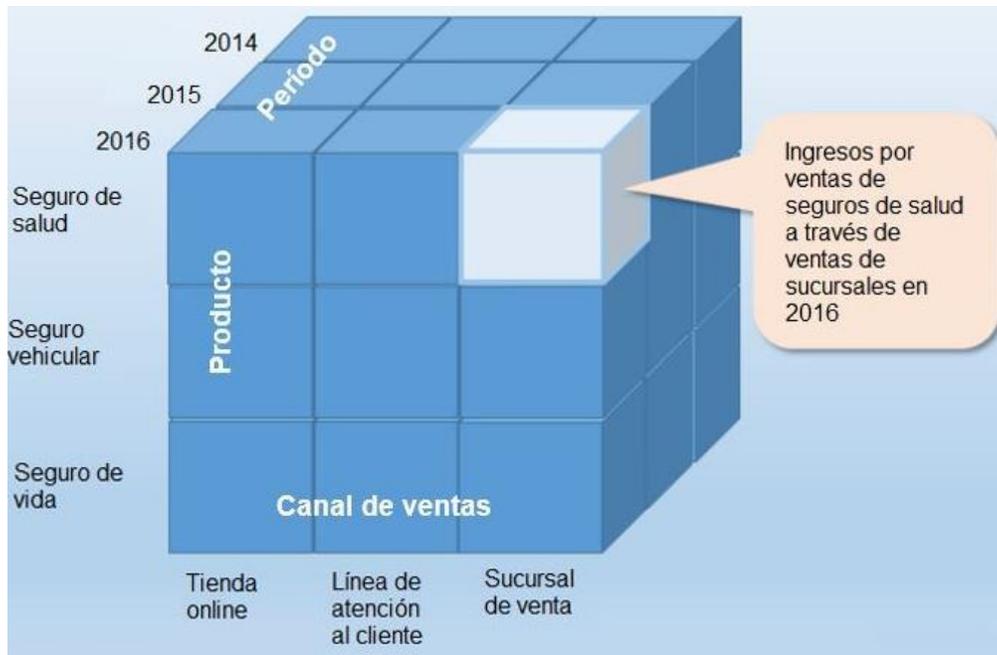


Ilustración 3. Cubo dimensional con 3 dimensiones. Basado en [16].

#### 2.2.4 Procesamiento de transacciones en línea

El procesamiento de transacciones en línea (OLTP) es un sistema que permite manipular los datos relacionados con las operaciones diarias de las organizaciones, principalmente para insertar, consultar, actualizar y eliminar información de una base de datos [17]. Se utilizan en muchos campos de aplicación, como: ventas, salud, reservas online, etc. Las características principales de estos sistemas son [18]:

- Poco tiempo de respuesta al momento de realizar operaciones.
- Alta disponibilidad.
- Gran número de usuarios pueden acceder al mismo tiempo a los datos.
- Alta concurrencia.
- Puede contener millones de datos, dependiendo del tipo de aplicación utilizada.
- Generalmente leen y manipulan pequeñas cantidades de datos.
- Realiza actualización de datos, en segundo plano, por lo cual los usuarios pueden seguir desarrollando y ejecutando tareas normalmente.

#### 2.2.5 Técnicas de diseño

Las técnicas de diseño son una guía específica para crear modelos de datos de acuerdo a un enfoque dimensional, dependiendo del contexto del proceso de negocio [19]. A continuación, se detallan dichas técnicas [10]:

- **Subdimensión:** es una dimensión incorporada o referenciada en otra dimensión. Esta técnica de diseño permite separar un subconjunto de atributos de una dimensión base en una nueva dimensión, con el fin de dividir dimensiones grandes en estructuras más manejables y de esta forma limitar espacio de almacenamiento. Un ejemplo es la dimensión persona, la cual puede contener una referencia de la dimensión demografía, esto permite almacenar los datos demográficos de las personas en una dimensión

diferente, lo cual es útil en términos de almacenamiento si la dimensión persona es grande.

- **Dimensiones Degeneradas:** es una dimensión que únicamente contiene su clave principal. Esta se agrega a la tabla de hechos como un atributo, con la connotación DD, debido a que no está relacionado a ninguna dimensión, pues no cuenta con más datos que puedan relacionarse.
- **Dimensión fecha:** la dimensión fecha contiene atributos que permiten analizar la información a través de períodos de tiempo o intervalos de fecha, en ella se tiene atributos como el número de semana, nombre del mes, período fiscal o el día individual. Esta dimensión se asocia generalmente a todas las tablas de hechos para permitir la navegación a través de jerarquías de fecha. La clave principal no suele ser un valor entero incremental, sino un número entero que representa AAAAMMDD (Año, mes y día).
- **Juego de Roles:** se presenta cuando una dimensión tiene más de una referencia en la tabla de hechos, cada referencia determina un rol diferente para la dimensión. Cada rol representa una clave externa que referencia a una vista separada de la dimensión.
- **Dimensión Basura:** es una agrupación de indicadores tipo bandera o atributos de texto de baja cardinalidad, por ejemplo: verdadero/falso, que no pueden ser organizados de coherentemente en una dimensión identificada previamente en la fase de diseño. La dimensión puede ser generada a partir de todas las posibles combinaciones de atributos o a partir de las combinaciones existentes en los datos, además posee su propia clave sustituta, la cual tendrá una referencia en la tabla de hechos.
- **Dimensión conformada:** es una dimensión única que se puede asociar a varias tablas de hechos en una DW, manteniendo sus atributos y sus registros para cada tabla de hechos.
- **Dimensión multivaluada y tabla puente:** generalmente una dimensión tiene relación de uno a muchos con la tabla de hechos, pero en algunos casos esta relación es de muchos a muchos, debido a que un registro de la tabla de hechos se relaciona a varios miembros de una dimensión. Para estos casos, la dimensión multivaluada se relaciona con la tabla de hechos a través de una tabla puente, la cual contiene la clave de la dimensión multivaluada y la clave compuesta de la tabla de hechos.
- **Dimensiones que cambian lentamente:** son dimensiones para las cuales los valores de sus atributos no cambian de manera frecuente. Por lo tanto, se han definido una serie de técnicas que permiten realizar un seguimiento histórico a los datos, de acuerdo con las solicitudes o requerimientos del proceso de negocio. Algunos tipos que presentan Kimball son:
  - **Tipo 0 (Conservar original):** los valores de los atributos de la dimensión nunca cambian, por esto, siempre se mantienen los valores originales.
  - **Tipo 1 (Sobrescribir):** cuando se produce un cambio en un atributo, se sobrescribe por el nuevo valor y se destruye el seguimiento histórico para ese atributo.
  - **Tipo 2 (Agregar un nuevo registro):** cuando se presenta un cambio en el valor de un atributo, se genera un nuevo registro en la dimensión y se le asigna una clave sustituta (clave primaria), la cual será usada como llave foránea en la tabla de

hechos, hasta el momento en que ocurra un nuevo cambio para el atributo. Al implementar atributos tipo 2, es necesario agregar tres nuevos atributos que permitan realizar análisis de la información actual e histórica, estos son: Fecha de inicio de la vigencia de los atributos cambiantes; Fecha de finalización de la vigencia; y un Indicador de fila actual para referenciar el registro que contiene la información actual.

- **Tipo 3 (Agregar un nuevo atributo):** agrega una nueva columna en la dimensión cuando ocurre un cambio en un atributo, con el objetivo de conservar su valor antiguo. Este tipo de cambio se usa con poca frecuencia.
- **Tipo 4 (Minidimensión):** se utiliza para separar los atributos que cambian con frecuencia dentro de una dimensión y ubicarlos en una nueva dimensión, la cual contiene una clave primaria que se asocia a la tabla de hechos, y así no sobrecargar la dimensión base con mucha información.
- **Tipo 5 (Minidimensión y subdimensión tipo 1):** permite almacenar los valores históricos y actuales de los atributos, en una minidimensión (tipo 4) y una subdimensión tipo 1 (Sobrescribe) con base a la minidimensión, respectivamente, es decir, que ambas son una misma tabla. En el cargue de datos, se debe sobrescribir la referencia a la minidimensión tipo 1, siempre que se modifique la asignación actual.
- **Tipo 6 (Adicionar atributos tipo 1 a una dimensión tipo 2):** al igual que tipo 5, se mantienen los valores históricos y actuales de los atributos, esta técnica incorpora atributos tipo 1 con base a los atributos históricos tipo 2 presentes en una dimensión. Por ejemplo, si una dimensión cuenta con un atributo estado tipo 2, se incluirá un atributo estado tipo 1, esto permite capturar los cambios, además de rastrear la asignación actual. En este caso, el atributo tipo 1 se sobrescribirá en todos los registros asociados cuando se presente una actualización del atributo tipo 2.
- **Tipo 7 (Dimensión tipo 1 y tipo 2):** permite acceder a la tabla de hechos por medio de una dimensión modelada como tipo 1 que muestra los valores actuales, o como una dimensión tipo 2 que muestra los valores históricos, es decir, la misma tabla permite ambas presentaciones, las cuales se despliegan como vistas separadas en las aplicaciones de inteligencia de negocios.

## **2.2.6 Proceso de extracción, transformación y carga (ETL)**

Es el proceso mediante el cual se recopilan datos de diversas fuentes, se transforman los datos de acuerdo a reglas establecidas y son cargados a un sistema destino [20]. El proceso ETL consta de tres pasos ordenados, y deben abordarse de manera individual y secuencial [21] y son los siguientes:

- **Extraer:** los datos fuente deben extraerse de diferentes tipos de origen que incluyen bases de datos, archivos de texto, etc.
- **Transformación:** es la etapa donde se realiza el trabajo más decisivo, pues deben aplicarse diversas operaciones sobre los datos extraídos desde los sistemas fuente, para transformarlos en información adecuada y útil para la bodega de datos. Estas

operaciones incluyen el filtrado, ordenamiento, clasificación, agrupación, unión, limpieza, duplicación y validación de datos [20].

- **Carga:** consiste en cargar la información extraída y transformada a la bodega de datos, donde los datos pueden ser consultados.

### **2.2.7 Visualización de datos**

Los usuarios pueden consultar y analizar la información de la DW, utilizando:

- **Reportes estándar:** Se refiere a informes ya definidos que presentan la información en estadísticas, gráficos o contenidos textual, por ejemplo, en tablas[22].
- **Consultas analíticas Ad-Hoc:** Una herramienta que va a permitir al usuario crear las consultas usando directamente las tablas relacionales y las combinaciones posibles entre ellas[10].

### **2.2.8 Norma ISO/IEC 25000.**

Esta norma brinda una guía para el uso de una serie de estándares internacionales denominados *Requisitos y Evaluación de Calidad de Productos Software* (SQuaRE). Su objetivo es establecer una guía para el desarrollo de productos software mediante la especificación de requisitos y la evaluación de características de calidad. La norma ISO/IEC 25000 se compone de cinco divisiones: 2500n, 2501n, 2502n, 2503n y 2504n (n representa otro dígito de la norma), y estas a su vez se componen de una serie de subdivisiones o estándares [23].

En este trabajo de grado se utilizan dos divisiones de esta norma: la ISO/IEC 2501n y la ISO/IEC 2502n.

- La primera división contiene el estándar ISO/IEC 25010, Modelo de Calidad Genérico, que especifica características de calidad de un producto y las clasifica en dos modelos de calidad, en este caso, se selecciona el modelo para la calidad en uso, el cual tiene 6 características, de las cuales se toma la característica *satisfacción*, y a partir de ésta, la subcaracterística *utilidad* con el fin de evaluar los reportes estándar y consultas analíticas Ad-hoc generados en el prototipo de bodegas de datos. A continuación, se define esta característica y subcaracterística:
  - **Satisfacción:** capacidad del sistema para satisfacer las necesidades mínimas de los usuarios al utilizarlo.
  - **Utilidad:** subcaracterística de Satisfacción y se refiere al grado de satisfacción en que un usuario logra los objetivos planteados.
- La segunda división contiene el estándar ISO/IEC 25022, Medidas de Calidad en Uso. Este estándar proporciona una serie de métricas que permiten evaluar el cumplimiento de las características de calidad en uso de la norma ISO/IEC 25010 en un producto o sistema software. En este trabajo, se selecciona la métrica *nivel de satisfacción* (Ilustración 4), la cual permite evaluar la característica seleccionada.

<b>MÉTRICAS DE CALIDAD EN USO</b>		
<b>Características</b>	<b>Subcaracterísticas</b>	<b>Métricas</b>
<b>Satisfacción</b>	Utilidad	<ul style="list-style-type: none"><li>• Nivel de satisfacción.</li><li>• Uso discrecional de las funciones.</li><li>• Porcentaje de quejas de los clientes.</li></ul>

**Ilustración 4. Métricas subcaracterística Utilidad. Tomada de [23]**

## Capítulo 3

### 3 METODOLOGÍA

Para llevar a cabo el desarrollo de este proyecto se utiliza la metodología de desarrollo de bodegas de datos para MiPymes (MBD)[24], la cual está basada en la metodología de Ralph Kimball y una metodología ágil de bodegas de datos. MBD presenta un enfoque iterativo e incremental, participación por parte del usuario, y detalle de sus fases y actividades. MBD es una metodología más reducida para proyectos, pues establece grupos pequeños de trabajo con pocos roles y responsabilidades concretas, además, presenta menor creación de artefactos.

MBD está diseñada entre seis fases (Ver Ilustración 5): Iniciación, Planeación, Análisis y Diseño, Desarrollo, Mantenimiento y Crecimiento, y Gestión del Proyecto. Estas fases se ejecutan para cada proceso de negocio de una organización. Cabe resaltar que la fase de Desarrollo se compone de cinco subfases, de las cuales tres se realizan de forma iterativa para cada proceso de negocio; y la fase de Gestión del proyecto tiene cabida en todo su ciclo de vida. Para el desarrollo de este proyecto se realiza un ciclo de vida de esta metodología.

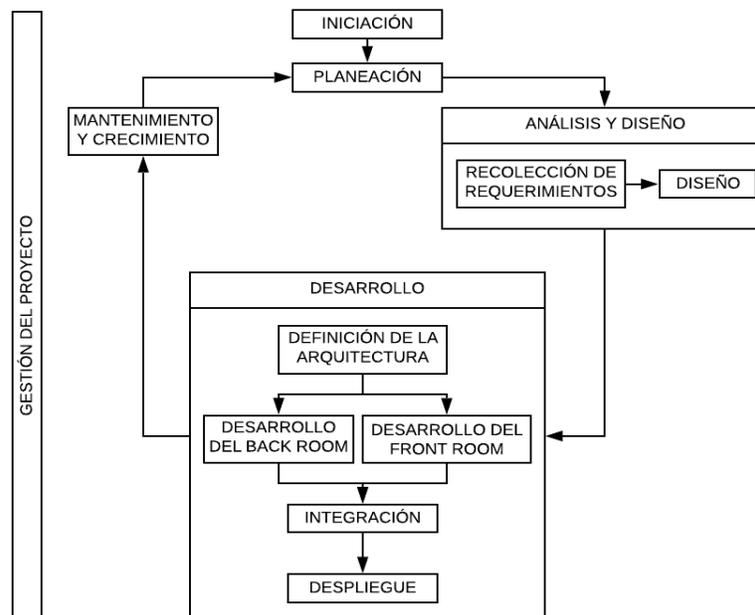


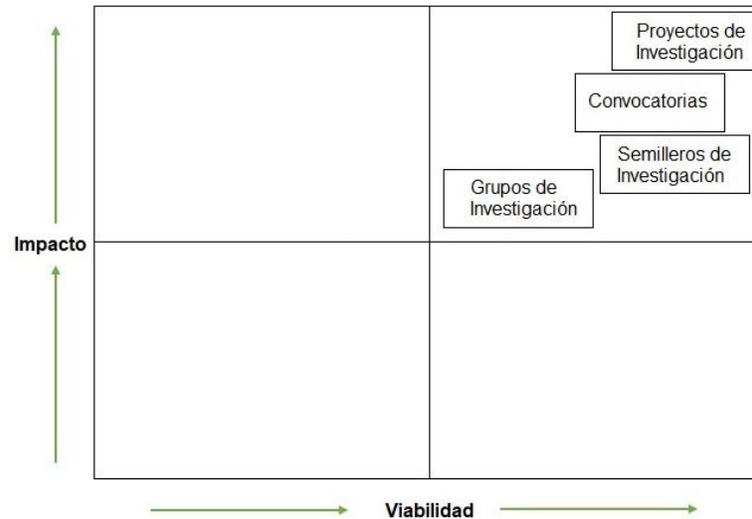
Ilustración 5. Métricas subcaracterística Utilidad. Tomada de [24]

A continuación, se presenta una descripción de cada una de las fases del ciclo de vida de la metodología MBD, teniendo en cuenta que no se describirá la fase de Mantenimiento y Crecimiento, debido a que está fuera del alcance del proyecto.

#### 3.1 FASE DE INICIACIÓN

En esta fase se identifican y priorizan los procesos de mayor impacto y viabilidad para la VRI, para esto se realizaron las siguientes actividades:

- **Identificación de procesos de negocio:** se efectuaron entrevistas a los usuarios para identificar los procesos de negocio de la División de Gestión de la investigación, junto con el impacto y viabilidad de estos.
- **Priorización de procesos de negocio:** se realizó un análisis del impacto de los procesos identificados para la División y la viabilidad de los datos, y se definió la prioridad de estos. En la Ilustración 6, se presentan los procesos de negocio identificados y priorizados, de acuerdo con las entrevistas realizadas.



**Ilustración 6. Matriz Impacto - Viabilidad**

El proceso seleccionado para realizar el prototipo de DW fue *Proyectos de investigación*, que incluye información general, cumplimiento de compromisos e información financiera; teniendo en cuenta que este proceso tiene mayor impacto para la División de Gestión de la investigación y mayor viabilidad para el acceso a las fuentes de datos e información. Sin embargo, en el transcurso de esta fase se identifican otros procesos de negocio, los cuales son analizados en el Anexo A.

### **3.2 FASE DE PLANEACIÓN**

En esta fase se define el alcance, justificación, roles y responsabilidades, y plan del proyecto. Del mismo modo que en la fase de Iniciación, las actividades de esta fase se realizaron durante la elaboración del anteproyecto:

- **Establecer la identidad del proyecto:** se definió el nombre del proyecto de investigación, el cual corresponde al título de este documento.
- **Desarrollar el alcance del proyecto:** se estableció lo que se quiere lograr al finalizar este proyecto, lo cual especifica principalmente cuatro modelos dimensionales, un prototipo de DW para el modelo de proyectos de investigación; y los resultados de la aplicación de una prueba de aceptación de reportes y consultas a los usuarios finales.

- **Desarrollar la justificación del proyecto:** se realizó un estudio costo-beneficio del proyecto para tener claridad acerca de los beneficios que se obtienen con la creación del prototipo de DW.
- **Asignación de roles y responsabilidades:** se asignó a cada integrante de este proyecto un rol y actividades a realizar, de acuerdo con sus habilidades.
- **Elaboración del plan del proyecto:** se especificaron las fases y actividades realizadas para desarrollar este proyecto, teniendo en cuenta fechas de iniciación, duración estimada. Este plan quedó incluido en el anteproyecto aprobado.
- **Reunión de iniciación del proyecto:** se realizó a cabo con el equipo de desarrollo y la directora del trabajo de grado. Además, se notificó a los usuarios.

Los artefactos de esta fase se encuentran en el Anexo B.

### **3.3 FASE DE ANÁLISIS Y DISEÑO**

En esta fase se realiza la identificación de los requerimientos analíticos de los procesos de proyectos, grupos, semilleros de investigación y convocatorias, seguidamente se diseñan los modelos dimensionales, también se especifica la forma en que se carga la información a la DW. Esta fase se compone por dos subfases: Recolección de requerimientos y Diseño.

#### **3.3.1 Recolección de requerimientos**

En esta subfase se identifican las necesidades analíticas de cada proceso de negocio, donde es muy importante contar con la participación de los funcionarios de la División de Gestión de la Investigación que conocen y ejercen actividades laborales dentro de estos procesos. De este modo, se realizaron entrevistas con el fin de obtener una mayor comprensión de los requerimientos y su correcta priorización. En esta subfase se realizaron las siguientes actividades:

- **Preparación de las entrevistas:** se definieron los roles de entrevistador y digitador para llevar a cabo las entrevistas, los cuales se intercambian entre los integrantes del equipo de desarrollo para cada una de las entrevistas. Luego, se identificaron los funcionarios a entrevistar, que hacen parte de los diferentes procesos de la División de Gestión de la Investigación. En esta actividad se crean los artefactos para realizar las entrevistas (ver Anexo C): *Cuestionario para el ejecutivo del negocio*, *Cuestionario para el analista del negocio* y *Cuestionario para encargado del sistema de información*.
- **Realizar las entrevistas:** antes de esta actividad se acordó con los funcionarios la disponibilidad para establecer la programación de las entrevistas. Para la asignación de los cuestionarios se tiene en cuenta el rol desempeñado por cada funcionario dentro de la División de Gestión de la Investigación, los cuales son: Encargado del sistema de información (SI), donde se aplica el cuestionario *Entrevista Personal SI*, el cual brinda información sobre la estructura del sistema y software utilizado, posteriormente el encargado deja vacante su puesto; Jefe de la División, a quien se le aplicaron en conjunto los cuestionarios *Entrevista Ejecutivo-Analista del negocio*, pues es la persona que conoce y dirige los aspectos de proyectos de investigación. Las entrevistas se

llevan a cabo de manera virtual y fuera de la programación establecida debido a los paros universitarios.

- **Documentación de los requerimientos:** las entrevistas realizadas se documentaron y se identificaron los requerimientos expresados por los entrevistados.
- **Priorización de requerimientos:** se realizó una priorización de los requerimientos identificados, para lo cual fue necesario programar una reunión con cada uno de los usuarios, quienes asignaron el nivel de prioridad a cada requerimiento, obteniendo un listado los requerimientos con priorizados (ver Tabla 4).
- **Revisión y aceptación del usuario:** por medio de esta actividad se consiguió la aceptación por parte de los funcionarios de la División de Gestión de la Investigación, de los requerimientos identificados y priorizados.

**Tabla 4. Priorización de requerimientos**

<b>Solicitudes de Información</b>	<b>Comentarios</b>	<b>Prioridad</b>	<b>Viabilidad</b>	<b>Rol</b>
Proyectos financiados por el sistema general de regalías (SGR)	Cantidad de proyectos que son financiados por el SGR, vinculados por facultad, para un determinado año.	Alta	Alta	Responsable SI
Proyectos financiados por Colciencias	Cantidad de proyectos que son financiados por Colciencias, vinculados por facultad, para un determinado año.	Alta	Alta	Responsable SI
Proyectos financiados por entidades externas	Cantidad de proyectos que son financiados por entidades externas, detallando los proyectos por cada facultad, para un determinado año.	Alta	Alta	Responsable SI
Entidades que han financiado proyectos de investigación	Cantidad de entidades que han financiado proyectos de investigación, detallando el monto y discriminando por el año de financiación.	Alta	Alta	Analista
Compromisos vencidos de los proyectos de investigación	Cantidad de compromisos vencidos y pendientes por entregar, discriminando el estado del proyecto y el rango de vencimiento del compromiso.	Alta	Alta	Responsable SI
Compromisos de los proyectos de investigación	Cantidad de compromisos, discriminando su responsable y fecha de vencimiento.	Alta	Alta	Responsable SI
Número de proyectos de investigación de la Universidad	Cantidad de proyectos de investigación que se tienen en la Universidad en un determinado año.	Alta	Alta	Responsable SI Ejecutivo
Número de proyectos de investigación por cada facultad de la Universidad	Cantidad de proyectos de investigación que se tienen en la Universidad en un determinado año, discriminando por la facultad a la que pertenecen.	Alta	Alta	Responsable SI Ejecutivo
Líneas de investigación de los proyectos de investigación en una convocatoria	Cantidad de proyectos de investigación discriminados por su línea de investigación y su convocatoria.	Alta	Alta	Analista Ejecutivo

Palabras clave de los proyectos de investigación por cada facultad de la Universidad	Cantidad de proyectos de investigación discriminados por sus palabras clave y facultad.	Alta	Alta	Analista Ejecutivo
Convocatorias y sus proyectos de investigación	Cantidad de proyectos, discriminados por su estado, para una determinada convocatoria.	Media	Alta	Analista
Clasificación de grupos – Medición Colciencias (Consolidado)	Cantidad de grupos de investigación discriminados por su escalafón para un determinado año.	Alta	Media	Responsable SI
Clasificación de grupos – Medición Colciencias (General)	Cantidad de grupos de investigación, discriminados por su escalafón y facultad, para un determinado año.	Alta	Media	Responsable SI
Clasificación de investigadores en Colciencias	Cantidad de investigadores según su escalafón en un determinado año.	Media	Media	Responsable SI
Grupos de investigación y sus integrantes	Número de integrantes de un grupo de investigación en un año determinado, discriminando por su escalafón y facultad	Alta	Alta	Analista
Número de jóvenes investigadores vinculados por facultad	Cantidad de jóvenes investigadores en un determinado año, discriminando por facultad.	Media	Alta	Analista
Disciplina de estudio de los grupos de investigación	Cantidad de grupos de investigación, discriminando su disciplina de estudio, para un periodo determinado.	Media	Alta	Analista Ejecutivo
Semilleros de investigación por facultad	Cantidad de semilleros de investigación, discriminados por facultad, para un determinado periodo.	Media	Alta	Analista

### **3.3.2 Diseño**

Esta subfase tiene como objetivo crear los modelos dimensionales para los procesos de proyectos, grupos, semilleros de investigación y convocatorias, con lo cual se busca cumplir los requerimientos identificados en las entrevistas. También, se identifican los sistemas fuente y los detalles de la información que será cargada a la DW. Las actividades realizadas son las siguientes:

- **Crear el modelo dimensional:** se definió la granularidad de las tablas de hechos de los esquemas estrella definidos para los cuatro procesos de negocios, las dimensiones y las medidas. Además, se describen las técnicas de diseño identificadas en estos esquemas. El detalle de los modelos creados se encuentra especificado en el Capítulo 4.
- **Crear el diseño físico de la base de datos:** se define DW relacional especificando para cada atributo que la compone el tipo de dato, la longitud, si permite nulos y si es llave primaria o foránea.
- **Diseñar el proceso ETL:** se realiza un mapeo de los datos, donde se define la fuente de datos, las transformaciones necesarias y el destino, para cada uno de los atributos y sus datos.

### **3.4 FASE DE DESARROLLO**

En esta fase se busca definir una arquitectura técnica en base a los requerimientos técnicos del sistema, implementar el prototipo de bodega de datos de proyectos de investigación para la VRI y brindar capacitación a los usuarios sobre el manejo del sistema. Esta fase se compone de cinco subfases: definición de arquitectura técnica, posteriormente y de forma iterativa, las subfases de desarrollo del back room, desarrollo de front room e integración, y finalmente se lleva a cabo la subfase de despliegue.

#### **3.4.1 Definición de arquitectura**

En esta subfase se realiza la selección e instalación de los productos del prototipo de DW teniendo en cuenta los requerimientos técnicos de la VRI, con respecto a herramientas para el motor relacional y OLAP, proceso ETL y reportes. Las actividades que se realizan en esta subfase son:

- **Reunir y documentar los requerimientos técnicos:** se revisan las entrevistas realizadas a los funcionarios de la VRI con el fin de identificar requerimientos técnicos.
- **Se especifican los requisitos de hardware, red y escritorio, e infraestructura:** necesarios para desplegar el prototipo de bodega de datos.
- **Selección e instalación de productos:** el principal objetivo de esta actividad es seleccionar las herramientas adecuadas para implementar el prototipo de DW, para lo cual se hace una revisión de herramientas disponibles en el mercado para los servicios (componentes) de motor relacional y OLAP, proceso ETL y reportes, con el fin de seleccionar las que cuentan con las características y funcionalidades que se requieren para realizar el prototipo en mención. Además, se revisa costos de herramientas para una futura puesta a producción.

Por cada componente se realiza una matriz de comparación, la cual está conformada por las siguientes columnas: *Característica*, representa las características evaluadas, *Peso* hace referencia al grado de importancia que se le asignó a la característica con base en las necesidades del proyecto, y la columna *Total* que representa el resultado de la multiplicación del puntaje y peso anteriores. La fila *Total*, representa el puntaje final que obtuvo la herramienta. Las herramientas con el puntaje total más alto por componente fueron las seleccionadas para realizar el prototipo de DW.

La Tabla 5 presenta la comparación de herramientas para motor relacional, como se puede observar las herramientas con mayor puntuación son MySQL, seguido de SQL Server y Oracle, esto debido principalmente a detalles de compatibilidad con más herramientas.

La Tabla 6 presenta la comparación de herramientas para OLAP, como se puede observar las herramientas con mayor puntuación son Pentaho Schema Workbenche, seguido de SQL Server Analysis Services, esto debido cuenta con gran variedad de funcionalidades para la implementación del cubo.

La Tabla 7 presenta la comparación de herramientas para ETL, como se puede observar las herramientas con mayor puntuación son Pentaho Data Integration seguido de SQL Server Integration Services, lo cual se debe a que cuenta con amplias características para el proceso de ETL.

La Tabla 8 presenta la comparación de herramientas para reportes, como se puede observar las herramientas con mayor puntuación son Power BI, seguido de Pentaho Reporting, esto debido principalmente a que cuentan con características técnicas para crear gran variedad de reportes.

Las herramientas seleccionadas, con mayor puntaje, son: MySQL, Pentaho Schema Workbench, Pentaho Data Integration, Power BI y Pentaho Report Designer, siendo estas últimas dos herramientas para reportes de consultas analíticas y reportes estándar, respectivamente.

En la Ilustración 7 se muestra de manera gráfica la arquitectura del prototipo de DW para proyectos de investigación de la División de gestión de la Investigación, en la cual se incluye los sistemas fuente utilizados: la base de datos relacional “BDVRI” y archivos Excel. La DW relacional denominada “VRI” se crea en Mysql, el cubo dimensional llamado “cubo\_vri” se crea en Pentaho Schema Workbench, el proceso de ETL se realiza en la herramienta Pentaho Data Integration, los reportes estándar en “Pentaho Report Designer” y las consultas analíticas se crean en Power BI.

**Tabla 5. Matriz de comparación de herramientas para motor relacional. Fuente propia**

Característica	Peso	SQL Server	Oracle	Postgresql	Mysql	Firebird	Total (SS)	Total (Oracle)	Total (PS)	Total (MS)	Total (F)
<b>Costo</b>	10	80	100	100	100	100	8	10	10	10	10
<b>Documentación</b>	20	100	100	100	100	80,4	20	20	20	20	16,1
<b>Características técnicas</b>	20	99,8	100	100	93	97	20	18,6	20	20	4
<b>Funcionalidad</b>	25	100	100	100	100	50	25	25	25	25	12,5
<b>Seguridad</b>	25	100	100	90	92,5	77,5	25	23,1	22,5	25	19,4
<b>TOTAL</b>	100						98	96,7	97,5	100	61,4

**Tabla 6. Matriz de comparación de herramientas OLAP. Fuente propia**

Característica	Peso	SQL Server Analysis Services Enterprise (SSASE)	Pentaho Schema Workbench Community (PSWC)	Iccube Enterprise (IE)	JASPERSOFT Community (JC)	Total SSASE	Total PSWC	Total IE	Total JC
<b>Costo</b>	20	80	100	80	80	16	20	16	16
<b>Documentación</b>	15	100	90	68	69,6	15	13,5	10,2	10,44
<b>Técnicas de Diseño</b>	30	100	94	88	80	30	28,2	26,4	24
<b>Funcionalidad</b>	20	95	92	83,2	65,4	19	18,4	16,64	13,08
<b>Características Técnicas</b>	15	93	95	90	69,5	13,95	14,25	13,5	10,425
<b>TOTAL</b>	100					93,95	94,35	82,74	73,945

Tabla 7. Matriz de comparación herramientas para ETL. Fuente propia.

Característica	Peso	SQL Server Integration Services Enterprise (SSIS)	Pentaho Data Integration (PDI)	Talend For Data Integration (TOSFDI)	Infosphere Datastage (IDS)	Total (SSIS)	Total (PDI)	Total (TOSFDI)	Total (IDS)
<b>Costo</b>	20	80	100	80	80	16	20	16	20
<b>Documentación</b>	15	100	80	100	80	15	12	15	12
<b>Características técnicas</b>	15	99.88	100	97.5	80	14.982	15	14.625	12
<b>Componentes de transformación</b>	35	100	98.4	98	100	35	34.44	34.3	35
<b>Funcionalidades</b>	15	100	100	95.2	100	15	15	14.28	15
<b>TOTAL</b>	100					95.982	<b>96.44</b>	94.205	94

Tabla 8. Matriz de comparación herramientas para reportes. Fuente propia.

Característica	Peso	SQL Server Reporting Services (SQL SSRS)	Power BI (PBI)	Pentaho Reporting (PR)	SQL Power Wabbit (PW)	Total SQL SSRS	Total PBI	Total PR	Total PW
<b>Costo</b>	15	80	100	80	100	12	15	12	15
<b>Documentación</b>	15	100	92.8	81.2	58.4	15	14.73	12.18	8.76
<b>Gráficos</b>	20	100	100	100	45	20	20	20	9
<b>Características técnicas</b>	15	75	100	100	100	11.25	15	15	15
<b>TOTAL</b>	100					93.25	<b>98.53</b>	93.32	74.26

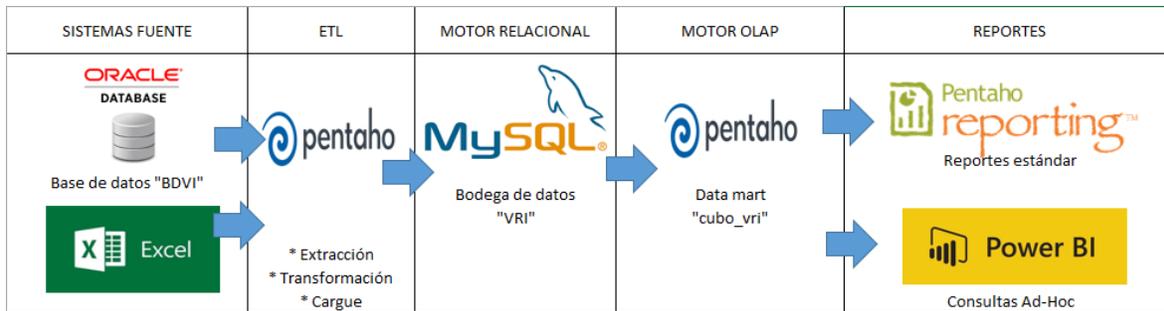


Ilustración 7. Arquitectura prototipo de DW. Fuente propia.

### 3.4.2 Desarrollo del Back Room

En esta subfase se crea el diseño físico de la DW relacional y dimensional, además el proceso de ETL, por medio de las siguientes actividades:

- **Diseño físico de la DW:** se define el estándar de nombrado a utilizar y se desarrolla el modelo físico de la DW.
- **Implementar la base datos:** se crea la base de datos BDVRI en el motor relacional de MYSQL.
- **Desarrollar el proceso ETL:** se crean trabajos y transformaciones en PDI y se ejecutan para cargar los datos en la DW relacional.

Esta subfase se especifica en la Sección 5.1, Desarrollo de Back Room.

### 3.4.3 Desarrollo del Front Room

En esta subfase se definen y crean los reportes del usuario final con base en los requerimientos identificados en la fase de análisis y diseño, para lo cual se realizan las siguientes actividades:

- **Identificar y priorizar los reportes candidatos:** se seleccionan los reportes que tienen mayor prioridad para los usuarios.
- **Diseñar la estrategia de navegación:** se organizan los reportes en carpetas, por temas analíticos, para facilitar a los usuarios la búsqueda de la información que desean.
- **Desarrollar un estándar para las aplicaciones de usuario final:** se define la presentación de los datos en forma de matriz, los estilos de las gráficas y se distribuye la información de los datos, permitiendo que el usuario identifique rápidamente lo que visualiza.
- **Desarrollar y documentar las aplicaciones de usuario final:** el proceso de creación de los reportes se encuentra especificado en la Sección 5.2, *Desarrollo del Front Room*.

### **3.4.4 Integración**

El principal objetivo de esta fase es consolidar el trabajo realizado en las subfases de Back Room y Front Room, para obtener un data mart completo, para lo cual se realizan las siguientes actividades:

- **Integración del Back Room y Front Room:** se realizan las configuraciones necesarias en las herramientas de reportes para que los usuarios finales puedan acceder a los datos cargados en la DW y también que se puedan visualizar los reportes.
- **Realizar pruebas a los datos:** se hace con el objetivo de identificar problemas de lógica y consistencia en la información y en los datos visualizados en los reportes.

### **3.4.5 Despliegue**

En esta fase se realiza una capacitación de uso del prototipo a los usuarios finales de la VRI, y su respectiva evaluación del despliegue de la DW, por medio de la aplicación de un test de satisfacción para evaluar la utilidad de los reportes estándar y las consultas analíticas que se generan en el prototipo de DW. Para esta evaluación se hace uso de la métrica *Nivel de satisfacción* (ISO/IEC 25022) [23]. El desarrollo y resultado de la evaluación de satisfacción está detallada en el Capítulo 6 de este documento.

## **3.5 FASE DE MANTENIMIENTO Y CRECIMIENTO**

En esta fase, se plantea un seguimiento al prototipo de DW puesto en producción, del mismo modo, se realizan actividades para el crecimiento del prototipo y ejecutan evaluaciones del desempeño que aseguren el correcto funcionamiento del prototipo en el transcurso del tiempo. Esta fase está fuera del alcance del proyecto.

## **3.6 FASE DE GESTIÓN DEL PROYECTO**

Esta fase se ejecuta en todo el proyecto, de tal manera que se obtiene un seguimiento a la ejecución de este, para lo cual se realizan las siguientes actividades:

- **Revisar el plan de proyecto:** se llevan a cabo reuniones de dirección del proyecto de grado con la directora de este, bajo el rol de gerente, para verificar el correcto cumplimiento de las actividades de cada fase y estar informados de los problemas que se presentan.
- **Reunión de información del estado del proyecto:** esta actividad se refiere a las diferentes reuniones de dirección del proyecto de grado. Las reuniones de revisión con los funcionarios de la División de Gestión de la Investigación se realizan solo cuando se presenta un adelanto extenso y significativo, debido a la limitada disposición de tiempo de los funcionarios.

## Capítulo 4

---

### 4 MODELADO DIMENSIONAL

El modelado dimensional permite satisfacer los requerimientos de los usuarios, proporcionando datos comprensibles y un rápido rendimiento de consultas mediante una herramienta de consulta OLAP [25].

En esta sección se presentan la descripción de los modelos dimensionales obtenidos para los procesos de negocio: *proyectos, grupos, semilleros de investigación y convocatorias*, elaborados a partir de la recolección de requerimientos. Inicialmente se presenta la matriz bus de estos procesos, se listan las dimensiones conformadas y no conformadas. Luego por cada modelo se presenta el esquema estrella, las medidas y el detalle de las dimensiones. Además, las técnicas de diseño identificados durante la elaboración de estos.

#### 4.1 MATRIZ BUS

La matriz bus muestra la relación que tienen los procesos de negocio y las dimensiones, siendo el primer paso para la elaboración de una bodega de datos [25]. La Tabla 9 presenta los modelos dimensionales obtenidos para los procesos de negocio mencionados, con dieciocho dimensiones relacionadas.

En esta matriz cada fila representa los modelos dimensionales obtenidos y las columnas representan las dimensiones involucradas. Las intersecciones marcadas (✓) indican la relación entre las dimensiones y los modelos dimensionales, siendo dimensiones conformadas aquella que presenten dos o más intersecciones.

Aunque la matriz bus no se encuentra incluida como un artefacto en la metodología de desarrollo de Mipymes, es creada debido a que es una herramienta importante en la planificación e identificación de las dimensiones conformadas asociadas a cada modelo dimensional perteneciente a un proceso de negocio.

#### 4.2 LISTA DE DIMENSIONES

Las dimensiones descritas a continuación, hacen parte los procesos de negocio mencionados anteriormente, donde se identifican las dimensiones conformadas y las no conformadas presentes en un solo modelo dimensional.

- Dimensiones conformadas:
  - Proyecto: proyectos de investigación inscritos en la Vicerrectoría de investigaciones (VRI) de la Universidad de Cauca.
  - Grupo: grupo de investigación registrado en la VRI, conformado por un conjunto de personas que interactúan para investigar y generar productos de conocimiento en uno o varios temas.
  - Persona: estudiante, docente, administrativo de la universidad u otra persona externa a la misma.
  - Fecha: fecha de inicio o de finalización de un evento.
  - Estado proyecto: modo en el que se encuentra un proyecto.
  - Datos socio demográficos: variables estadísticas con información asociada a una zona geográfica.
  - Localización: niveles de organización territorial del país.

- Convocatoria: proceso por el cual se convoca a interesados en participar en la ejecución de proyectos de investigación.
  
- Dimensiones no conformadas:
  - Tipo compromiso: clase de compromiso de los proyectos de investigación.
  - Estado compromiso: modo en el que se encuentra un compromiso.
  - Línea de investigación: enfoque de conocimiento, metodología y práctica que permiten la consolidación de teorías, paradigmas y aportes a la humanidad alrededor de un tema.
  - Indicadores de documento: provee información de una determinada condición de los documentos solicitados a un proyecto de investigación.
  - Palabra clave: palabra que permiten clasificar y direccionar las entradas en los sistemas de indexación y de recuperación de la información en las bases de datos de un manuscrito o área temática en particular.
  - Rol integrante: función que una persona desempeña como integrante en un proyecto de investigación.
  - Tipo egreso: tipo de salida de recursos financieros con el fin de cumplir un pago.
  - Rubro: conceptos en que se divide el presupuesto del proyecto de investigación, caracterizado por corresponder a un concepto específico de ingresos o gastos.
  - Entidad financiera: organización que se dedica a ofrecer servicios financieros.
  - Tipo Financiación: clase de financiación.
  - Disciplina: disciplina de estudio de los grupos de investigación.
  - Rol integrante grupo: función que una persona desempeña como integrante en un grupo de investigación.
  - Escalafón: escalafón de un grupo de investigación.
  - Estado convocatoria: modo en el cual se encuentra una convocatoria.
  - Estado semillero: modo en el cual se encuentra un semillero.
  - Programa: programa académico al cual pertenece un semillero.
  - Documentos semillero: provee información de una determinada condición de los documentos solicitados a un semillero de investigación.
  - Rol integrante semillero: función que una persona desempeña como integrante en un semillero de investigación.
  - Semillero: célula investigativa perteneciente a la Universidad.
  - Plan Trabajo: plan de trabajo y/o actividades que ejecutan los semilleros.

Tabla 9. Matriz bus. Fuente propia

Modelos dimensionales	Dimensiones																											
	Proyecto	Grupo	Persona	Fecha	Estado proyecto	Tipo compromiso	Estado compromiso	Línea de investigación	Indicadores de documento	Palabra clave	Datos socio demográficos	Rol integrante proyecto	Localización	Convocatoria	Tipo egreso	Rubro	Entidad financiera	Tipo Financiación	Disciplina	Rol integrante grupo	Escalafón	Estado convocatoria	Estado semillero	Programa	Documentos semillero	Rol integrante semillero	Semillero	Plan Trabajo
Cumplimiento de proyectos	✓	✓	✓	✓	✓	✓	✓																					
Proyectos	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	✓	✓														
Financiación de proyectos	✓	✓	✓	✓	✓										✓	✓	✓	✓										
Grupos	✓	✓	✓										✓						✓	✓	✓							
Semilleros	✓	✓	✓								✓		✓									✓		✓	✓	✓	✓	✓
Convocatorias	✓		✓	✓							✓			✓								✓						

### 4.3 MODELOS DIMENSIONALES

En esta sección se presentan los modelos dimensionales propuestos para el proceso de negocio *proyectos de investigación*, donde cada modelo cuenta con su esquema estrella, dimensiones y medidas, donde las dimensiones no conformadas se presentan en su correspondiente modelo y las conformadas se presentan una única vez. Las dimensiones conformadas se encuentran en *Azul* y las no conformadas se encuentran en *Negrita*. Los esquemas de los modelos dimensionales para *grupos*, *semilleros de investigación* y *convocatorias* se encuentran en el Anexo C.

Para cada uno de los modelos se describen las medidas presentes en las tablas de hechos y en tablas puente, si es el caso.

Estos modelos incluyen información sociodemográfica de los integrantes de proyectos de investigación que permiten una alta capacidad de análisis de estos.

#### 4.3.1 Modelo de Cumplimiento de los proyectos de investigación

El modelo de Cumplimiento de los proyectos de investigación satisface las solicitudes analíticas hechas por los usuarios, sobre el cumplimiento de los proyectos de investigación (descritas en la Sección 3.3.1), donde su esquema estrella (ver Ilustración 8) almacena información acerca de los compromisos asignados a cada proyecto de investigación.

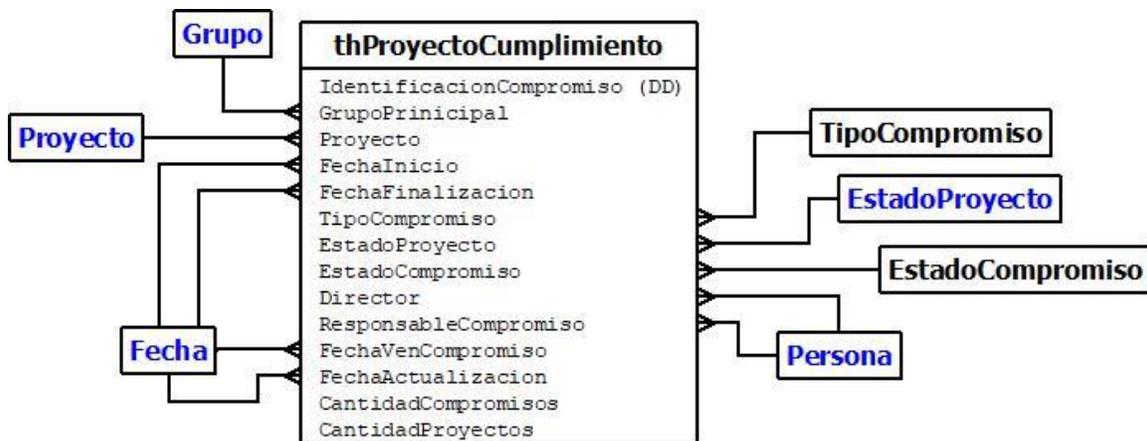


Ilustración 8. Esquema Estrella Cumplimiento de proyectos. Fuente propia.

##### 4.3.1.1 Lista de medidas

El esquema estrella de Cumplimiento de los proyectos de investigación presenta dos medidas básicas y visibles (ver Tabla 10), que son:

- **CantidadCompromisos**: cantidad de compromisos distintos en una consulta.
- **CantidadProyectos**: cantidad de proyectos de investigación distintos en una consulta.

La Tabla 10 presenta estas medidas, la función de agregación asociada a la misma y si son visibles o no al usuario.

**Tabla 10. Medidas Cumplimiento proyectos de investigación. Fuente propia.**

Medida	Función de agregación	Visibilidad
CantidadCompromisos	Recuento distinto	Visible
CantidadProyectos	Recuento distinto	Visible

#### 4.3.1.2 Detalle de dimensiones

El modelo de Cumplimiento de los proyectos de investigación está conformado por siete dimensiones, en las cuales se consolidan datos de los proyectos de investigación y su cumplimiento respecto a los compromisos asignados.

De la Tabla 11 a la Tabla 17, se describen las dimensiones del modelo detallando atributos, una descripción y ejemplo de los datos que contiene, junto a las jerarquías definidas para el mismo.

**Tabla 11. Dimensión Proyecto. Fuente propia.**

Dimensión Proyecto		
Almacena información del proyecto de investigación.		
Atributos	Descripción	Ejemplo
IDProyecto	Clave primaria de esta dimensión, es una secuencia auto incremental.	1, 2, 3, ...
IdentificadorProyecto	Clave natural de esta dimensión, es el código que trae desde la base de datos de origen.	17, 18, 19, ...
TipoProyecto	Tipo de proyecto de investigación.	Innovación
Título	Nombre del proyecto de investigación.	LINEAMIENTOS DE USABILIDAD PARA EL DISEÑO DE APLICACIONES, ANÁLISIS DE REDES...
Email	Correo electrónico del proyecto	trojas@unicauca.edu.co, gramirez@unicauca.edu.co, ...
Teléfono	Número de contacto del responsable.	82090000, 8233221, ...
ObjetivoSocioeconómico	Objetivo socioeconómico del proyecto de investigación	Brindar apoyo agrícola, ...
Lugar ejecución	Lugar de ejecución del proyecto de investigación	Finca El Sol, Cajibío, ...
Clasificación	Clasificación del proyecto de investigación	Funcionamiento VRI , Jóvenes Investigadores, ...
Subclasificación	Subclasificación del proyecto de investigación	Proyectos de investigación, Proyección social, ...
Jerarquías	Descripción	Ejemplo
Jerarquía Clasificación Proyecto	Jerarquía compuesta por: Clasificación, Subclasificación	Jóvenes investigadores->Proyección social.

**Tabla 12. Dimensión grupo. Fuente propia**

<b>Dimensión Grupo</b>		
Almacena información de los grupos de investigación de la Universidad.		
<b>Atributos</b>	<b>Descripción</b>	<b>Ejemplo</b>
IDGrupo	Clave primaria de esta dimensión, es una secuencia auto incremental.	1, 2, 3, ...
IdentificadorGrupo	Clave natural de esta dimensión, es el código que trae desde la base de datos de origen.	1569, 1570, 1577, ...
NombreGrupo	Nombre del grupo de investigación.	ARTES 2000, DISEÑO Y DESARROLLO, ...
NombreFacultad	Nombre de la facultad.	FACULTAD DE ARTES, FACULTAD DE CIENCIAS DE LA SALUD, ...
SiglaFacultad	Siglas de la facultad.	FCS, FACA, ...
DirGrupoLac	Dirección web en Colciencias.	www.minciencias.org/grupa12/
Tipo	Tipo de grupo de investigación	Grupo
Dirección	Dirección del grupo de investigación.	Cra 1 No. 12-12, ...
Teléfono	Número del teléfono del grupo de investigación	8209000, ...
Fax	Número del fax del grupo de investigación.	8209001, ...
Email	Correo electrónico del grupo de investigación.	correo@unicauca.edu.co, ...
www	Página web del grupo	www.grupo.com, ...
CodigoColciencias	Código del grupo dentro de Colciencias.	Grup10A, ...
<b>Jerarquías</b>	<b>Descripción</b>	<b>Ejemplo</b>
Jerarquía Grupo	Jerarquía compuesta por: SiglaFacultad, IdentificadorGrupo	FIET->18...

**Tabla 13. Dimensión Fecha. Fuente propia.**

<b>Dimensión Fecha</b>		
Almacena información de fecha.		
<b>Atributos</b>	<b>Descripción</b>	<b>Ejemplo</b>
ID Fecha	Clave primaria de esta dimensión, es una secuencia auto incremental.	20000116, 20000117, 20000118, ...
Fecha Completa	Año, mes y día.	2000-01-16, 2000-01-17, 2000-01-18 ...
Año	Año que pertenece a la fecha.	2001, 2002, 2003, ...
Semestre	Periodo de tiempo, de seis meses, en el que se encuentra el año.	1,2.
Periodo Académico	Parte de un año académico.	2000-1, 2000-2, ...
Mes	Representación numérica del mes.	1,2, 3, ...
Nombre Mes	Nombre del mes.	Enero, febrero, ...
<b>Jerarquías</b>	<b>Descripción</b>	<b>Ejemplo</b>

Jerarquía Fecha	Jerarquía compuesta por: año, semestre, mes y día.	2021-> 2-> Octubre-> 1, 2022->1-> Enero-> 4...
-----------------	--	--

**Tabla 14. Dimensión Persona. Fuente propia.**

<b>Dimensión Persona</b>		
Almacena información de la persona.		
<b>Atributos</b>	<b>Descripción</b>	<b>Ejemplo</b>
IDPersona	Clave primaria de esta dimensión, es una secuencia auto incremental.	1,2,3, ...
IdentificaciónPersona	Clave natural de la dimensión Persona, es el código que trae desde la base de datos de origen.	1003801239, 1004082960, ...
Tipoidentificacion	Tipo de identificación de la persona.	Cédula de ciudadanía.
NombreCompleto	Nombres y apellidos completos de la persona.	Juan Diego Losada Caballero, Valentina Muñoz Méndez, ...
Teléfono	Número telefónico de la persona.	8233333, 8244343, 8255643, ...
Celular	Número celular de la persona.	3123456432, 2111342356, ...
Dirección	Dirección de residencia de la persona.	Cra 1 No 12-12, Calle 3 No 12-12 ...

**Tabla 15. Dimensión Estado Compromiso. Fuente propia.**

<b>Dimensión Estado Compromiso</b>		
Almacena información del estado de un compromiso.		
<b>Atributos</b>	<b>Descripción</b>	<b>Ejemplo</b>
IDEstadoCompromiso	Clave primaria de esta dimensión, es una secuencia auto incremental.	1,2,3, ...
Categoría	Categoría que tiene un compromiso.	Por entregar, cumplido.
Subcategoría	Subcategoría que tiene un compromiso.	A tiempo, vencido.
Rango	Rango de días para el estado de un compromiso, de acuerdo a su fecha de vencimiento.	0-10 días 11-30 días >30 días.
<b>Jerarquías</b>	<b>Descripción</b>	<b>Ejemplo</b>
Jerarquía Estado Compromiso	Jerarquía compuesta por: Categoría, Subcategoría y Rango.	Pendiente-> vencido -> "> 30 días"

**Tabla 16. Dimensión Estado Proyecto. Fuente propia.**

<b>Dimensión Estado Proyecto</b>		
Almacena información del estado de un proyecto.		
<b>Atributos</b>	<b>Descripción</b>	<b>Ejemplo</b>
IDEstadoProyecto	Clave primaria de esta dimensión, es una secuencia auto incremental.	1,2,3, ...
EstadoProyecto	Estado que tiene un proyecto de investigación.	Formulado, Aprobado, En Ejecución, ...
<b>Jerarquías</b>	<b>Descripción</b>	<b>Ejemplo</b>

Jerarquía Estado Proyecto	Jerarquía compuesta por: EstadoProyecto, IDEstadoProyecto	Formulado->3.
---------------------------	---	---------------

Tabla 17. Dimensión Tipo Compromiso. Fuente propia.

Dimensión Tipo Compromiso		
Almacena información del tipo de un compromiso.		
Atributos	Descripción	Ejemplo
IDTipoCompromiso	Clave primaria de esta dimensión, es una secuencia auto incremental.	1,2,3, ...
TipoCompromiso	Tipo que tiene un compromiso.	Informe técnico parcial, Informe técnico final, ...
Jerarquías	Descripción	Ejemplo
Jerarquía Tipo compromiso	Jerarquía compuesta por: TipoCompromiso, IDTipoCompromiso	Informe técnico final->2

### 4.3.2 Modelo de Proyectos de investigación

El modelo de Proyectos de investigación satisface a las solicitudes analíticas, hechas por los usuarios, sobre los proyectos de investigación (Sección 3.3.1), donde su esquema estrella (ver Ilustración 9) muestra las dimensiones que lo conforman y sus medidas.

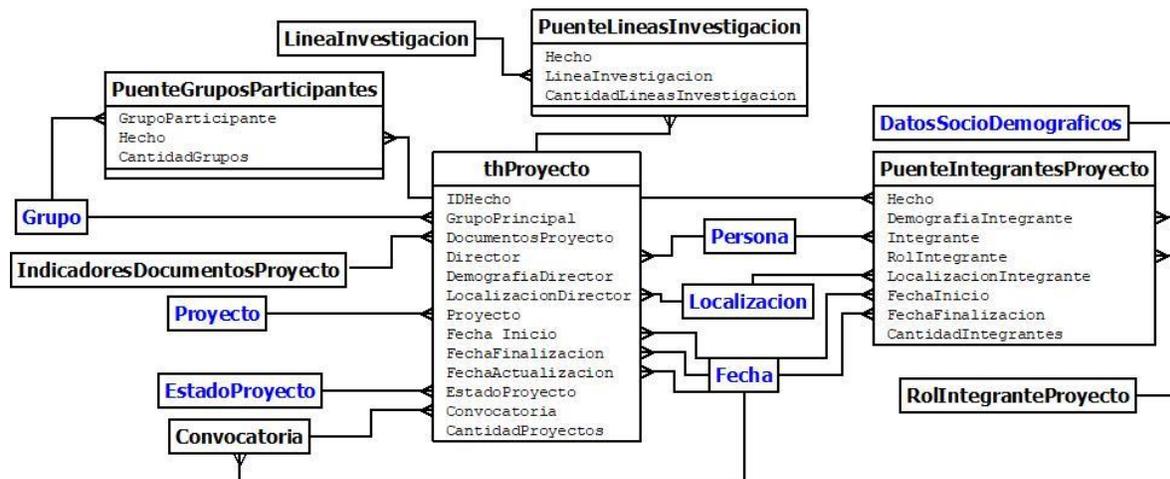


Ilustración 9. Esquema Estrella de Proyectos. Fuente propia.

#### 4.3.2.1 Lista de medidas

El esquema estrella de Proyectos de investigación presenta cuatro medidas básicas, visibles (ver Tabla 18).

Las medidas de la tabla de hechos son:

- CantidadProyectos: cantidad de proyectos de investigación distintos en una consulta.

La tabla *PuenteGruposParticipantes* tiene una sola medida:

- CantidadGrupos: cantidad de grupos distintos en una consulta.

La tabla *PuenteIntegrantesProyecto* tiene una sola medida:

- CantidadIntegrantes: cantidad de integrantes distintos en una consulta.

La tabla *PuenteLíneas investigación* tiene una sola medida:

- CantidadLíneasInvestigación: cantidad de líneas de investigación distintas en una consulta.

**Tabla 18. Medidas Proyectos de investigación. Fuente propia.**

Tabla	Medida	Función de Agregación	Visibilidad
<i>thProyecto</i>	CantidadProyectos	Recuento distinto	Visible
<i>PuenteGrupos Participantes</i>	CantidadGrupos	Recuento distinto	Visible
<i>PuenteIntegrantes Proyecto</i>	CantidadIntegrantes	Recuento distinto	Visible
<i>PuenteLíneasInvestigacion</i>	CantidadLineasInvestigacion	Recuento distinto	Visible

#### 4.3.2.2 Detalle de dimensiones

El modelo de Proyectos de investigación está conformado por nueve dimensiones, en las cuales se consolidan datos de los proyectos de investigación. De la Tabla 19 a la

Tabla 25 se presenta el detalle de las dimensiones propias del modelo, ya que las dimensiones conformadas están detalladas en la Sección 4.3.1.

**Tabla 19. Dimensión convocatoria. Fuente propia**

Dimensión Convocatoria		
Almacena información de la convocatoria almacenada		
Atributos	Descripción	Ejemplo
IDConvocatoria	Clave primaria de esta dimensión, es una secuencia auto incremental.	1, 2, 3, ...
Consecutivo	Clave natural de esta dimensión, es el código que trae desde la base de datos de origen.	17, 18, 19, ...
Convocatoria	Nombre de la convocatoria.	PROYECTO DE INVESTIGACIÓN 2005, PROYECTO DE DESARROLLO INTERNO, ...
InternaExterna	Tipo de convocatoria.	Interna, externa.
FechaInicio	Clave foránea de la dimensión <i>Fecha</i> para hacer referencia a la fecha de inicio de la convocatoria.	20000116, 20000117, 20000118, ...
FechaCierre	Clave foránea de la dimensión <i>Fecha</i> para hacer referencia a la fecha de cierre de la convocatoria.	20000116, 20000117, 20000118, ...

FechaResultados	Clave foránea de la dimensión <i>Fecha</i> para hacer referencia a la fecha de resultados de la convocatoria.	20000116, 20000117, 20000118, ...
Estado	Estado de la convocatoria.	Abierta, cerrada, terminada.
<b>Jerarquías</b>	<b>Descripción</b>	<b>Ejemplo</b>
Jerarquía Convocatoria	Jerarquía compuesta por: InternaExterna, Estado	Interna, Terminada

**Tabla 20. Dimensión Localización. Fuente propia.**

<b>Dimensión Localización</b>		
Almacena información de la localización.		
<b>Atributos</b>	<b>Descripción</b>	<b>Ejemplo</b>
IDLocalización	Clave primaria de esta dimensión, es una secuencia auto incremental.	1, 2, 3, ...
País	Nombre del país de localización	COLOMBIA, ECUADOR, ...
Departamento	Nombre del departamento de localización	CAUCA, ANTIOQUIA, ...
Ciudad	Nombre de la ciudad de localización	POPAYÁN, MEDELLIN, ...
<b>Jerarquías</b>	<b>Descripción</b>	<b>Ejemplo</b>
Jerarquía Localización	Jerarquía compuesta por: país, departamento y municipio.	Colombia-> Cauca-> Popayán, ...

**Tabla 21. Dimensión Indicadores Documentos Proyecto. Fuente propia**

<b>Dimensión Indicadores Documentos Proyecto</b>		
Almacena información de los indicadores de documento del proyecto.		
<b>Atributos</b>	<b>Descripción</b>	<b>Ejemplo</b>
IDDocumentoProyecto	Clave primaria de esta dimensión, es una secuencia auto incremental.	1, 2, 3, ...
Convenio	Indica si el proyecto ha entregado el documento Convenio	0, 1
AvalDepto	Indica si el proyecto ha entregado el documento Aval de Departamento	0, 1
ActaInicio	Indica si el proyecto ha entregado el documento Acta de Inicio	0, 1
CartaEntidades	Indica si el proyecto ha entregado el documento Carta de Entidades	0, 1
Memorandun	Indica si el proyecto ha entregado el documento Memorandun	0, 1

ViabilidadEconomica	Indica si el proyecto ha entregado el documento Viabilidad Económica	0, 1
ViabilidadAcademica	Indica si el proyecto ha entregado el documento Viabilidad Académica	0, 1
SolicitudViabilidad	Indica si el proyecto ha entregado el documento Solicitud Viabilidad	0, 1
ActaLiquidacion	Indica si el proyecto ha entregado el documento Acta de Liquidación	0, 1
Cronograma	Indica si el proyecto ha entregado el documento Cronograma	0, 1
AvalFacultad	Indica si el proyecto ha entregado el documento Aval de Facultad	0, 1
SoportesEticos	Indica si el proyecto ha entregado el documento Soportes éticos	0, 1

**Tabla 22. Dimensión Línea de Investigación. Fuente propia.**

<b>Dimensión Línea Investigación</b>		
Almacena información sobre las líneas de investigación de los proyectos		
<b>Atributos</b>	<b>Descripción</b>	<b>Ejemplo</b>
IDLineaInvestigacion	Clave primaria de esta dimensión, es una secuencia auto incremental.	1, 2, 3, ...
NombreLineaInvestigacion	Nombre de la línea de investigación	Bases de datos, Ingeniería del Software, ...

**Tabla 23. Dimensión Palabra Clave. Fuente propia.**

<b>Dimensión Palabra Clave</b>		
Almacena información sobre palabras clave de los proyectos		
<b>Atributos</b>	<b>Descripción</b>	<b>Ejemplo</b>
IDPalabraClave	Clave primaria de esta dimensión, es una secuencia auto incremental.	1, 2, 3, ...
PalabraClave	Nombre de la palabra clave	Datos, ETL, Innovación, ...

**Tabla 24. Dimensión Datos Socio Demográficos. Fuente propia.**

<b>Dimensión Datos Socio Demográficos</b>		
Almacena información sobre los datos socio demográficos de las personas		
<b>Atributos</b>	<b>Descripción</b>	<b>Ejemplo</b>
IDDatosSocioDemograficos	Clave primaria de esta dimensión, es una secuencia auto incremental.	1, 2, 3, ...
Genero	Género de la persona	Femenino, Masculino
MaximoNivelEscolar	Máximo nivel de escolaridad de la persona	Bachillerato, Primaria, ...

TipoVinculacion	Tipo de vinculación de una persona con la Universidad	Docente, Estudiante, ...
-----------------	---	--------------------------

Tabla 25. Dimensión Rol Integrante Proyecto. Fuente propia.

Dimensión Rol Integrante Proyecto		
Almacena información sobre los roles de los proyectos		
Atributos	Descripción	Ejemplo
IDRollIntegranteProyecto	Clave primaria de esta dimensión, es una secuencia auto incremental.	1, 2, 3, ...
RollIntegranteProyecto	Nombre del rol de integrante	Director de proyecto, asesor, estudiante pregrado, ...

### 4.3.3 Modelo de Financiación de proyectos de investigación

El modelo de Financiación de proyectos de investigación satisface a las solicitudes analíticas, hechas por los usuarios, sobre la financiación de proyectos de investigación (ver Sección 3.3.1) donde su esquema estrella (ver Ilustración 10) muestra las dimensiones que lo conforman y sus medidas.

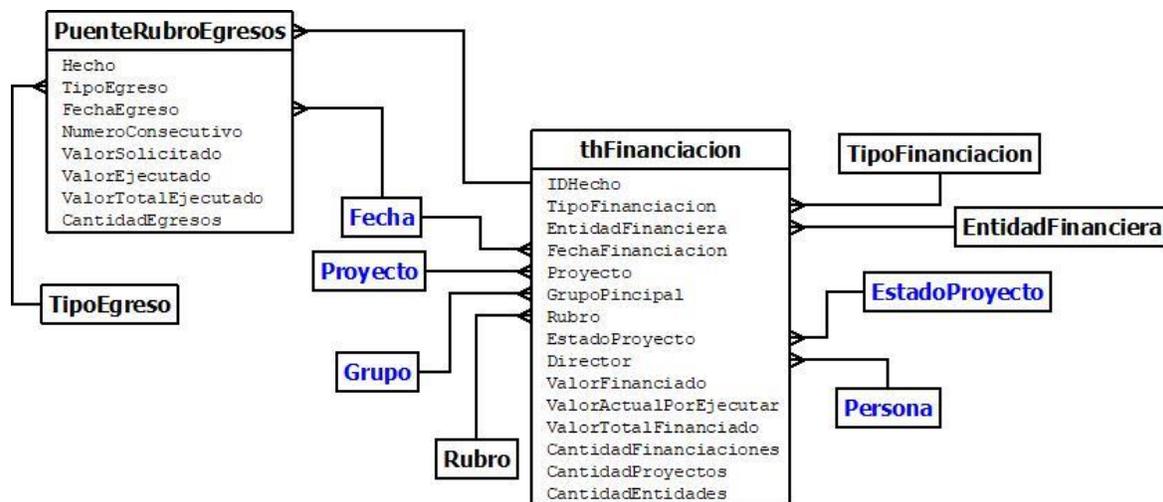


Ilustración 10. Esquema Estrella Financiación proyectos. Fuente propia.

#### 4.3.3.1 Lista de medidas

El esquema estrella de Financiación de proyectos de investigación (ver Tabla 26) presenta cuatro medidas básicas, visibles.

Las medidas de la tabla de hechos son:

- CantidadFinanciaciones: cantidad de financiadores distintos en una consulta.
- CantidadProyectos: cantidad de proyectos de investigación distintos en una consulta.
- CantidadEntidades: cantidad de entidades financiera distintas en una consulta.

La tabla *PueneteRubroEgresos* tiene una sola medida:

- *CantidadEgresos*: cantidad de Egresos distintos en una consulta.

**Tabla 26. Medidas Financiación de proyectos de investigación. Fuente propia.**

Tabla	Medida	Función de agregación	Visibilidad
<i>thFinanciación</i>	CantidadFinanciaciones	Recuento distinto	Visible
	CantidadProyectos	Recuento distinto	Visible
	CantidadEntidades	Recuento distinto	Visible
<i>PuenteRubrosEgresos</i>	CantidadEgresos	Recuento distinto	Visible

#### 4.3.3.2 Detalle de dimensiones

El modelo de Financiación de proyectos de investigación está conformado por nueve dimensiones, en las cuales se consolidan datos de las financiaciones asociadas a los proyectos de investigación. De la Tabla 27 a la Tabla 30 se presenta el detalle de las dimensiones propias del modelo, ya que las dimensiones conformadas están detalladas en la Sección 4.3.1.

**Tabla 27. Dimensión Rubro. Fuente propia.**

Dimensión Rubro		
Almacena información del rubro		
Atributos	Descripción	Ejemplo
IDRubro	Clave primaria de esta dimensión, es una secuencia auto incremental.	1, 2, 3, ...
Rubro	Nombre del rubro.	Viaje, equipo, capacitación, ...

**Tabla 28. Dimensión Entidad Financiadora. Fuente propia.**

Dimensión Entidad Financiadora		
Almacena información de la entidad financiadora		
Atributos	Descripción	Ejemplo
IDEntidadFinanciera	Clave primaria de esta dimensión, es una secuencia auto incremental.	1, 2, 3, ...
EntidadFinanciera	Nombre de la entidad financiera	Minciencias, Gobernación del Cauca, ...
TipoEntidad	Tipo de entidad financiera	Interna, Externa.
Jerarquías	Descripción	Ejemplo
Jerarquía Entidad Financiera	Jerarquía compuesta por: TipoEntidad, EntidadFinanciera	Externa, Minciencias

**Tabla 29. Dimensión Tipo de Egreso. Fuente propia.**

Dimensión Tipo Egreso		
Almacena información del tipo de egreso		
Atributos	Descripción	Ejemplo
IDTipoEgreso	Clave primaria de esta dimensión, es una secuencia auto incremental.	1, 2, 3, ...

TipoEgreso	Tipo de egreso	Compra, comisión, ...
------------	----------------	-----------------------

**Tabla 30. Dimensión Tipo de Financiación. Fuente propia**

Dimensión Tipo Financiación		
Almacena información del tipo de financiación		
Atributos	Descripción	Ejemplo
IDTipoFinanciación	Clave primaria de esta dimensión, es una secuencia auto incremental.	1, 2, 3, ...
TipoFinanciación	Tipo de financiación	Financiación, intercambio, ...

#### 4.4 ESPECIFICACIÓN DE LAS TÉCNICAS DE DISEÑO

En esta sección se presentan las técnicas de diseño identificadas en el proceso de elaboración de los modelos dimensionales presentados en la matriz bus (ver Tabla 9). Estas técnicas de diseño indican la forma de manejar situaciones de modelado que se presentan habitualmente en diferentes tipos de negocios, donde su correcta aplicación permite un ahorro significativo en cuanto a espacio de almacenamiento de información, disminuye la complejidad de uso para los usuarios finales reduciendo el número de dimensiones (dimensión basura) y permite incluir correctamente aquellas dimensiones que no poseen atributos evitando adicionarlas al modelo (dimensión degenerada); además, brinda la posibilidad de realizar consultas que involucren diferente granularidad entre dimensiones y tablas de hechos (dimensión multivaluada y tabla puente) [25].

A continuación, se presentan las técnicas de diseño identificadas en los modelos dimensionales, acompañadas de su descripción y en que parte del modelo es aplicada.

##### 4.4.1 Minidimensión

Esta técnica de diseño está presente en todos los modelos dimensionales, en el que la dimensión cuenta con atributos que cambian en el tiempo, y es necesario almacenar su histórico.

La técnica de diseño minidimensión, se identificó en las dimensiones *Estado Proyecto* (ver Ilustración 11), *Estado Compromiso*, *Tipo Compromiso*, *Indicadores Documentos Proyecto*, *Localización* y *Datos Socio Demográficos*. La dimensión *Estado Proyecto* contiene información del estado de los proyectos de investigación, permitiendo obtener información de acuerdo con este.

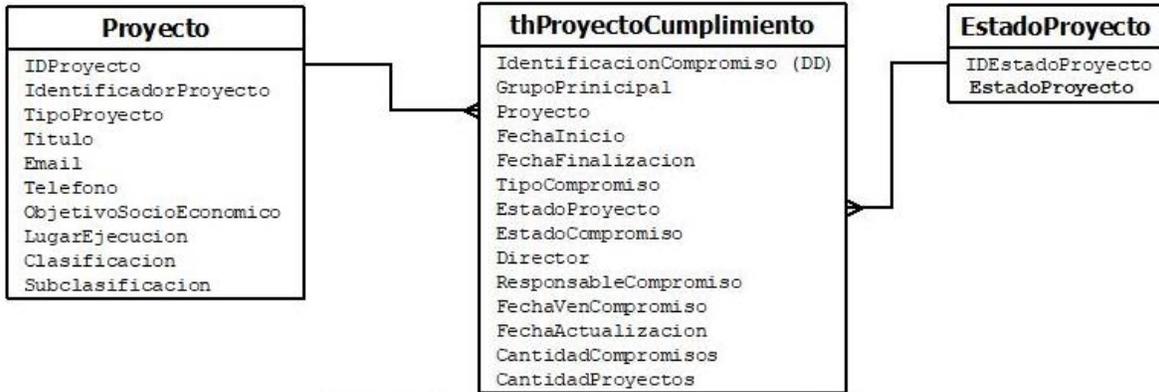


Ilustración 11. Minidimensión. Fuente propia

#### 4.4.2 Dimensión degenerada

La técnica de diseño dimensión degenerada solo contiene su clave principal, la cual es incluida en la tabla de hechos y lleva consigo un “DD” al final, lo cual indica que no tiene asociada dimensión alguna.

Esta técnica de diseño se identificó en la dimensión *Identificación Compromiso* (ver Ilustración 12) lo que permite realizar agrupaciones en las consultas.



Ilustración 12. Dimensión degenerada. Fuente propia.

#### 4.4.3 Dimensión multivaluada y tabla puente

La técnica de diseño multivaluada y tabla puente, mediante una tabla puente representa una relación muchos a muchos entre una tabla de hechos y una dimensión, donde dicha tabla puente contiene las claves primarias de las partes relacionadas, es decir, de la tabla de hechos (a quien se le agrega una clave alternativa) y la dimensión involucrada.

Esta técnica de diseño se identificó en las dimensiones: *Línea Investigación* (Ver Ilustración 13), *Grupo*, *Línea Investigación*, *Palabra Clave*, *Persona*, *Datos Sociodemográficos*, *Rol Integrante Proyecto* y *Tipo Egreso*. La Ilustración 13 muestra como esta relación muchos a

muchos se realiza mediante la tabla *PuenteLineasInvestigacion*, la cual contiene las claves de las tablas involucradas.

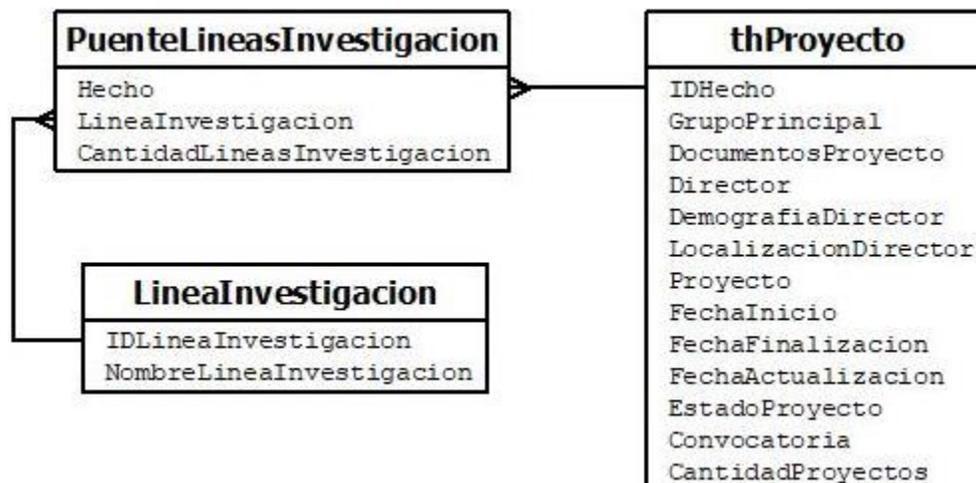


Ilustración 13. Dimensión multivaluada y tabla puente. Fuente propia.

#### 4.4.4 Juego de roles

En esta técnica de diseño se presenta cuando existe una relación múltiple entre dos tablas. Se puede observar en la Ilustración 14 una relación doble entre la dimensión *Persona* y la tabla de hechos *TH\_ProyectoCumplimiento*, para considerar información relacionada con el director del proyecto de investigación y el responsable del compromiso, donde los roles de director y responsable son de relevancia para el análisis del cumplimiento de los proyectos.

Esta técnica de diseño se identificó en las dimensiones *Persona* (Ver Ilustración 14) y *Fecha*.



Ilustración 14. Juego de roles. Fuente propia.

## Capítulo 5

---

### 5 PROTOTIPO BODEGAS DE DATOS PARA PROYECTOS DE INVESTIGACIÓN

En este capítulo se hace una descripción de las subfases de desarrollo, Back Room, Front Room, Integración y Despliegue, mediante las cuales se desarrolla el prototipo de bodegas de datos (DW) para el proceso de negocio de *proyectos de investigación*, relacionado con los modelos dimensionales presentados en el Capítulo 4.

#### 5.1 DESARROLLO DE BACK ROOM

Del desarrollo de esta subfase se presentan las actividades de desarrollo del proceso de extracción, transformación y carga (ETL), población y validación de los datos que permiten poblar la DW, haciendo uso de las fuentes de datos identificadas en la fase de análisis y diseño, estas son:

- **SIVRI:** permite gestionar la información de los grupos de investigación y sus semilleros, los proyectos de investigación, los integrantes, áreas de conocimiento, así como también, las convocatorias internas o externas que se presentan en la Universidad.
- **Archivos Excel:** almacenan información adicional, acerca de los procesos de negocio, que no se encuentra en SIVRI. Es decir, información que se encuentra en fuentes externas y es proporcionada en documentos de este formato. Por ejemplo, información del sistema de Minciencias.

Aunque se identificaron los sistemas fuentes necesarios para obtener la información para cargar la DW con información real, no fue posible hacerlo en su totalidad dada la confidencialidad y restricciones de acceso. Para los esquemas de *Proyectos de Investigación y Cumplimiento de los Proyectos de Investigación* se carga información real desde las fuentes de datos. Sin embargo, esto no es posible para el esquema de *Financiación de los Proyectos de Investigación*. A continuación, se presentan las actividades realizadas en esta subfase.

##### 5.1.1 Desarrollo del proceso ETL

En este proceso se crean y ejecutan transformaciones y trabajos de ETL de los sistemas fuente a la DW. Se identifican y solucionan problemas en cuanto a la calidad de los datos con el fin de extraer información viable, lo cual se realiza con el apoyo del encargado del sistema de información (SI) de la VRI, quien resuelve dudas con respecto a la estructura de la base de datos e información que almacena.

El proceso ETL abarca la creación de los trabajos y transformaciones para el cargue inicial e incremental para las tablas de dimensión, tablas puente y tablas de hechos. El cargue inicial sólo se ejecuta una vez durante todo el ciclo de vida de la DW, y los cargues incrementales en un período de tiempo establecido, en este caso, diariamente para los esquemas de *Cumplimiento de los Proyectos de Investigación* y *Financiación de los Proyectos de Investigación* (la información debe ser actualizada día a día por aspectos

relacionados a rangos en fecha de cumplimientos y fechas de movimientos financieros); y semanal para el esquema de *Proyectos de Investigación*.

La herramienta seleccionada para realizar el cargue de información (ver sección 3.4.1 *Definición de la arquitectura*) fue es *Pentaho Data Integration* (PDI), la cual brinda dos tipos de controles de flujo: *Trabajos*, como flujo principal dentro de la herramienta; y *Transformaciones*, las cuales definen funciones a realizar para el manejo de los datos. A continuación se presenta una descripción de los controles utilizados [26]:

- **Trabajos:** Los trabajos son modelos similares a flujos de trabajo para coordinar recursos, ejecución y dependencias de actividades ETL. Los trabajos agregan piezas individuales de funcionalidad para implementar un proceso completo, ejemplos de tareas comunes realizadas en un trabajo incluyen: obtener archivos, verificar condiciones como la existencia de una tabla de base de datos de destino necesaria, ejecutar una transformación que llene esa tabla y enviar por correo electrónico un registro de errores si falla una transformación. Los archivos de trabajos tienen una extensión *.kjb*.
- **Transformación:** Es una red de tareas lógicas llamadas pasos. Estas son esencialmente flujos de datos y se definen como un conjunto lógico de configuraciones de transformación de datos. Estos archivos tienen una extensión *.ktr*. Los dos componentes principales asociados con las transformaciones son los pasos y los saltos:
  - **Pasos:** son los componentes básicos de una transformación, por ejemplo, una entrada de archivo de texto o una salida de una tabla. Hay más de 140 pasos disponibles en PDI y están agrupados según su función; por ejemplo, entrada, salida, secuencias de comandos, etc. Cada paso de una transformación está diseñado para realizar una tarea específica, como leer datos de un archivo sin formato, filtrar filas y/ modificar datos de una base de datos. Los pasos se pueden configurar para realizar las tareas que se necesitan.
  - **Saltos:** son rutas de datos que conectan pasos entre sí y permiten que los metadatos del esquema pasen de un paso a otro. En Ilustración 15, parece que se está produciendo una ejecución secuencial; sin embargo, eso no es cierto. Los saltos determinan el flujo de datos a través de los pasos, no necesariamente la secuencia en la que se ejecutan. Cuando se ejecuta una transformación, cada paso se inicia en su propio subproceso e inserta, y pasa datos.

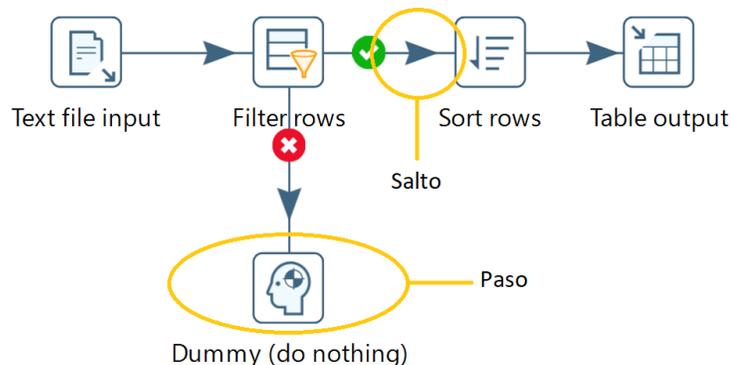


Ilustración 15. Paso y salto en una transformación. Tomado de [27]

Teniendo en cuenta los controles y elementos descritos se realiza el cargue inicial, donde se comienza con dimensiones que requieren un solo sistema fuente, luego las demás dimensiones, y se finaliza con las tablas de hechos y las dimensiones puente.

En la mayoría de los modelos diseñados, se cuenta con el elemento *FechaActualización* que almacena las fechas en las que se actualizan los registros de las tablas de hechos para el cargue incremental. La herramienta de PDI permite definir en un mismo flujo de pasos el cargue inicial e incremental de una tabla de hechos.

A continuación, se describe el flujo de datos para el proceso ETL para las tablas: *Proyecto*, *Localización*, *Fecha*, *TH\_ProyectoCumplimiento* y *PuentePalabrasClave*.

### 1. Tabla de dimensión Proyecto

En la Ilustración 16 se observa el cargue de la dimensión *Proyecto*, el cual describe el siguiente flujo:

**Extracción:** se extraen datos provenientes de la fuente VRI, mediante el uso del paso “*Proyecto*”, donde se obtiene información como *título del proyecto*, *tipo de proyecto*, *ID del proyecto*, etc.

**Transformación:** se realiza la conversión de datos (*TituloMayuscula-TipolNiciaMayuscula*), se intercambian valores *NULL* por “*No hay datos*” (*ReemplazarNULL*), esto con el fin de cargar la información sin tener información nula.

**Carga:** las columnas del flujo de la transformación se asocian con los registros y los atributos de la dimensión *Proyecto* de la DW, donde se verifica la existencia del registro para que sea actualizado o en su defecto, ingresado como nuevo registro, asignando una nueva llave y cargando sus datos (*Insertar/Actualizar\_DimProyecto*).



**Ilustración 16. Cargue dimensión Proyecto. Fuente propia.**

### 2. Tabla de dimensión Localización

En la Ilustración 17 se observa el cargue de la dimensión *Localización*, el cual describe el siguiente flujo:

**Extracción:** se extraen los datos a partir de un archivo Excel generado de forma manual, esto realizado con el paso “*Localizacion*”, pues no se contaba con esta información en la base datos VRI. La información de este archivo está relacionada con el país *Colombia*, sus departamentos y ciudades.

**Transformación:** se realiza transformación de datos, por medio del paso “*Mayusculas*”, donde se definen en mayúscula todos los atributos extraídos para la dimensión, de esta manera definir un estándar para las localizaciones registradas en la DW.

**Carga:** las columnas del flujo de la transformación se asocian con los registros y los atributos de la dimensión *Localización* de la DW, donde se verifica la existencia del registro para que sea actualizado o en su defecto, ingresado como nuevo registro, asignando una nueva llave y cargando sus datos (*Insertar/Actualizar\_DimLocalización*).



**Ilustración 17. Cargue dimensión Localización. Fuente propia**

### 3. Tabla de dimensión Fecha

En la Ilustración 18 se observa el cargue de la dimensión *Fecha*, el cual describe el siguiente flujo:

**Extracción:** se extraen datos provenientes de VRI, mediante el uso de los pasos “*FechasInicioProyecto*”, “*FechasFinProyecto*”, “*FechasInicioConvocatoria*”, “*FechasFinConvocatoria*”, “*FechasResultadosConvocatoria*”, “*FechasInicioRolIntegrante*”, “*FechasFinRolIntegrante*” y “*SYSDATE*”, donde se tienen fechas de inicio y fin de proyectos, inicio, cierre y resultado de convocatorias, y fechas de inicio y fin del rol desempeñado por un integrante dentro de un proyecto, además de la fecha del sistema.

**Transformación:** con el fin de transformar el tipo de dato al formato “YYYYMMDD” (año, mes, día), se realizan procesos como “*FechaAMD*” y “*CambioNombre*” para posteriormente obtener el ID de las fechas obtenidas, usando los pasos “*OrdenFecha*”, “*SepararHorasMinsSegs*”, “*SepararAtributos*” y “*ObtenerID*”. Luego se asignan los semestres, mes y periodo académico para cada fecha obtenida, haciendo uso de los pasos “*AñadirSemestre*”, “*AsignarMes*” y “*CrearPeriodoAcadémico*”. Por último, se comparan fechas por medio de su ID, con el cual se buscan dentro de la dimensión y verificar si se crea un nuevo registro o se actualiza, esto con uso del paso “*Insertar/ActualizarFecha*”.

**Carga:** las columnas del flujo de la transformación se asocian con los registros y los atributos de la dimensión *Fecha* de la DW, donde se verifica la existencia del registro para que sea actualizado o en su defecto, ingresado como nuevo registro, asignando una nueva llave y cargando sus datos (*Insertar/Actualizar\_DimFecha*).

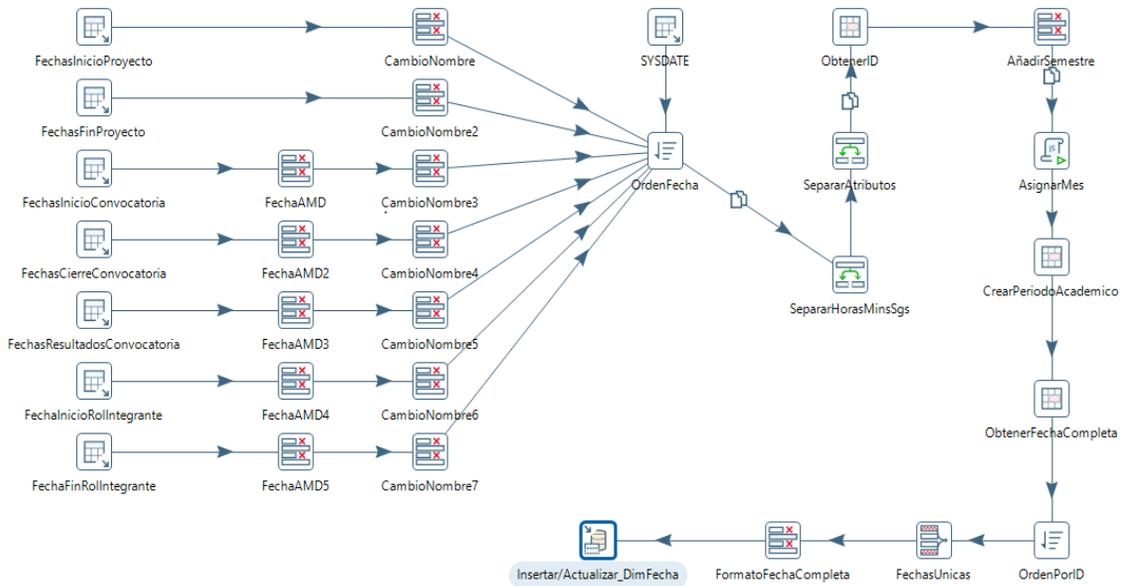


Ilustración 18. Cargue dimensión Fecha. Fuente propia.

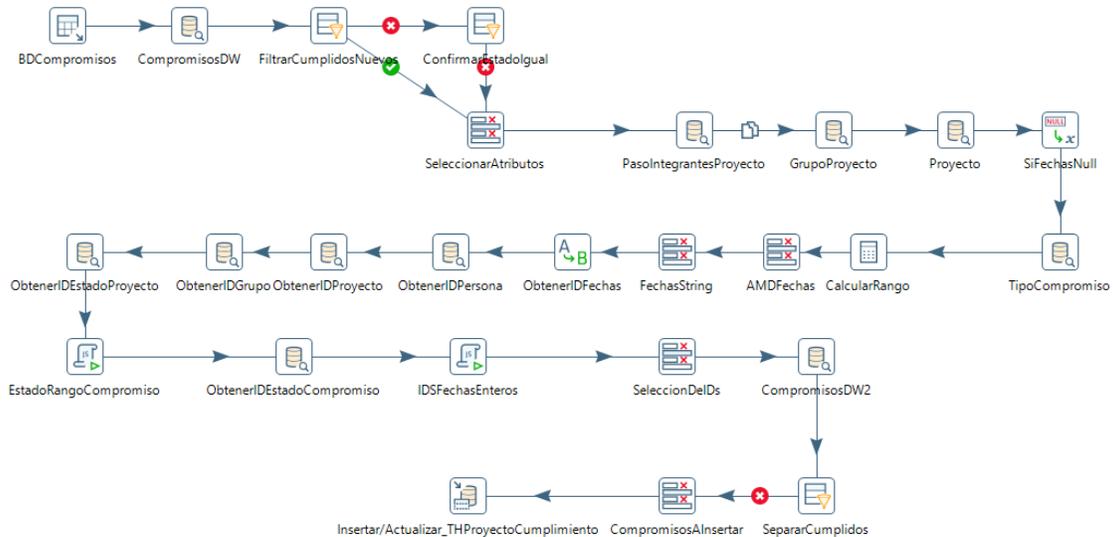
#### 4. Tabla de hechos TH\_ProyectoCumplimiento

Para realizar el proceso ETL de la tabla de hechos *TH\_ProyectoCumplimiento*, es necesario ejecutar una transformación, donde se definen ciertos pasos, lo cual se puede observar en la Ilustración 19.

**Extracción:** los datos se extraen de la fuente VRI, haciendo uso del paso “*BDCompromisos*”. El proceso inicia consultando los compromisos que se tienen registrados en la base de datos, donde se descarta registros ya existentes para insertar únicamente registros nuevos en la DW. Durante la extracción para esta tabla de hecho, se accede a tablas para obtener datos relacionados a *Proyectos*, *grupos*, *integrantes*, *fechas*, *tipo de compromisos*, entre otros, utilizando los pasos “*CompromisosDW*”, “*PasoIntegrantesProyecto*”, “*GrupoProyecto*”, “*Proyecto*”, “*TipoCompromiso*”, “*ObtenerIDPersona*”, “*ObtenerIDProyecto*”, “*ObtenerIDGrupo*”, “*ObtenerIDEstadoProyecto*” y “*ObtenerIDEstadoCompromiso*”.

**Transformación:** en este proceso dentro del flujo, se define el paso “*SiFechasNull*” para convertir las fechas nulas en valor 0, y que así pueda ser almacenado en la DW. También, se realiza cálculo de rangos y asignación de estados de rangos, usando los pasos “*CalcularRango*” y “*ObtenerIDRango*” para los compromisos de los proyectos, pues son significativos a nivel de consulta y análisis por parte del usuario, lo cual se obtiene utilizando la fecha de vencimiento y de actualización para los compromisos.

**Carga:** las columnas del flujo de la transformación se asocian con los registros y los atributos de la tabla de hechos *TH\_ProyectoCumplimiento* de la DW, donde se verifica la existencia del registro para que sea actualizado, descartando los compromisos de estado “*Cumplido*”, usando el paso “*DescartarCumplidos*”, o en su defecto, ingresado como nuevo registro, asignando una nueva llave y cargando sus datos, por medio del paso “*Insertar/Actualizar\_THProyectoCumplimiento*”



**Ilustración 19. Cargue tabla de hechos TH\_ProyectoCumplimiento**

En la Ilustración 19, se observa que los pasos “FiltrarCumplidosNuevos”, “ConfirmarEstadolgual” y “SepararCumplidos” son para realizar filtros de datos por medio de condicionales, presentando seguidamente círculos de color verde y rojo que representan “Verdadero” y “Falso” respectivamente, según corresponda la condición expresada.

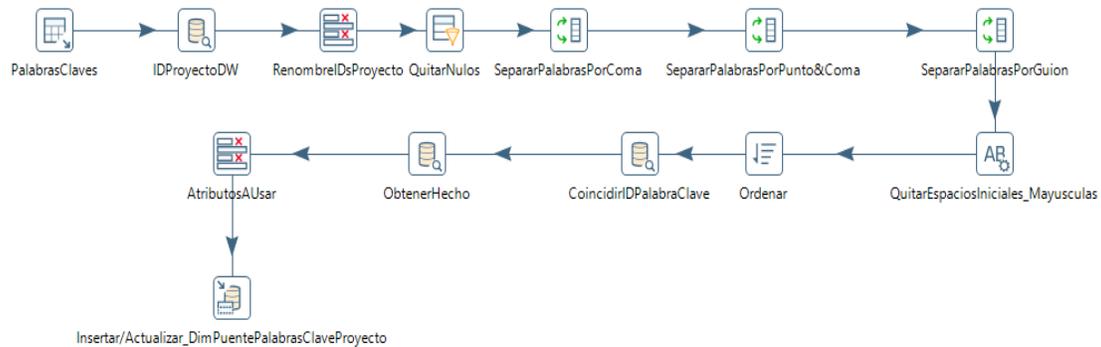
## 5. Tabla puente **PuentePalabrasClave**

En la Ilustración 20 se observa el cargue de la tabla puente *PalabrasClave*, el cual describe el siguiente flujo:

**Extracción:** los datos se extraen de la base de datos VRI, utilizando el paso “Proyecto”. La tabla obtenida contiene un atributo denominado *Palabras clave*, que contiene varios valores separados por coma, guion, punto y coma.

**Transformación:** se realiza un proceso de separación de las diferentes palabras clave asociadas a un proyecto, haciendo uso de los pasos “SepararPorComa”, “SepararPorPunto&Coma” y “SepararPorGuion”, con el fin de tenerlas de manera independiente en el flujo. A estas palabras también se les quita espacios iniciales y se dejan con su primera letra mayúscula, utilizando el paso “QuitarEspaciosIniciales\_Mayusculas”, para evitar duplicidad de palabras. En el paso “AtributosAUsar” se realiza la conversión de los tipos de datos con el objetivo de ejecutar búsquedas para obtener el identificador de la dimensión *Palabra Clave* y de la tabla de hechos *TH\_Proyecto*.

**Cargue:** las columnas del flujo de la transformación se asocian con los registros y los atributos de la tabla puente *PuentePalabrasClave* de la DW, donde se verifica la existencia del registro para que sea actualizado o en su defecto, ingresado como nuevo registro, asignando una nueva llave y cargando sus datos en el paso “Insertar/Actualizar\_DimPuentePalabrasClaveProyecto”.



**Ilustración 20. Cargue tabla puente Palabras Clave**

### 5.1.2 Población y validación de datos

Al terminar el desarrollo de las transformaciones para el cargue de los datos, se definen trabajos con los cuales se busca definir un flujo lógico, para cargar primero las dimensiones, luego las tablas de hecho, y finalmente las tablas puente, para cada uno de los modelos.

La Ilustración 21 muestra la declaración del trabajo para realizar el cargue de dimensiones del modelo de *Proyectos de Investigación*, donde se realiza una transformación por cada una de las dimensiones, las cuales pasan al control de pasos, para finalmente completar de manera exitosa el cargue.

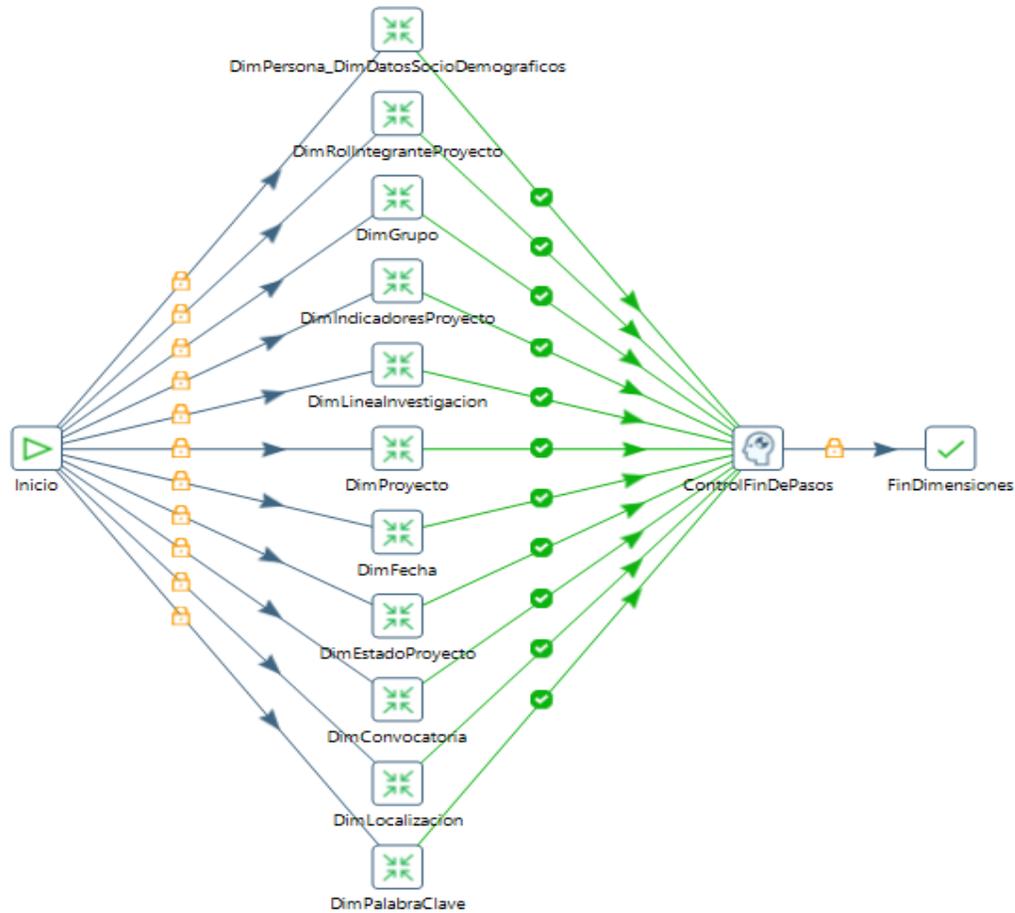
Luego, en la Ilustración 22, se muestra el flujo de trabajo definido para realizar el cargue de los datos de la tabla de hechos y las tablas puente del modelo.

## 5.2 DESARROLLO DEL FRONT ROOM

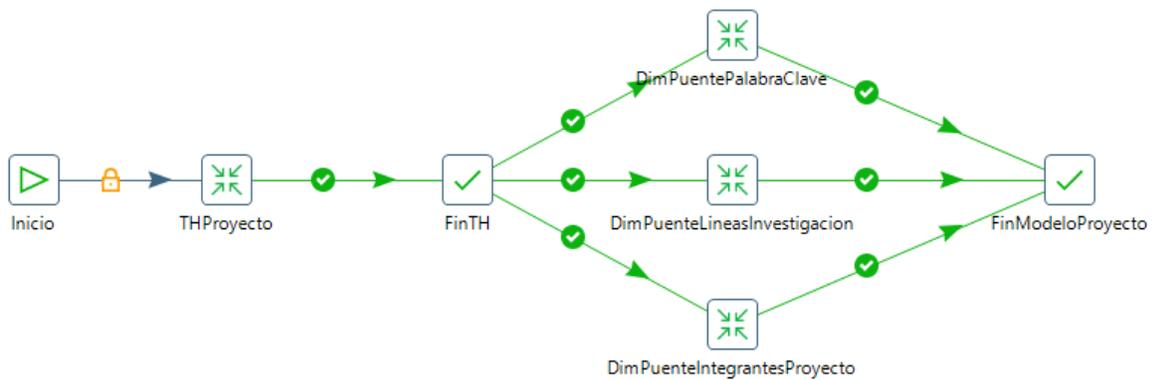
En esta sección se presentan las actividades requeridas para la visualización de los reportes y las consultas para los usuarios finales. Estas actividades son:

### 5.2.1 Identificación y priorización de reportes candidatos

La Tabla 31 presenta los reportes candidatos obtenidos a partir de los requerimientos analíticos.



**Ilustración 21. Trabajo para cargue de dimensiones - Modelo Proyecto**



**Ilustración 22. Trabajo cargue tabla de hecho y tablas puente - Modelo Proyecto**

**Tabla 31. Reportes candidatos. Fuente propia**

Requerimiento analítico	Comentarios	Prioridad	Usuario	Tipo	Título
Proyectos financiados por el sistema general de regalías (SGR)	Cantidad de proyectos que son financiados por el SGR, vinculados por facultad, para un determinado año.	Alta	Encargado SI	AD-Hoc	Cantidad de proyectos financiados por el SGR
Proyectos financiados por Minciencias	Cantidad de proyectos que son financiados por Colciencias, vinculados por facultad, para un determinado año.	Alta	Encargado SI	AD-Hoc	Cantidad de proyectos financiados por Colciencias
Proyectos financiados por otras entidades externas	Cantidad de proyectos que son financiados por entidades externas, vinculados por facultad, para un determinado año.	Alta	Encargado SI	AD-Hoc	Cantidad de proyectos financiados por entidades externas
Entidades que han financiado proyectos de investigación	Cantidad de entidades que han financiado proyectos de investigación, detallando el monto y discriminando por el año de financiación.	Alta	Secretaria	Reporte estándar AD-Hoc	Cantidad de entidades que han financiado proyectos de investigación
Compromisos vencidos de los proyectos de investigación	Cantidad de compromisos vencidos y pendientes por entregar, discriminando el estado del proyecto y el rango de vencimiento del compromiso.	Alta	Encargado SI	AD-Hoc	Cantidad de compromisos vencidos y pendientes por entregar
Compromisos de los proyectos de investigación	Cantidad de compromisos, discriminando su responsable y fecha de vencimiento.	Alta	Encargado SI	Reporte estándar AD-Hoc	Cantidad de compromisos de los proyectos de investigación

Proyectos de investigación de la Universidad	Cantidad de proyectos de investigación que se tienen en la Universidad en un determinado año.	Alta	Encargado SI Jefe de División	Reporte estándar AD-Hoc	Cantidad de proyectos de investigación
Líneas de investigación de los proyectos de investigación en una convocatoria	Cantidad de proyectos de investigación discriminados por su línea de investigación y su convocatoria.	Alta	Secretaria Jefe de División	Reporte estándar AD-Hoc	Cantidad de proyectos por línea de investigación
Palabras clave de los proyectos de investigación por cada facultad de la Universidad	Cantidad de proyectos de investigación discriminados por sus palabras clave y facultad.	Alta	Secretaria Jefe de División	AD-Hoc	Cantidad de proyectos por palabra clave

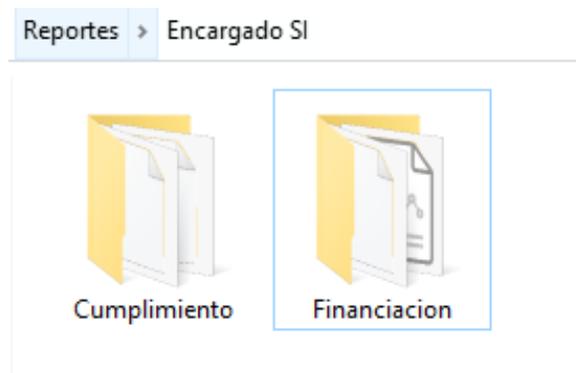
## 5.2.2 Diseño de la estrategia de navegación

En la estrategia de navegación se define el agrupamiento de los reportes y consultas Ad-Hoc para que los usuarios puedan acceder a ellas de una forma organizada y sencilla. Se crean áreas de trabajo para cada uno de los usuarios. La carpeta *Jefe* para el uso de la Jefe de la División, la carpeta *Secretaría* por la secretaria de la División, y la carpeta *Encargado SI* para el encargado del sistema de información de la División.

### 5.2.2.1 Reportes estándar

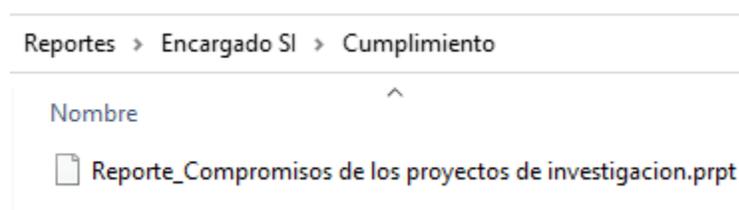
Se crea una carpeta para almacenar los reportes estándar para el encargado del sistema de información, otra para los reportes estándar para la jefe y otra carpeta para los reportes estándar de la secretaria. Internamente, cada carpeta está compuesta por los temas analíticos requeridos por los funcionarios, y dentro de esta encontrará los reportes requeridos.

La Ilustración 23, muestra las carpetas que contienen los reportes estándar que son requeridos por el *Encargado del sistema de información*.



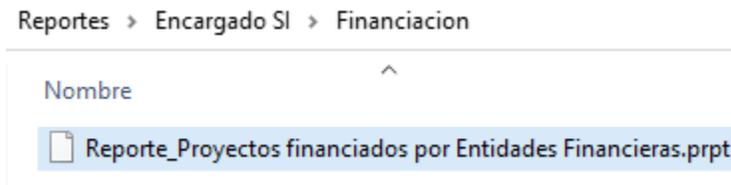
**Ilustración 23. Carpetas de reportes listadas para el encargado del SI**

Al seleccionar la carpeta *Cumplimiento*, se muestran los reportes estándar correspondientes a dicho tema analítico, como lo muestra la Ilustración 24.



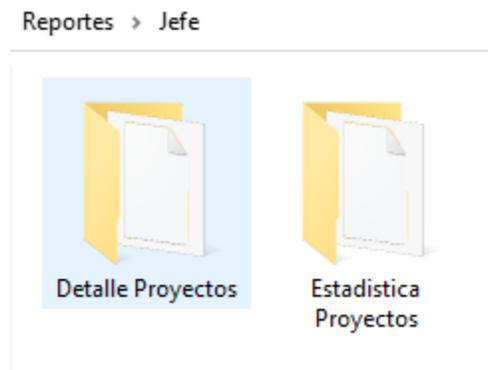
**Ilustración 24. Reportes sobre Cumplimiento para el encargado del SI**

Al seleccionar la carpeta *Financiación*, se muestran los reportes estándar de ese tema, como la presenta la Ilustración 25.



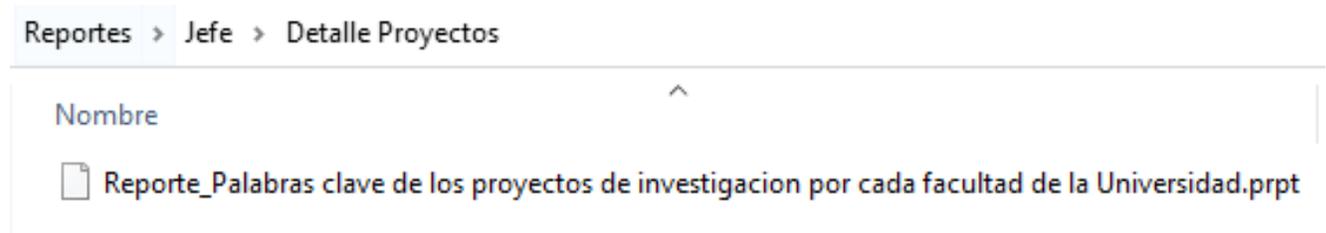
**Ilustración 25. Reportes sobre Financiación para el encargado del SI**

La Ilustración 26, muestra las carpetas que contienen los reportes estándar que son requeridos por la *Jefe*.



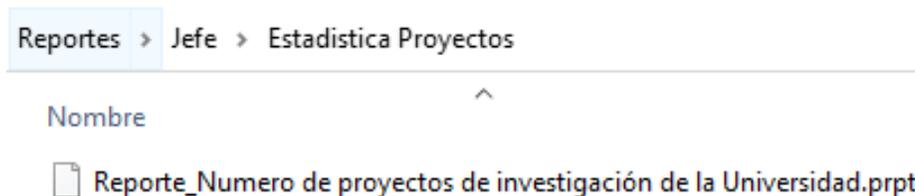
**Ilustración 26. Carpetas de reportes listadas para la jefe**

Al seleccionar la carpeta *Detalle Proyectos*, se muestran los reportes estándar correspondientes a dicho tema analítico, como lo muestra la Ilustración 27.



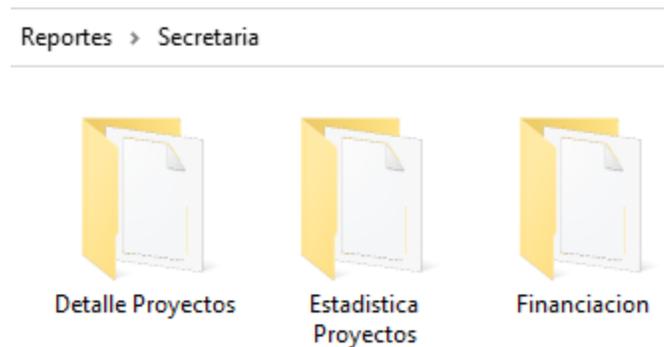
**Ilustración 27. Reportes sobre Detalle de Proyectos para la jefe**

Al seleccionar la carpeta *Estadística de Proyectos*, se muestran los reportes estándar de ese tema, como la presenta la Ilustración 28.



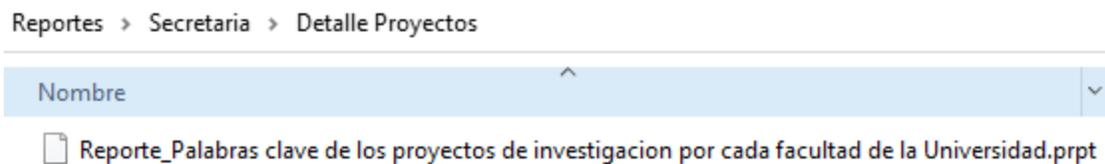
**Ilustración 28. Reportes sobre Estadística de Proyectos para la jefe**

La Ilustración 29, muestra las carpetas que contienen los reportes estándar que son requeridos por la *Secretaria*.



**Ilustración 29. Carpetas de reportes listadas para la secretaria**

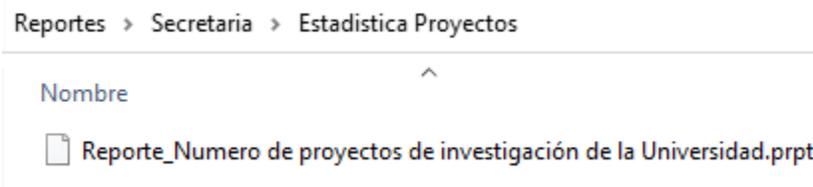
Al seleccionar la carpeta *Detalle de Proyectos*, se muestran los reportes estándar correspondientes a dicho tema analítico, como lo muestra la Ilustración 30.



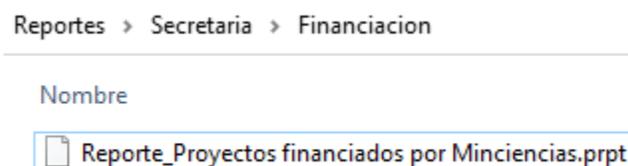
**Ilustración 30. Reportes sobre Detalle de Proyectos para la secretaria**

Al seleccionar la carpeta *Estadística de Proyectos*, se muestran los reportes estándar de ese tema, como la presenta la

Ilustración 31, y al seleccionar la carpeta *Financiación*, se muestran los reportes estándar de ese tema, como la presenta la Ilustración 32.



**Ilustración 31. Reportes sobre Estadística de Proyectos para la secretaria**



**Ilustración 32. Reportes sobre Financiación para la secretaria**

### 5.2.2.2 Consultas Ad-Hoc

Para la visualización de las consultas Ad-hoc, se hace uso de la aplicación web de la herramienta POWER BI. En este caso se tiene el mismo orden de carpetas especificados para reportes estándar (Encargado SI, Jefe y Secretaria), pero relacionado en un área de trabajo, como lo muestra la Ilustración 33.

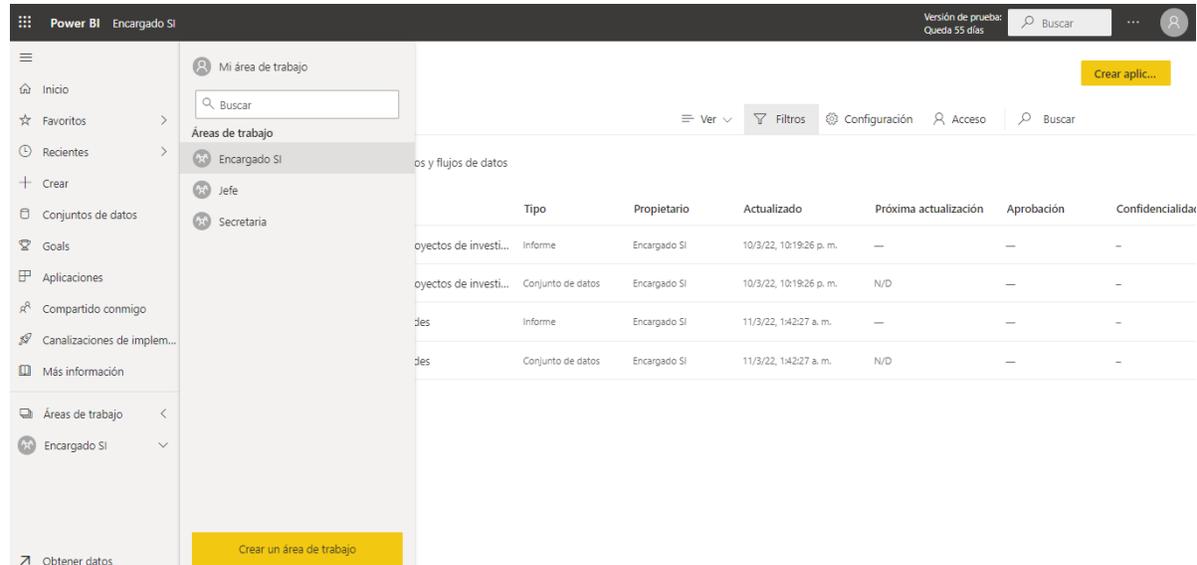


Ilustración 33. Área de trabajo de usuarios. Fuente propia

Al seleccionar área de trabajo *Encargado SI*, se visualizan las consultas de la Ilustración 34.

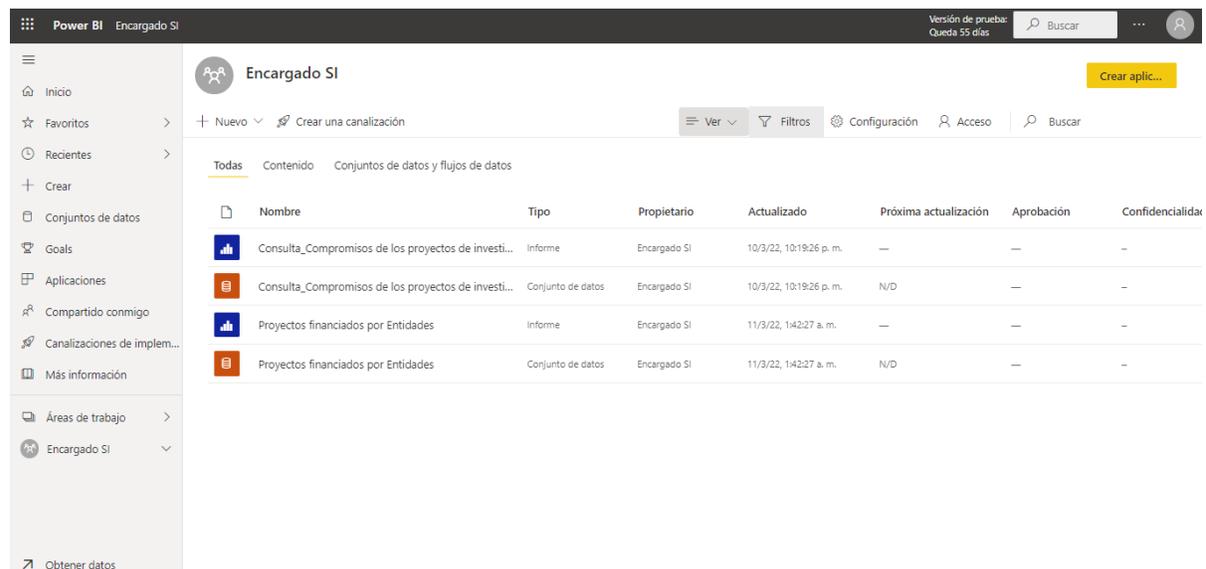


Ilustración 34. Consultas Encargado SI. Fuente propia.

El área de trabajo *Secretaria*, contiene las consultas de la Ilustración 35.

Power BI Secretaria

Secretaria

+ Nuevo  Crear una canalización

Ver  Filtros Configuración Acceso Buscar

Todas	Contenido	Conjuntos de datos y flujos de datos		Nombre	Tipo	Propietario	Actualizado	Próxima actualización	Aprobación	Confidencialidad
				CConsulta_Lineas de investigación de los proyectos...	Conjunto de datos	Secretaria	11/3/22, 11:17:29 a. m.	N/D	—	—
				Consulta_Compromisos de los proyectos de investi...	Informe	Secretaria	11/3/22, 12:24:5 a. m.	—	—	—
				Consulta_Compromisos de los proyectos de investi...	Conjunto de datos	Secretaria	11/3/22, 12:24:5 a. m.	N/D	—	—
				Consulta_Numero de proyectos de investigación d...	Informe	Secretaria	11/3/22, 12:6:58 a. m.	—	—	—
				Consulta_Numero de proyectos de investigación d...	Conjunto de datos	Secretaria	11/3/22, 12:6:58 a. m.	N/D	—	—
				Consulta_Palabras clave de los proyectos de investi...	Informe	Secretaria	11/3/22, 12:37:01 a. m.	—	—	—
				Consulta_Palabras clave de los proyectos de investi...	Conjunto de datos	Secretaria	11/3/22, 12:37:01 a. m.	N/D	—	—
				Proyectos financiados por Entidades	Informe	Secretaria	11/3/22, 14:31:1 a. m.	—	—	—
				Proyectos financiados por Entidades	Conjunto de datos	Secretaria	11/3/22, 14:31:1 a. m.	N/D	—	—

**Ilustración 35. Consultas Secretaria. Fuente propia**

El área de trabajo *Jefe*, contiene las consultas de la Ilustración 36.

Power BI Jefe

Jefe

+ Nuevo  Crear una canalización

Ver  Filtros Configuración Acceso Buscar

Todas	Contenido	Conjuntos de datos y flujos de datos		Nombre	Tipo	Propietario	Actualizado	Próxima actualización	Aprobación	Confidencialidad
				CConsulta_Lineas de investigación de los proyectos...	Conjunto de datos	Jefe	11/3/22, 11:16:59 a. m.	N/D	—	—
				Consulta_Compromisos de los proyectos de investi...	Informe	Jefe	11/3/22, 12:22:1 a. m.	—	—	—
				Consulta_Compromisos de los proyectos de investi...	Conjunto de datos	Jefe	11/3/22, 12:22:1 a. m.	N/D	—	—
				Consulta_Numero de proyectos de investigación d...	Informe	Jefe	11/3/22, 12:6:35 a. m.	—	—	—
				Consulta_Numero de proyectos de investigación d...	Conjunto de datos	Jefe	11/3/22, 12:6:35 a. m.	N/D	—	—
				Consulta_Palabras clave de los proyectos de investi...	Informe	Jefe	11/3/22, 12:36:40 a. m.	—	—	—
				Consulta_Palabras clave de los proyectos de investi...	Conjunto de datos	Jefe	11/3/22, 12:36:40 a. m.	N/D	—	—
				Proyectos financiados por Entidades	Informe	Jefe	11/3/22, 14:24:9 a. m.	—	—	—
				Proyectos financiados por Entidades	Conjunto de datos	Jefe	11/3/22, 14:24:9 a. m.	N/D	—	—

**Ilustración 36. Consultas Jefe. Fuente propia**

Las consultas Ad-hoc se realizan desde el portal web de Power BI, donde los usuarios pueden identificar la información de dimensiones y medidas que pueden ser usadas para estas consultas según sea la necesidad del usuario, cómo lo muestran los ejemplos de la Ilustración 37 y la Ilustración 38, para la información tanto en tabla cómo en gráfico, respectivamente.

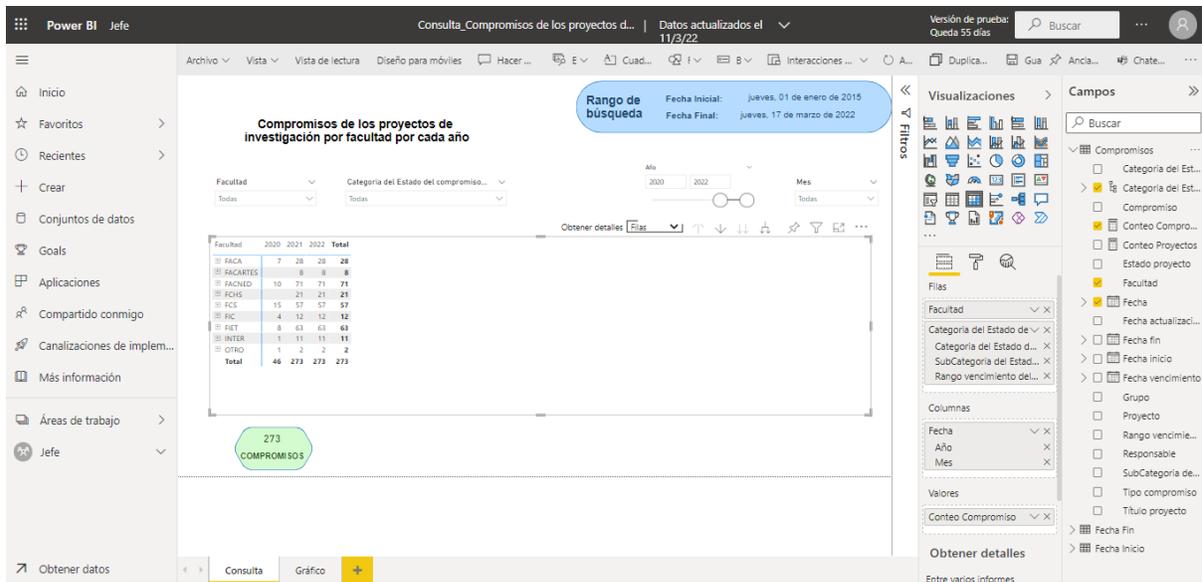


Ilustración 37. Tabla de consulta Ad-hoc en Power BI

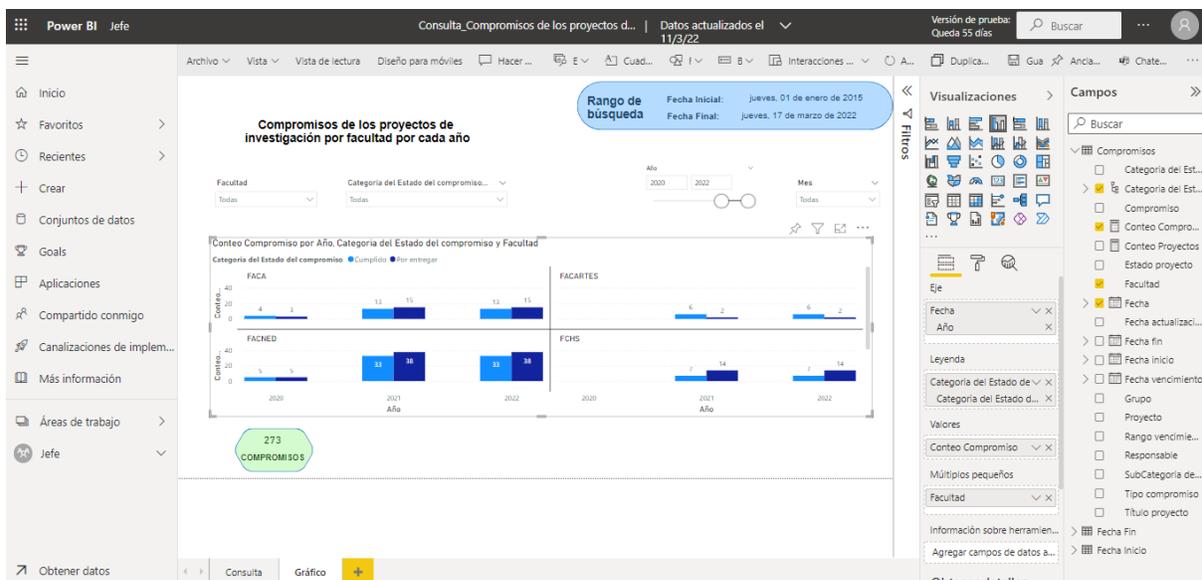


Ilustración 38. Gráfico de consulta Ad-hoc en Power BI

### 5.2.3 Formato y nombrado de los reportes

El formato en los reportes brinda la opción de definir sus elementos y distribución, para que sean de fácil interpretación por parte de los usuarios. En la Ilustración 39 se indica el formato general definido para los reportes estándar, compuesto por:

- **Encabezado de página:** contiene el logo de la Vicerrectoría Universidad del Cauca seguido del nombre de la División de Gestión de la Investigación, el título se define

según la solicitud realizada para dicho reporte y el total de la cantidad requerida para el mismo.

- **Encabezado del reporte:** contiene el nombre las columnas asociadas a la información a mostrar.
- **Detalles:** es el cuerpo del reporte que contiene filas con la información solicitada en el reporte.

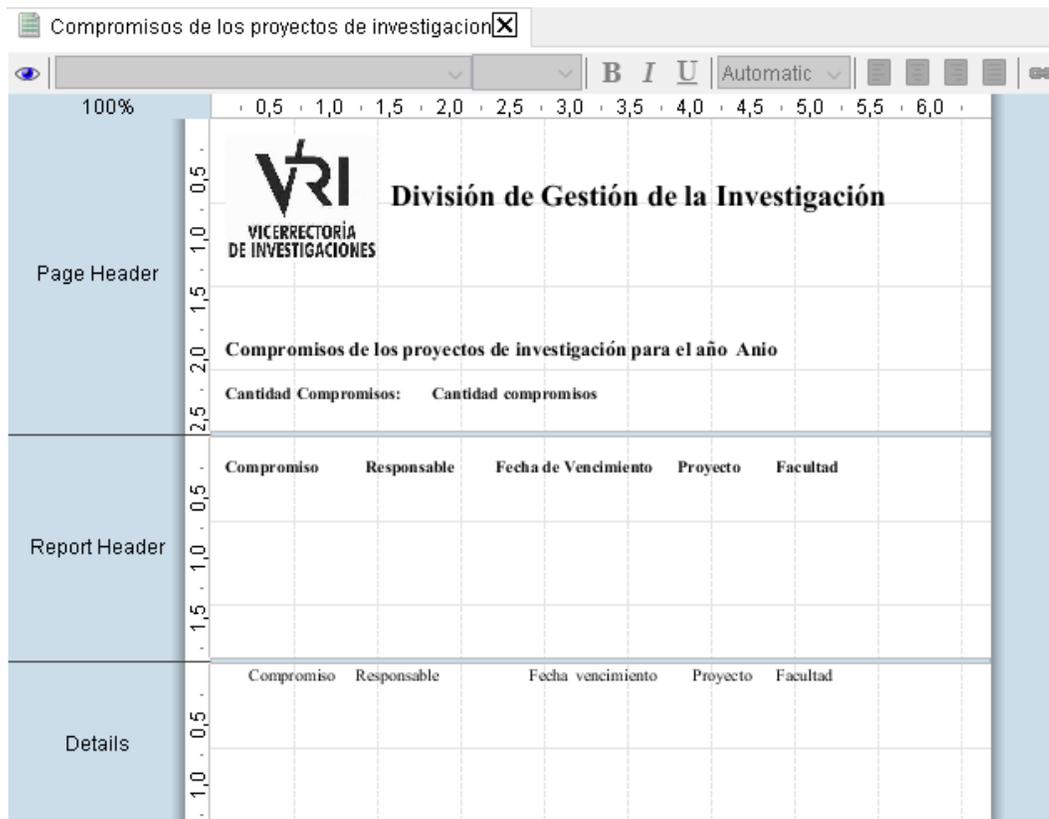


Ilustración 39. Formato general reportes estándar. Fuente propia

## 5.2.4 Selección de un enfoque de implementación

Se selecciona una herramienta de escritorio, *Pentaho Report Designer* (PRD), para el manejo de reportes estándar, y para el acceso a las consultas se utiliza la herramienta Power BI, debido a que es fácil e intuitiva para el uso de los usuarios.

## 5.2.5 Desarrollo de la aplicación de usuario final

En este apartado se presentan los pasos utilizados para realizar la aplicación.

### 5.2.5.1 Creación de reportes estándar

La creación de reportes estándar comienza creando un proyecto nuevo en *PRD*, estableciendo la conexión con la DW y definiendo una consulta para obtener la información necesaria para satisfacer la solicitud del reporte. Al finalizar este, se le da un formato según lo descrito en la Sección 5.2.3.

### 5.2.5.2 Creación de consultas Ad-Hoc

La creación de consultas Ad-hoc inicia creando un nuevo informe en *POWER BI*, estableciendo la conexión con la DW y definiendo una consulta SQL para obtener la información necesaria para satisfacer la solicitud de la consulta Ad-hoc. Al finalizar este procedimiento, la herramienta permite al usuario añadir en la pestaña “Campos” las dimensiones, atributos, jerarquías y medidas a utilizar. En la pestaña “Visualizaciones”, se escoge la forma en que se desea ver la información en el área del trabajo para la consulta, tales como tablas, matrices, gráficos, etc. Además, se permite agregar y utilizar segmentadores y filtros para los datos, según los requerimientos del usuario. Lo descrito se presenta en la Ilustración 40.

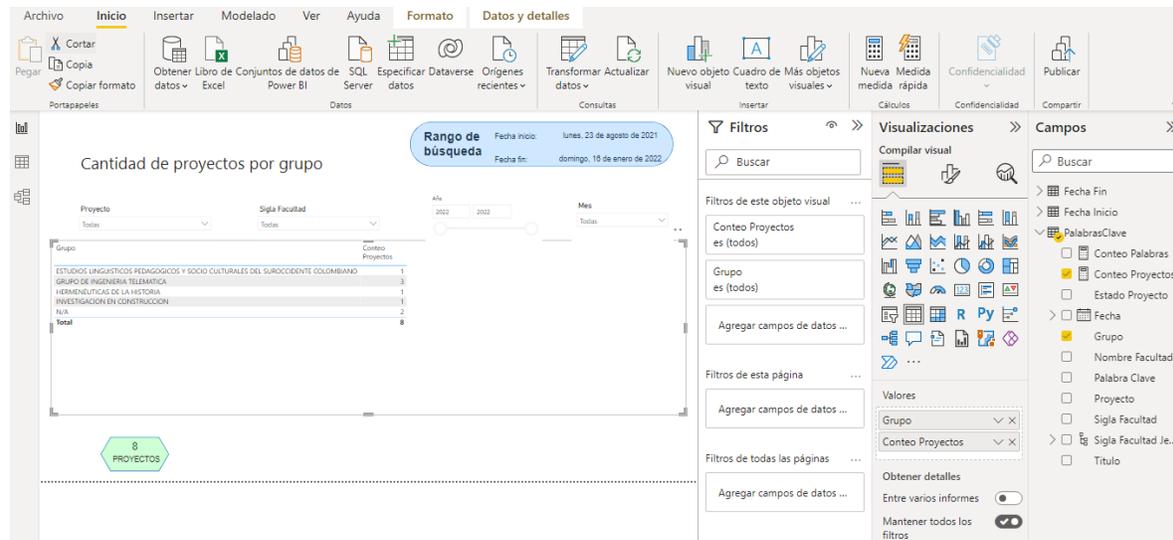


Ilustración 40. Área de trabajo consultas Ad-Hoc

## 5.3 INTEGRACIÓN

En esta subfase se pretende obtener un prototipo de DW completo, mediante la integración del Back Room y Front Room, para lo cual se desarrollan las siguientes actividades:

### 5.3.1 Procesar el cubo

Haciendo uso de *Pentaho Schema Workbench* se realiza la definición del cubo mediante un esquema (ver Ilustración 41), el cual es un contenedor de cubos, donde los datos son obtenidos de un sistema de gestión de bases de datos relacionales *MySQL*, por esto en primer lugar se define la conexión entre estos (ver Ilustración 42).

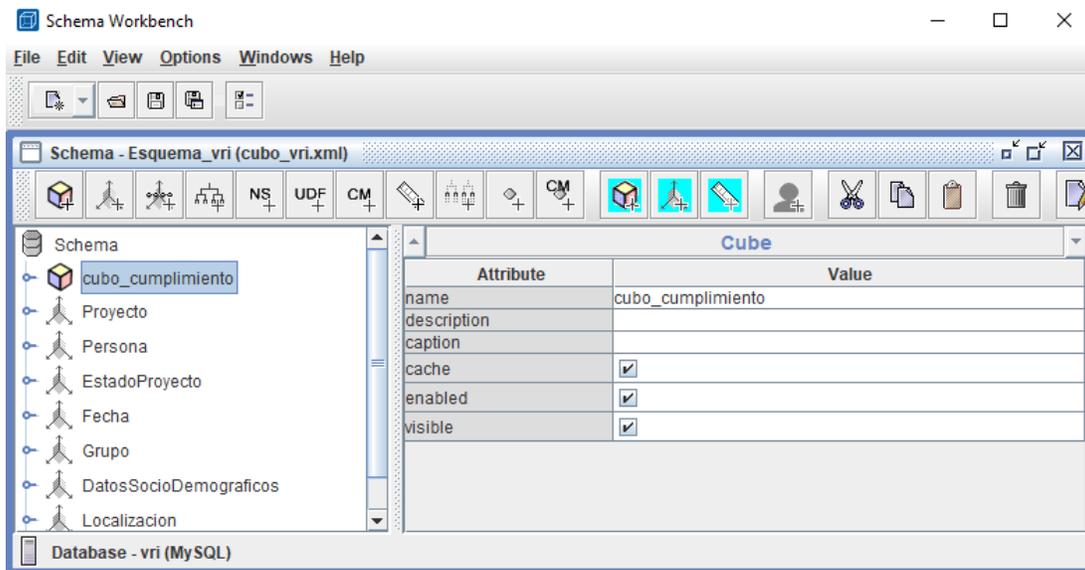


Ilustración 41. Esquema en Schema Workbench

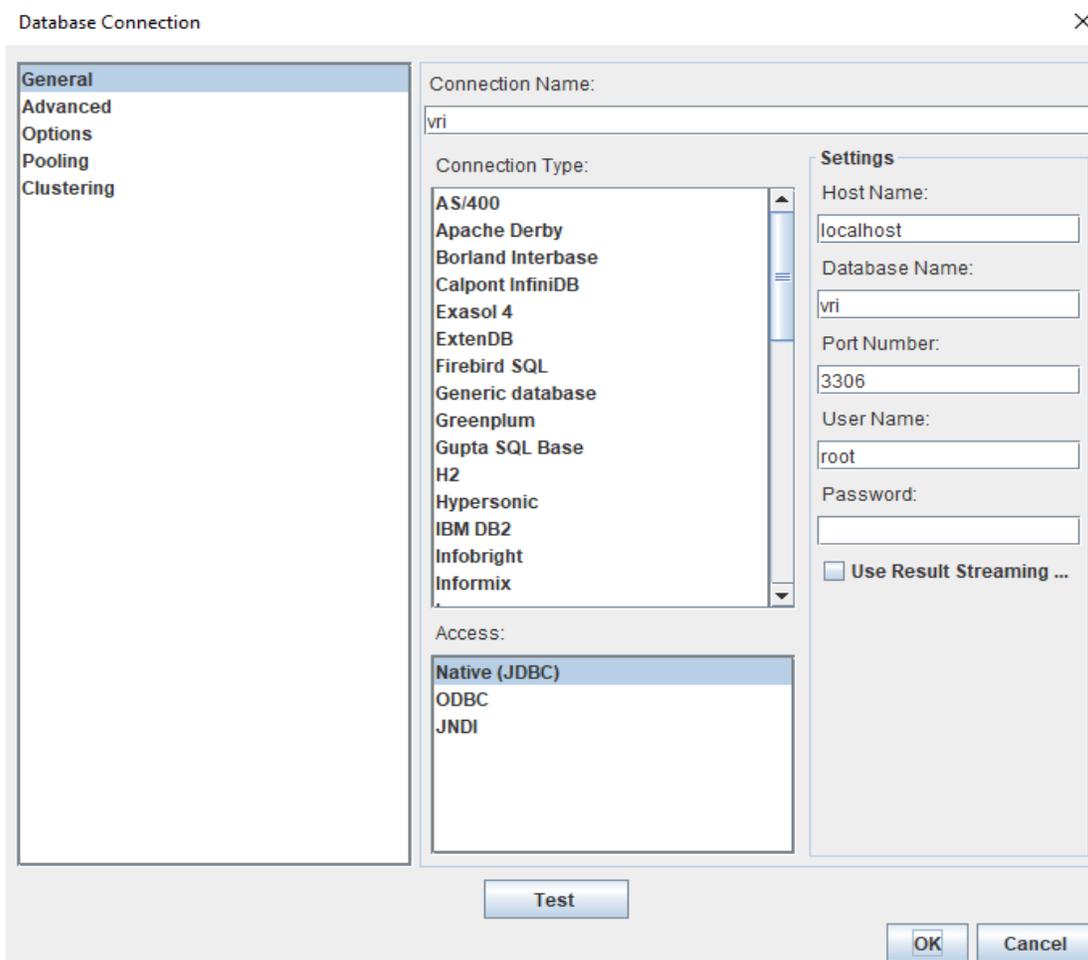


Ilustración 42. Conexión del esquema con el motor relacional

Una vez definido el cubo se obtiene un fichero XML, que representa los esquemas dimensionales del cubo, que puede ser procesado por un visor analítico, en este caso, *La Azada* (ver Ilustración 43)

FechaActualizacion_Aprobado.FechaCompleta()	-Estado	Aprobado	En Ejecución	Formulado	No aprobado	Revisión VRI	Suspendido	Terminado	Terminado con pendiente	
All Fecha.FechaCompleta()		1.588	9	219	135	69	56	39	728	333
+2000		37		15	17					1
+2001		31		24	7					
+2002		27		17	10					
+2003		42		23	19					
+2004		27		17	10					
+2005		31		18	13					
+2006		31		18	13					
+2007		31		22	9					
+2008		28		18	10					
+2009		40		29	11					
+2010		34		18	16					
+2015		3				1			1	1
+2022		1.226	9			68	56	38	724	331

Ilustración 43. Procesamiento del cubo en visor analítico La Azada

Para realizar el procesamiento del cubo, es necesario definir una conexión con la base de datos relacional y el esquema de datos del XML, como se observa en la Ilustración 44.

**Gestor de fuentes de datos (Datasources)**

Esta pantalla le permite configurar sus fuentes de datos.

**Fuentes de datos**

- MSAS
- Scenario
- SmallFoodMart\_Derby
- XMLA FoodMart
- prueba
- ventasWilliam
- vri**

**Datos generales**

Nombre: vri

Descripción:

**Información sobre la conexión**

Tipo de conexión:  Mondrian  XMLA

Base de datos: MySQL

Servidor: Parametros de conexión

URL: jdbc:mysql://localhost/vri

Usuario: vri

Password: ●●●

**Esquema de datos**

	Esquema
-	file:catalogs/cubo_vri.xml
+	

Terminar Cancelar

Ilustración 44. Conexión base de datos relacional y esquema de datos

### 5.3.2 Conexión a la DW con las herramientas de reportes

La conexión a la DW es establecida con el fin de publicar los reportes en un servidor de informes para que sean visualizados mediante un portal web o para que sean alojados de manera local accediendo a ellos a través de una interfaz de usuario de una aplicación. Estas dos conexiones se definen en las herramientas *POWER BI* y *Pentaho Report Designer*, como lo muestran la Ilustración 45 y la Ilustración 46, respectivamente.

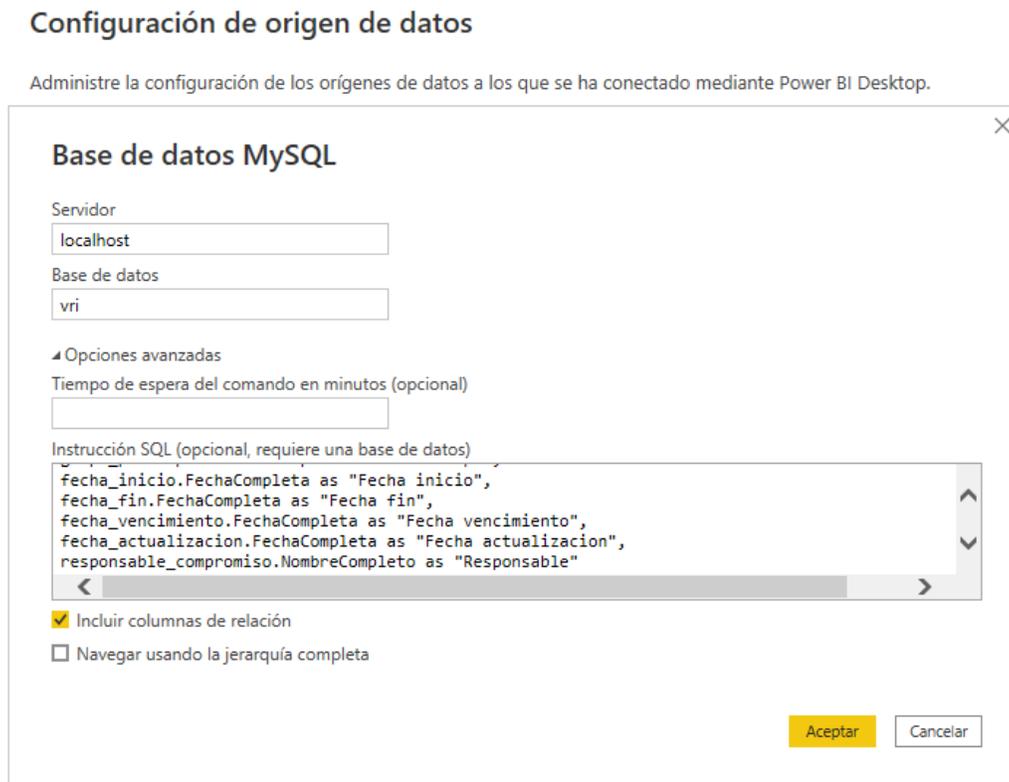


Ilustración 45. Conexión a la DW con POWER BI

### 5.3.3 Pruebas sobre el prototipo

Con el fin de identificar problemas en el prototipo, se ejecutaron consultas sobre la DW comparando los resultados con el sistema fuente y la DW, de forma manual, para verificar la consistencia de los datos.

### 5.3.4 Presentación del prototipo a los usuarios

El prototipo es presentado en uno de los ordenadores del equipo de desarrollo, iniciando con una breve inducción sobre la generación de consultas Ad-hoc en la herramienta seleccionada para tal fin, y la forma en la que se accede a los reportes estándar al igual que la visualización de los mismos, esto con el fin de que los usuarios hicieran observaciones para mejorar el prototipo y determinar la necesidad de realizar una nueva iteración sobre subfases de Back Room y Front Room para dar solución a errores o cambios pertinentes (no se presentaron correcciones).

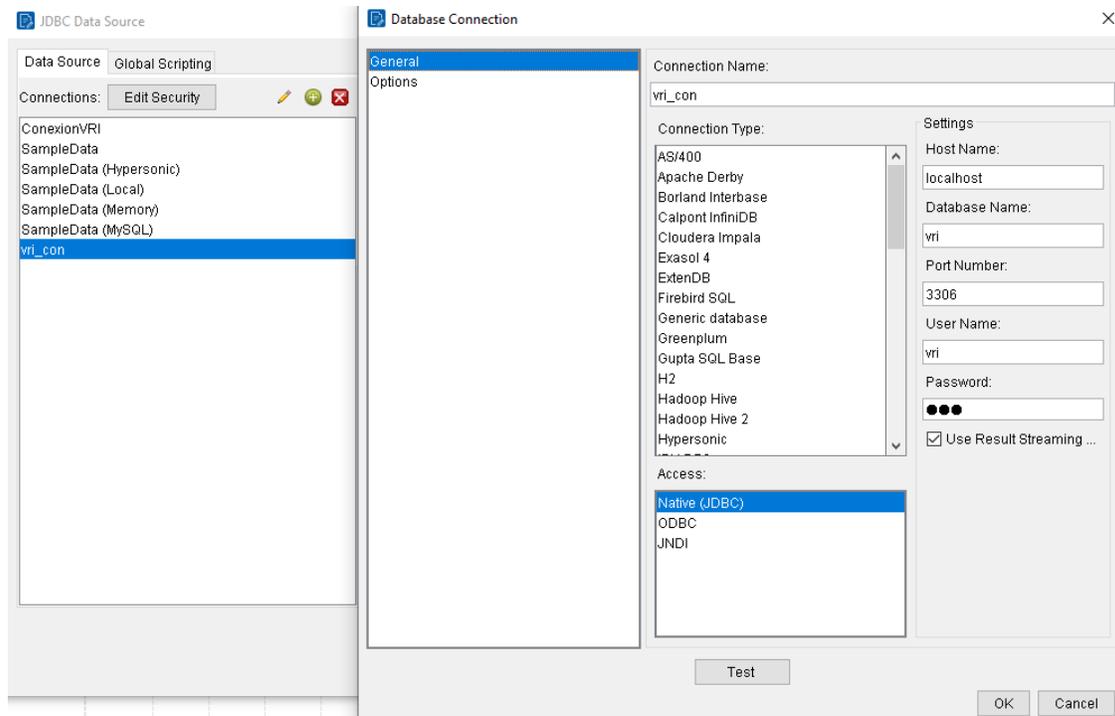


Ilustración 46. Conexión a la DW con Pentaho Report Designer

### 5.3.5 Proceso de aceptación de usuario

Se realiza por medio del test de aceptación, elaborado con base en la característica de satisfacción, *Modelo para la calidad en uso* de la norma ISO/IEC 25010, este proceso se describe en el Capítulo 6.

## 5.4 DESPLIEGUE

Esta subfase no se realizó debido a que el prototipo no cuenta con la totalidad de los datos reales, lo cual implica que los usuarios no puedan realizar un análisis real de la información. Además, no es posible acceder a un servidor para poner en producción el prototipo creado. Las actividades que se realizaron fueron:

- **Diseño de una estrategia de capacitación de usuarios:** explicación del objetivo del modelo y creación de la DW, de la información almacenada en el prototipo y los posibles análisis a generar. Para las consultas Ad-hoc, se realiza una descripción de las medidas y las dimensiones que se pueden utilizar. Finalmente se realiza una introducción sobre la visualización de reportes estándar, el acceso a los mismos, la organización de sus carpetas, los parámetros de cada reporte y la organización estructural de los mismos.
- **Capacitación de los usuarios:** de acuerdo con lo planteado en la actividad anterior, se realiza en el mismo espacio programado para la ejecución de la evaluación de satisfacción (ver resultados en el Capítulo 6), debido a la poca disponibilidad de tiempo de los usuarios.

## Capítulo 6

---

### 6 EVALUACIÓN DEL PROTOTIPO

En este capítulo se describe el proceso de evaluación del prototipo de bodegas de datos realizado para medir el nivel de satisfacción de los usuarios finales con respecto a los reportes estándar y las consultas Ad-hoc, mediante la aplicación de un test de nivel de satisfacción.

Primero se realiza una descripción acerca del objetivo de la evaluación, seguido de la normativa, descripción y aplicación del test; luego se describe el proceso de evaluación y se presentan los resultados obtenidos.

#### 6.1 OBJETIVO DE LA EVALUACIÓN

El objetivo de la evaluación consiste en comprobar la utilidad de disponer de un sistema DW dentro de la División de Gestión de la Investigación de la Universidad del Cauca, brindando una vista de los beneficios que aporta esta clase de sistemas para los usuarios en las dependencias de proyectos investigación de las universidades que cuenten con ellas.

#### 6.2 NORMATIVA PARA EL NIVEL DE SATISFACCIÓN DE USUARIO

La norma ISO/IEC 25010, para evaluación interna y externa de calidad de un software, trata en su mayoría características orientadas al código fuente y puesta en producción para un producto [28]. Dado que para el prototipo de bodegas de datos no fue necesario escribir código fuente y que no fue posible realizar el despliegue del prototipo, su evaluación se centra en la *calidad en uso*, haciendo énfasis a consultas.

De las características referentes a la *calidad en uso* de la norma ISO/IEC 25010, se elige la característica de *satisfacción con Utilidad* como subcaracterística, donde su métrica de *Nivel de satisfacción* permite evaluar el nivel de satisfacción que tiene el usuario con respecto al prototipo de bodega de datos.

Dado que, durante la revisión de metodologías de DW, no se encontró alguna norma o test que permita evaluar la utilidad de las consultas analíticas y/o reportes de una DW, se realiza una revisión de las normas ISO/IEC 25000 [28], con el fin de identificar una medida que sirva de base para la evaluación.

Dentro de los cinco grupos en que se divide las normas de ISO/IEC 25000, se encuentra el grupo de *Medición de calidad* donde pertenece la norma ISO/IEC 25022 *Medidas de calidad de uso*, la cual permite evaluar la calidad en uso del producto.

La característica que nos permite evaluar el prototipo, en cuanto a métricas calidad de uso, es la *Satisfacción* que cuenta con *Utilidad* como única subcaracterística (ver Tabla 32), la cual nos permite determinar el *nivel de satisfacción* en que se encuentra un usuario al momento de hacer uso del producto software.

Aunque la *Utilidad* cuenta con diferentes métricas, solo se considera el *nivel de satisfacción* para evaluar el prototipo, ya que para hacer uso de las demás métricas es necesario que el prototipo ya se encuentre en etapa de producción [28].

**Tabla 32. Métricas para la característica de calidad Satisfacción. Tomado de [28]**

Métricas para la característica de calidad Satisfacción	
Subcaracterística	Métrica
<i>Utilidad</i>	<ul style="list-style-type: none"> <li>• <b>Nivel de satisfacción.</b></li> <li>• Uso discrecional de las funciones.</li> <li>• Porcentajes de quejas de los clientes.</li> </ul>

### 6.3 IMPLEMENTACIÓN DEL TEST DE NIVEL DE SATISFACCIÓN DE USUARIO

Mediante un cuestionario o test se realiza la evaluación con respecto a la *Utilidad*, según lo sugiere [28]. Se definen tres cuestionarios de satisfacción de usuario para los funcionarios de la División de Gestión de la Investigación en relación con los *proyectos de investigación*, estos son: para el jefe, el encargado del SI y la secretaria.

Los cuestionarios están compuestos cuatro secciones:

- **Primera:** se indica cual es el objetivo de la evaluación seguido a una breve descripción de la información que se almacena en el prototipo.
- **Segunda:** evaluación de los reportes estándar que son presentados a los usuarios, y que fueron elaborados acuerdo a las solicitudes hechas por los mismos. El planteamiento de las preguntas busca evaluar el nivel analítico que brindan los reportes al igual que su facilidad de interpretación en cuanto al contenido. La evaluación de los reportes se realiza en una tabla, donde cada fila presenta una pregunta, la cual debe ser calificada por el usuario según la escala determinada [29].
- **Tercera:** De manera similar a los reportes estándar, se realiza la evaluación de las consultas analíticas Ad-hoc, donde las preguntas buscan evaluar la facilidad de navegación en los datos, interpretación del contenido y gráficos presentados.
- **Cuarta:** Al finalizar, el usuario hace constar la realización de la evaluación registrando su firma.

Los test de evaluación se encuentran especificados en el Anexo D.

Para cada una de las preguntas se usa la escala de calificación de Likert [29] (Ver Tabla 33), ya que permite establecer el nivel de conformidad de quien es encuestado y es la escala más empleada en cuestionarios y encuestas para la investigación.

**Tabla 33. Calificación de los cuestionarios según la escala de Likert. Tomado de [29]**

Valor numérico	Conformidad
5	Totalmente de acuerdo
4	Parcialmente de acuerdo
3	Indiferente
2	Algo en desacuerdo
1	Totalmente en desacuerdo

## **6.4 APLICACIÓN DEL TEST DE NIVEL DE SATISFACCIÓN**

Este test es aplicado a las siguientes funcionarias de la División de Gestión de la Investigación:

- Adriana Milena Hurtado, Secretaria.
- Lina María López, Jefe de División.
- Elizabeth González, Equipo Apoyo al Investigador.
- Susana Sarria, Secretaria.

Debido a la poca disponibilidad de tiempo de los funcionarios de la División, se realiza una sesión de socialización sobre los modelos dimensionales y prototipo de DW creados, reportes estándar y consultas Ad-hoc definidos. Posteriormente, se hace capacitación sobre el manejo de las herramientas de reportes para que los usuarios comprendan el proceso de creación y puedan después diseñarlos de acuerdo con sus necesidades. Finalmente se realiza el proceso de evaluación por parte de los funcionarios, sobre los reportes y consultas definidos por los desarrolladores.

Los reportes estándar y consultas Ah-hoc evaluadas por cada una de las funcionarias, son los siguientes:

- Reportes estándar.
  - Compromisos de los proyectos de investigación.
  - Líneas de investigación de los proyectos de investigación.
  - Proyectos de investigación de la Universidad.
  - Proyectos financiados por Entidades Financieras.
- Consulta Ad-hoc.
  - Proyectos financiados por Entidad Financiera por cada año.
  - Compromisos de los proyectos de investigación por facultad por cada año.
  - Proyectos de investigación de la Universidad por facultad por cada año.
  - Líneas de investigación de los proyectos de investigación por facultad y convocatoria.
  - Palabras clave de los proyectos de investigación por cada facultad de la Universidad.

## **6.5 PROCESO DE EVALUACIÓN**

Para responder a la pregunta “¿Qué tan satisfecho está el usuario?”, se hace uso de la métrica de calidad de *Nivel de satisfacción*, definida en [28]. Para medir esta métrica se utiliza la siguiente fórmula:

Donde:

$$X = A/B \quad \text{Ecuación 1}$$

*A*: Representa el número de preguntas con respuestas satisfactorias.

*B*: Representa la cantidad de preguntas realizadas en el cuestionario aplicado,  $B > 0$ .

El valor deseado de *X* está en el rango:  $0 \leq X \leq 1$ , donde 1 es el mayor valor de satisfacción.

Para tener una interpretación más clara del resultado de  $X$  se utiliza una escala de medición de 0 a 100%, donde el valor esperado para los resultados de la evaluación es 100%. En la Tabla 34 se muestra la clasificación de escalas, con su nivel de puntuación y su respectivo grado de satisfacción.

**Tabla 34. Grado de satisfacción según niveles de puntuación. Tomada de [28].**

Escala de medición	Niveles de puntuación	Grado de satisfacción
80,75% – 100%	Cumple con los requisitos	Muy satisfactorio
50,00% - 80,74%	Aceptable	Satisfactorio
20,75% - 40,90%	Mínimamente aceptable	Insatisfactorio
0,00% - 20,74%	Inaceptable	

## 6.6 RESULTADOS OBTENIDOS

El resultado de la evaluación de satisfacción se basa en el número de respuestas del funcionario en cada valor numérico de la escala de Likert. En la Tabla 35 se presenta por cada funcionario el total de preguntas evaluadas y la cantidad de preguntas por cada calificación dada.

**Tabla 35. Cantidad de preguntas evaluadas en los tests. Fuente propia.**

Funcionario	Cantidad de preguntas	Calificación				
		5	4	3	2	1
Adriana Hurtado	32	32				
Lina María López	32	32				
Elizabeth González	32	27	1	4		
Susana Sarria	32	32				
Total	128	123	1	4		

De acuerdo con [28] para realizar el análisis de calidad del producto software se utiliza una matriz de calidad, la cual permite evaluar el producto de manera completa y concisa. Se adaptó la matriz al propósito de evaluación con la característica elegida denominada *Satisfacción* y la subcaracterística *Utilidad*, no incluyendo las columnas nivel de importancia y la ponderación, debido a que solo se incluye una característica y los valores en estas columnas no alterarían los resultados finales.

Mediante la matriz calidad, descrita en [28] se realiza el análisis de calidad del producto software que permite evaluar el producto de una forma completa y concisa.

En la Tabla 36 se presenta una adaptación de la matriz de calidad donde se muestran las fórmulas empleadas en la evaluación de la métrica de *Nivel de Satisfacción*. Debido a que las respuestas de las funcionarias son *Totalmente de acuerdo*, con calificación 5, *Parcialmente de acuerdo*, con calificación 4 e *Indiferente*, con calificación 3, se definen las fórmulas para medir la proporción de estos valores en la escala de Likert, exceptuando las respuestas de calificación 3, debido a que es *Indiferente*, por lo tanto no es relevante para la evaluación de la métrica. El valor  $X1$ , es la proporción de la cantidad de preguntas que son calificadas con 5 ( $A1$ ) sobre el número total de preguntas aplicadas en las cuatro encuestas ( $B$ ); el valor  $X2$ , es la proporción de las preguntas que son calificadas con 4 ( $A2$ ) sobre el número total de preguntas aplicadas en las cuatro encuestas ( $B$ ). Los valores obtenidos en las fórmulas se presentan en la columna *Valor obtenido por variable* y el *Valor total obtenido* en la métrica  $X$  correspondería a la sumatoria de valores  $X1$  y  $X2$ . El *Valor*

deseado para la medida es 1, el cual se traduce en un porcentaje que representa el nivel de satisfacción.

**Tabla 36. Matriz de calidad. Fuente propia.**

Métrica	Fórmula	Valor obtenido por variable	Valor total obtenido	Valor deseado
Nivel de satisfacción	$X1 = A1/B$	$A1 = 123$ $B = 128$ $X1 = 0.961$	$X = X1 + X2$ $X = 0.969$	1
	$X2 = A2/B$	$A2 = 1$ $B = 128$ $X2 = 0.008$	96,9%	

En la Tabla 36 se observa que en el 96,1% (X1) de las preguntas de las funcionarias, manifiestan estar *Totalmente de acuerdo* y en el 0.8% (X2) están *Parcialmente de acuerdo*. La variable X1 y X2 se suman para obtener el valor total que corresponde a que los usuarios están de acuerdo (total y parcial) dando como resultado el 96.9% (X). Por lo tanto, con base en la Tabla 34, para el *valor total obtenido* el nivel de puntuación corresponde a *Cumple con los requisitos* y el grado de satisfacción es *Muy satisfactorio*.

Con lo anterior, se puede evidenciar que los reportes estándar y consultas analíticas, definidas en el prototipo de DW para *proyectos de investigación*, son elementos importantes para análisis de datos e información, representando un apoyo significativo en el proceso de toma de decisiones de la División.

## Capítulo 7

### 7 CONCLUSIONES, RECOMENDACIONES Y TRABAJO FUTURO

En este capítulo se realiza una descripción de las conclusiones obtenidas, recomendaciones para la División y propuesta de trabajo futuro, de acuerdo con el trabajado desarrollado en este proyecto.

#### 7.1 CONCLUSIONES

- Con la realización de este trabajo de grado, se proponen modelos dimensionales para Proyectos, Grupos, Semilleros y Convocatorias de investigación, los cuales incluyen información socio-demográfica de integrantes en estos procesos, que permite realizar importantes análisis mejorando la toma de decisiones en el tema de investigación. Para estos modelos, se detallan sus dimensiones, atributos y medidas, además de las técnicas de diseño empleadas para su creación, permitiendo que estos modelos sirvan como base, en situaciones similares de modelado, para demás instituciones que estén en el proceso de implementar diseños propios en los temas tratados.
- La identificación e inclusión en el prototipo de las técnicas de diseño: Minidimensión, Dimensión degenerada, Dimensión multivaluada y tabla puente, y Juego de roles; han permitido cumplir con los requerimientos analíticos presentados por los usuarios finales, presentando información ordenada y fácil de comprender.
- Es recomendable apoyarse en foros y tutoriales externos a la documentación de la suite de Pentano, ya que ésta no es detallada y explícita sobre cómo realizar la implementación de técnicas de diseño en la herramienta Schema Workbench.
- Fue necesario en la implementación del proceso ETL definir un atributo de fecha de actualización para cada registro de las dimensiones y tablas de hechos, para almacenar históricamente las modificaciones relacionadas con estos y permitir el análisis por parte de los funcionarios. Lo anterior, debido a que los sistemas fuente existentes solo guardan la información nueva o actualizada por última vez.
- Debido a que no se contó con la información real de los sistemas fuente, fue necesario implementar transformaciones para tener la menor cantidad de datos nulos o inconsistentes en la bodega de datos, pues de esto depende que se tenga mayor potencial de análisis por parte de los usuarios.
- El prototipo de bodega de datos para proyectos de investigación es evaluado con base a la medida *Nivel de satisfacción* de la norma ISO/IEC 25022: Medidas de *Calidad en Uso*, definida para la subcaracterística *Utilidad*, mostrando que el grado de satisfacción y por ende de utilidad del prototipo para los usuarios es del 96.9%, lo cual permite reconocer el poder analítico adquirido con la creación del modelo dimensional y la utilidad que la bodega de datos brinda a la División de Gestión de la Investigación de la Universidad del Cauca.

## **7.2 RECOMENDACIONES**

- Realizar una actualización constante de la documentación de la base de datos, ya que la existente presenta muchas inconsistencias con respecto a la definición de esta. Esta actualización debe aclarar qué tablas usa el sistema, una descripción sobre los atributos y sus tipos de datos, pues algunos no son usados en la base de datos o tienen nombres que no son comprensibles y se dificulta su interpretación.
- Modificar la base de datos existente en aras de mejorar el almacenamiento de la información de manera histórica, ya que al momento de realizar una actualización esta se sobrescribe, lo cual no permite que se puedan hacer tareas de seguimiento, control y análisis de datos.
- Es importante que los funcionarios de la División conozcan de manera detallada cómo debe ser ingresada la información a los diferentes sistemas fuente y la importancia de contar con datos completos y veraces, pues de esto depende en gran medida que los datos cargados durante el proceso de ETL sean de la mayor calidad posible, y así las consultas serán más completas y veraces, permitiendo que los análisis de la información sean correctos.
- Con el fin de ofrecer mayor visibilidad acerca de los indicadores y/o aspectos relevantes en los reportes y consultas propuestos, se recomienda implementar dashboards que permitan entender los procesos de negocio que se gestionan en la División de Gestión de la Investigación.

## **7.3 TRABAJO FUTURO**

- Implementar el prototipo de proyectos de investigación con datos reales y hacer el despliegue para que el personal de la División de investigación pueda acceder a los reportes y consultas analíticas con la información real para soportar la toma de decisiones.
- Ampliar el catálogo de reportes estándar y consultas analíticas para el proceso de proyectos de investigación, aunque se cumple con las expectativas de los usuarios, con la información que se tiene es posible realizar más análisis de los datos, apoyando así toma de decisiones sobre los aspectos generales de proyectos.
- Realizar el prototipado de los modelos dimensionales diseñados para los procesos de Grupos, Semilleros y Convocatorias de investigación; ya que también son importantes para la División debido a que cuentan con información relevante para el análisis de aspectos investigativos que se desarrollan en la Universidad.
- Adaptar los modelos dimensionales de Proyectos, Grupos, Semilleros y Convocatorias de investigación de la División a otras dependencias de universidades públicas colombianas y evaluar el nivel de adaptación.

## REFERENCIAS BIBLIOGRÁFICAS

- [1] “Vicerrectoría de Investigaciones - Universidad del Cauca,” Apoyo al Investigador. [Online]. Available: <http://vri.unicauca.edu.co/index.php/acerca-de/quienes-somos/item/77>. [Accessed: 23-Jan-2020].
- [2] N. D. Duque Méndez and A. Tamayo Alzate, “Data Warehouse (Bodega de Datos). Herramienta para la toma de decisiones (Parte I) Sistema de Procesamiento de Transacciones (OLTP: On-line Transaction Processing),” Univ. Nac. sede Manizales, no. 1, p. 126, 2001.
- [3] R. Kimball, R. Laura, M. Ross, and W. Thornthwaite, *The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing and Deploying Data Warehouse*, 2nd Ed. Indianapolis: John Wiley & Sons, Inc., 1998.
- [4] E. Miranda, E. Suryani, and R. Rudy, “Implementation Of Datawarehouse , Datamining and Dashboard for Higher Education,” *J. Theor. Appl. Inf. Technol.*, vol. 64, no. 3, pp. 710–717, 2014.
- [5] F. Di Tria, E. Lefons, F. Tangorra, D. Informatica, B. Aldo, and V. Orabona, “Academic Data Warehouse Design Using a Hybrid Methodology,” vol. 12, no. 1, pp. 135–160.
- [6] D. Solodovnikova, L. Niedrite, and A. Niedritis, “Architecture Enabling Adaptation of Data Integration Processes for a Research Information System,” *Found. Comput. Decis. Sci.*, vol. 43, no. 2, pp. 129–149, 2018.
- [7] E. Von F. Lapura, J. K. J. Fernandez, M. J. K Pagatpat, and D. D. Dinawanao, “Development of a University Financial Data Warehouse and its Visualization Tool,” *Procedia Comput. Sci.*, vol. 135, no. 1, pp. 587–595, 2018.
- [8] B. Kitchenham and S. Charters, “Procedures for performing systematic reviews.” Keele University, UK and National ICT Australia, 2004.
- [9] W. H. Inmon, *Building the Data Warehouse*, 3rd Ed. New York: John Wiley & Sons, Inc., 2002.
- [10] R. Kimball and M. Ross, *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeing*, 3rd Ed. Indianapolis: John Wiley & Sons, Inc., 2013.
- [11] R. A. ESPINOSA, “Data Warehouse Para La Gestión De Lista De Espera Sanitaria.” Univ. Politécnica Madrid, p. 148, 2018.
- [12] A. Cedeño Trujillo, “Modelo multidimensional,” *Ing. Ind.*, vol. 27, no. 1, p. 9, 2006.
- [13] P. Ponniah, *Data Warehousing Fundamentals*, 1st Ed. New York: John Wiley & Sons, Inc., 2001.
- [14] D. Bernabeu and G. Mattío, “Introducción DATA WAREHOUSING : Marco Conceptual HEFESTO : Metodología Data Warehouse,” p. 182, 2017.
- [15] C. S. Jensen, T. B. Pedersen, and C. Thomsen, “Multidimensional Databases and Data Warehousing,” *Synth. Lect. Data Manag.*, vol. 2, no. 1, pp. 1–111, 2010.
- [16] “Los data warehouses y la toma de decisiones en la empresa.” [Online]. Available: <https://www.ionos.es/digitalguide/online-marketing/analisis-web/los-data-warehouses-en-la-business-intelligence/>. [Accessed: 09-Dec-2019].

- [17] T. Setiadi and S. B. Premapasha, "Scrum Implementation for Online Transaction Processing (OLTP) in Hospital Management," in International Conference on Telecommunication Systems, Services, and Applications (TSSA), 2018, no. 1, pp. 6–8.
- [18] "Using Partitioning in an Online Transaction Processing Environment." .
- [19] J. Mundy and W. Thornthwaite, "The Microsoft Data Warehouse Toolkit: With SQL Server 2008 R2 and the Microsoft Business Intelligence Toolset (Google eBook)," p. 720, 2011.
- [20] "Extract, transform, and load (ETL) | Microsoft Docs." [Online]. Available: <https://docs.microsoft.com/en-us/azure/architecture/data-guide/relational-data/etl>. [Accessed: 16-Jul-2018].
- [21] "What is ETL (Extract, Transform, Load)? - Talend." [Online]. Available: <https://www.talend.com/resources/what-is-etl/>. [Accessed: 16-Jul-2018].
- [22] "K2 Standard Reports." [Online]. Available: [https://help.k2.com/onlinehelp/k2blackpearl/userguide/4.7/default.htm#5.Reports/W\\_S\\_Standard\\_Reports.htm](https://help.k2.com/onlinehelp/k2blackpearl/userguide/4.7/default.htm#5.Reports/W_S_Standard_Reports.htm). [Accessed: 28-Apr-2020].
- [23] "PORTAL ISO 25000." .
- [24] M. E. M. E. Mendoza, I. De Sistemas, M. Informática, I. De Sistemas, and I. De Sistemas, "Metodología de Desarrollo de Bodegas de Datos para Micro, Pequeñas y Medianas Empresas," no. 1, pp. 85–101, 2010.
- [25] "The Data Warehouse Lifecycle Toolkit - Ralph Kimball, Margy Ross, Warren Thornthwaite, Joy Mundy, Bob Becker - Google Libros." [Online]. Available: [https://books.google.com.co/books?id=ONQio9do\\_70C&printsec=frontcover&dq=The+Data+Warehouse+Lifecycle+Toolkit&hl=es&sa=X&ved=0ahUKEwiJuqWP5avnAhXkYN8KHRPRBOIQ6AEIKDAA#v=onepage&q=The+Data+Warehouse+Lifecycle+Toolkit&f=false](https://books.google.com.co/books?id=ONQio9do_70C&printsec=frontcover&dq=The+Data+Warehouse+Lifecycle+Toolkit&hl=es&sa=X&ved=0ahUKEwiJuqWP5avnAhXkYN8KHRPRBOIQ6AEIKDAA#v=onepage&q=The+Data+Warehouse+Lifecycle+Toolkit&f=false). [Accessed: 30-Jan-2020].
- [26] "Desde el Web Importer - Mendeley: guía básica - Biblioguías UCM at Universidad Complutense." [Online]. Available: <https://biblioguias.ucm.es/mendeley/anadir-desde-web-importer>. [Accessed: 07-Mar-2022].
- [27] "Data Integration Perspective in the PDI Client - Hitachi Vantara Lumada and Pentaho Documentation." [Online]. Available: [https://help.hitachivantara.com/Documentation/Pentaho/8.1/Products/Data\\_Integration/Data\\_Integration\\_Perspective](https://help.hitachivantara.com/Documentation/Pentaho/8.1/Products/Data_Integration/Data_Integration_Perspective). [Accessed: 12-Mar-2022].
- [28] S. Informáticos, Y. De Computación, B. Chisaguano, and E. Amparo, "ESCUELA POLITÉCNICA NACIONAL FACULTAD DE INGENIERÍA DE SISTEMAS EVALUACIÓN DE CALIDAD DE PRODUCTOS SOFTWARE EN EMPRESAS DE DESARROLLO DE SOFTWARE APLICANDO LA NORMA ISO/IEC 25000 PROYECTO PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO EN."
- [29] J. García Sánchez, J. Aguilera Terrats, and A. Castillo Rosas, "Guía técnica para la construcción de escalas de actitud," vol. 8, pp. 1–13, 2011.