

ANÁLISIS DE LAS PUBLICACIONES CIENTÍFICAS DE LA UNIVERSIDAD
PONTIFICIA BOLIVARIANA MEDIANTE UN MODELO DE DATOS
MULTIDIMENSIONAL

NATALIA HENAO SALDARRIAGA
LUNA CAROLINA BERNAL POSADA

UNIVERSIDAD PONTIFICIA BOLIVARIANA
ESCUELA INGENIERÍAS
FACULTAD DE INGENIERÍA EN TECNOLOGÍAS DE INFORMACIÓN Y
COMUNICACIÓN
MAESTRÍA EN TECNOLOGÍAS DE INFORMACIÓN Y COMUNICACIÓN
MEDELLÍN
2016

ANÁLISIS DE LAS PUBLICACIONES CIENTÍFICAS DE LA UNIVERSIDAD
PONTIFICIA BOLIVARIANA MEDIANTE UN MODELO DE DATOS
MULTIDIMENSIONAL

NATALIA HENAO SALDARRIAGA

LUNA CAROLINA BERNAL POSADA

Trabajo de grado para optar al título de Magíster en Tecnologías de la Información
y la Comunicación

Asesor

IVÁN AMÓN URIBE

Magíster en Ingeniería de Sistemas

UNIVERSIDAD PONTIFICIA BOLIVARIANA

ESCUELA INGENIERÍAS

FACULTAD DE INGENIERÍA EN TECNOLOGÍAS DE INFORMACIÓN Y
COMUNICACIÓN

MAESTRÍA EN TECNOLOGÍAS DE INFORMACIÓN Y COMUNICACIÓN

MEDELLÍN

2016

DECLARACIÓN ORIGINALIDAD

“Declaro que esta tesis (o trabajo de grado) no ha sido presentada para optar a un título, ya sea en igual forma o con variaciones, en esta o cualquier otra universidad”. Art. 82 Régimen Discente de Formación Avanzada, Universidad Pontificia Bolivariana.

FIRMA AUTOR (ES)

Luna Bernal R. Natalia Herazo S.

Ciudad y fecha

Medellín

21 de septiembre de 2016

CONTENIDO

INTRODUCCIÓN	5
1 PLANTEAMIENTO DEL PROBLEMA	6
1.1 PROBLEMA.....	6
1.2 JUSTIFICACIÓN	7
2 OBJETIVOS.....	8
2.1 OBJETIVO GENERAL.....	8
2.2 OBJETIVOS ESPECÍFICOS	8
3 MARCO REFERENCIAL	9
3.1 MARCO CONTEXTUAL.....	9
3.2 MARCO CONCEPTUAL.....	10
3.3 ESTADO DEL ARTE	15
3.3.1 Análisis de la producción científica de entidades universitarias.	15
3.3.2 Soluciones de inteligencia de negocios para instituciones académicas.	16
3.3.3 Aplicación de técnicas y herramientas de inteligencia de negocios para análisis de la producción científica	19
3.3.4 Análisis del estado del arte	20
4 METODOLOGÍA.....	22
5 PRESENTACIÓN Y ANÁLISIS DE RESULTADOS	25
5.1 IMPLEMENTACIÓN DE LA SOLUCIÓN	25
5.1.1 Levantamiento de requisitos	25
5.2 DISEÑO DEL MODELO MULTIDIMENSIONAL DE DATOS	28
5.2.1 Definir el proceso de negocio	30
5.2.2 Seleccionar el nivel de granularidad	30

5.2.3	Definir las dimensiones.....	30
5.2.4	Identificar las métricas	31
5.2.5	Definir la tabla de Hechos (Fact_Producto).....	31
5.3	IMPLEMENTACIÓN DEL MODELO MULTIDIMENSIONAL DE DATOS.....	36
5.3.1	Proceso de extracción	37
5.3.2	Proceso de transformación	39
5.3.3	Proceso de carga.....	42
5.4	PRESENTACIÓN DE RESULTADOS.....	46
5.4.1	Etapa 1: tablero de control.....	47
5.4.2	Etapa 2: análisis.....	48
5.4.3	Etapa 3: reporte	50
5.4.4	Análisis según la producción	52
5.4.5	Análisis según el impacto científico	59
6	CONCLUSIONES	63
7	TRABAJOS FUTUROS	65
8	REFERENCIAS	66
	ANEXO 1	70
	ANEXO 2	74

LISTA DE FIGURAS

FIGURA 1: ELEMENTOS BÁSICOS DE LA BODEGA DE DATOS	12
FIGURA 2: ADAPTACIÓN DE LA METODOLOGÍA DE KIMBALL PARA EL ANÁLISIS Y VISUALIZACIÓN DE LAS PUBLICACIONES CIENTÍFICAS DE LA UNIVERSIDAD PONTIFICIA BOLIVARIANA.....	22
FIGURA 3: <i>STAR NET</i> PARA DIMENSIÓN TIPO – HECHO PRODUCTO	28
FIGURA 4: <i>STAR NET</i> PARA DIMENSIÓN IDIOMA– HECHO PRODUCTO	29
FIGURA 5: <i>STAR NET</i> PARA DIMENSIÓN TIEMPO– HECHO PRODUCTO	29
FIGURA 6: <i>STAR NET</i> PARA DIMENSIÓN INSTITUCIÓN – HECHO PRODUCTO	29
FIGURA 7: <i>STAR NET</i> PARA DIMENSIÓN REVISTA – HECHO PRODUCTO....	29
FIGURA 8: <i>STAR NET</i> PARA DIMENSIÓN PERSONA – HECHO PRODUCTO..	29
FIGURA 9: MODELO DE DATOS MULTIDIMENSIONAL DISEÑADO	32
FIGURA 10: PROCESO DE <i>ETL</i> IMPLEMENTADO	36
FIGURA 11: MODELO DE DATOS MULTIDIMENSIONAL IMPLEMENTADO	46
FIGURA 12: TABLERO DE CONTROL GENERAL	48
FIGURA 13: CARACTERIZACIÓN DE LA PRODUCCIÓN	49
FIGURA 14: FUENTES DE PUBLICACIÓN.....	50
FIGURA 15: DETALLE PRODUCTO	51
FIGURA 16: CATEGORÍAS DE LAS FUENTES DE PUBLICACIÓN	51
FIGURA 17: KPI	52

LISTA DE GRÁFICAS

GRÁFICA 1: EVOLUCIÓN DE LA PRODUCCIÓN CIENTÍFICA	53
GRÁFICA 2: PARTICIPACIÓN EN PRODUCCIÓN POR NIVEL DE FORMACIÓN	54
GRÁFICA 3: EJEMPLO DE PRODUCTO POR UN SOLO NIVEL DE FORMACIÓN	55
GRÁFICA 4: EJEMPLO DE PRODUCTO POR VARIOS NIVELES DE FORMACIÓN	55
GRÁFICA 5: PARTICIPACIÓN EN PRODUCCIÓN POR TIPO DE VINCULACIÓN	56
GRÁFICA 6: PARTICIPACIÓN EN PRODUCCIÓN POR ESCUELA	57
GRÁFICA 7: PRODUCTOS POR INVESTIGADOR	57
GRÁFICA 8: PRODUCTOS POR GRUPO DE INVESTIGACIÓN	58
GRÁFICA 9: TIPOS DE PRODUCTOS.....	59
GRÁFICA 10: PRODUCTOS POR CUARTIL	60
GRÁFICA 11: CANTIDAD DE PRODUCTOS Y CITACIÓN POR INVESTIGADOR	61
GRÁFICA 12: PRODUCTOS POR FUENTES DE PUBLICACIÓN	61
GRÁFICA 13: CITACIONES POR FUENTE DE PUBLICACIÓN	62

LISTA DE TABLAS

TABLA 1: USUARIOS ENCUESTADOS.....	26
TABLA 2: REQUISITOS FUNCIONALES DEL SISTEMA	27
TABLA 3: REQUISITOS NO FUNCIONALES DEL SISTEMA	27
TABLA 4: INTEGRANTES DE GRUPOS DE INVESTIGACIÓN.....	34
TABLA 5: INVESTIGADORES	34
TABLA 6: ARTÍCULOS	35
TABLA 7: PAÍSES DE LA FUENTE DE PUBLICACIÓN.....	35
TABLA 8: DESCRIPCIÓN DE LOS NIVELES DE FORMACIÓN DE LOS INVESTIGADORES.....	35
TABLA 9: CUARTILES DE LAS FUENTES DE PUBLICACIÓN.....	35
TABLA 10: CATEGORÍAS DE LAS FUENTES DE PUBLICACIÓN	36
TABLA 11: IDIOMAS DE PUBLICACIÓN DE LOS PRODUCTOS	36
TABLA 12: TABLA EQUIVALENTE PARA INTEGRANTES DE GRUPOS DE INVESTIGACIÓN.....	37
TABLA 13: TABLA EQUIVALENTE PARA INVESTIGADORES.....	37
TABLA 14: TABLA EQUIVALENTE PARA ARTÍCULOS	38
TABLA 15: TABLA EQUIVALENTE PARA PAÍSES DE LA FUENTE DE PUBLICACIÓN	38
TABLA 16: TABLA EQUIVALENTE PARA NIVELES DE FORMACIÓN DE LOS INVESTIGADORES.....	38
TABLA 17: TABLA EQUIVALENTE PARA CUARTILES DE LAS FUENTES DE PUBLICACIÓN	39
TABLA 18: TABLA EQUIVALENTE PARA CATEGORÍAS DE LAS FUENTES DE PUBLICACIÓN	39
TABLA 19: TABLA EQUIVALENTE PARA IDIOMAS DE PUBLICACIÓN DE LOS PRODUCTOS.....	39
TABLA 20: PROCESO DE TRANSFORMACIÓN PARA LA DIMENSIÓN FUENTES	40
TABLA 21: PROCESO DE TRANSFORMACIÓN PARA LA DIMENSIÓN CATEGORÍA.....	40
TABLA 22: PROCESO DE TRANSFORMACIÓN PARA LA DIMENSIÓN CUARTILES.....	40
TABLA 23: PROCESO DE TRANSFORMACIÓN PARA LA DIMENSIÓN IDIOMA	40
TABLA 24: PROCESO DE TRANSFORMACIÓN PARA LA DIMENSIÓN TIPO PRODUCTO	41
TABLA 25: PROCESO DE TRANSFORMACIÓN PARA LA DIMENSIÓN TIEMPO	41

TABLA 26: PROCESO DE TRANSFORMACIÓN PARA LA TABLA COMPLEMENTARIA PRODUCTO	41
TABLA 27: PROCESO DE TRANSFORMACIÓN PARA LA TABLA COMPLEMENTARIA INVESTIGADORES	41
TABLA 28: PROCESO DE TRANSFORMACIÓN PARA LA TABLA DE HECHOS	42
TABLA 29: DIMENSIÓN CATEGORÍA	43
TABLA 30: DIMENSIÓN CUARTIL	43
TABLA 31: DIMENSIÓN IDIOMA.....	43
TABLA 32: DIMENSIÓN FUENTES.....	44
TABLA 33: DIMENSIÓN TIEMPO.....	44
TABLA 34: DIMENSIÓN TIPO PRODUCTO.....	44
TABLA 35: TABLA COMPLEMENTARIA PRODUCTO	44
TABLA 36: TABLA COMPLEMENTARIA INVESTIGADORES.....	45
TABLA 37: TABLA COMPLEMENTARIA INVESTIGADORES.....	45

GLOSARIO

ARTÍCULO CIENTÍFICO: es un documento redactado para difundir los resultados o nuevos conocimientos obtenidos a partir de una investigación científica realizada.

ARTICLE IN PRESS: artículo que ha sido aceptado para publicación, pero aún no ha sido publicado formalmente por que se encuentra a la espera de ser editado.

BANNER: sistema que se encarga de optimizar los procesos académicos, administrativos y financieros de la UPB soportados en tecnologías de la información y la comunicación

CONFERENCE PAPER: es un texto que fue redactado con el objetivo de ser aceptado en una conferencia

CUARTILES: es un indicador que permite conocer cuál es el posicionamiento de una revista científica con respecto a otras revistas similares.

ETL: proceso de extracción, transformación y carga de datos.

HECHO: es el objeto de análisis en el modelo de datos multidimensional.

INDEXACIÓN: proceso que permiten normalizar las publicaciones científicas bajo ciertos criterios técnicos que otorgan validez.

ISSN: número internacional normalizado de publicaciones seriadas, es un código único de 8 dígitos para identificar cualquier tipo de publicación seriada.

KPI: indicador clave de desempeño.

PROCEEDINGS: es una colección de *conference papers* o contribuciones hechas en una conferencia o reunión científica en particular.

QLIK SENSE: herramienta de inteligencia de negocios.

REVISTAS CIENTÍFICAS: son una recopilación de varios artículos científicos, su publicación es periódica y se especializan por áreas de conocimiento.

SCIMAGO JOURNAL & COUNTRY RANK: aplicativo web para la consulta de revistas e indicadores.

SCOPUS: base de datos que contiene bibliografía científica; resúmenes y citas de literatura académica a nivel mundial.

STAR NET: diagrama de las dimensiones a analizar.

RESUMEN

Lograr un alto posicionamiento en los principales rankings de medición según los indicadores de investigación y producción científica, es un reto constante para la Universidad Pontificia Bolivariana. Para alcanzar este reto la institución trabaja bajo el concepto de universidad docente con énfasis en investigación, este modelo de trabajo y la constante importancia que se le da a la actividad investigativa han permitido aumentar los niveles de generación de productos de investigación, tales como publicaciones científicas.

Es importante recalcar que, aunque la Universidad está en constante producción de investigación, hoy en día, la institución cuenta solo con revisiones básicas sobre el estado actual de la producción científica, pero aún no cuenta con procesos rutinarios que permitan organizar y conocer el estado real de dicha producción.

A partir de la necesidad anterior surge este proyecto, el cual contribuirá para que la Universidad Pontificia Bolivariana cuente con un sistema que facilite el análisis y permita a los usuarios acceder a información para generar conocimiento y realizar seguimiento a las publicaciones científicas desarrolladas por los investigadores. Para este trabajo se consideran los trabajos realizados en los años 2009-2015.

El proyecto permitió obtener visualizaciones a partir de las cuales se conoció y analizó el comportamiento de la producción de publicaciones científicas, estos resultados serán útiles para la elaboración futura de estrategias que permitan mejorar el posicionamiento actual en los rankings, fortalecer la actividad investigativa de la institución y mejorar la visibilidad e impacto de las publicaciones científicas de la Universidad Pontificia Bolivariana.

PALABRAS CLAVE: modelo multidimensional de datos; inteligencia de negocios; Qlik Sense; publicación científica; Scopus.

ABSTRACT

Achieving a high position in the major research rankings is an ongoing challenge for the Universidad Pontificia Bolivariana, that is why the institution has implemented the concept of teaching with emphasis on research. This concept and the continued importance given to research activity have helped to increase the research products generation, such as scientific publications.

Even though the university is in a continuous process of research generation, nowadays the institution has only basic reviews about the current status of the research activity and has not yet any process that could help organize and analyze the real research production status.

This project implemented a system to facilitate the access and analysis to all the information related to research generation activity accomplished in the years 2009 to 2015.

The obtained results were multiple dashboards used to analyze the behavior and evolution of the research activity, the results of the analysis will be helpful in a future for the creation of new strategies to reach a better ranking position and enhance the research activity and production.

KEY WORDS: multidimensional data model; business intelligence; Qlik Sense; Scientific publications; Scopus.

INTRODUCCIÓN

En la actualidad todas las organizaciones, sin importar su enfoque, se enfrentan diariamente a la toma rápida de decisiones impactando todas las áreas de negocio.

La inteligencia de negocios es un conjunto de estrategias y herramientas que permiten adquirir conocimiento a partir de datos, tienen como objetivo facilitar a los usuarios un acceso integrado, rápido y fácil a la información requerida por estos, para optimizar la toma de decisiones. Los sistemas de inteligencia de negocios deben integrar datos de diferentes fuentes y entregarlos buscando la integridad, exactitud, conformidad y oportunidad de los datos.

Con este proyecto se construyó un mecanismo de análisis de la información de la producción investigativa proveniente de diferentes fuentes de datos (Scopus, SCImago, entre otros) relacionada con la Universidad Pontificia Bolivariana en los años 2009 al 2015. Para ello se utilizó la herramienta de inteligencia de negocios Qlik Sense, para analizar y lograr visualizar cómo ha sido el desempeño asociado a esta actividad.

Lo anterior se hizo ya que el análisis de la actividad científica es una acción estratégica que ayuda a conocer el desempeño de los programas de investigación de cualquier entidad. Ejecutar dichos análisis, permite a las entidades impulsar sus líneas de mayor impacto científico (Comunidad de Madrid, 2002) y mejorar su visibilidad y posicionamiento en el contexto mundial o regional (Arencibia & de Moya Anegón, 2008).

Este informe está estructurado de la siguiente manera: en el capítulo 2 se exponen el problema principal y la justificación del proyecto. En el capítulo 3 se detallan los objetivos propuestos. En el capítulo 4 se presenta los marcos contextual y teórico del proyecto, luego en el capítulo 5 se expone la metodología base para la ejecución del proyecto y en el capítulo 6 se exponen los resultados y análisis realizados. Finalmente, en el capítulo 7 se presentan las conclusiones logradas y en el capítulo 8 se plantean trabajos futuros.

1 PLANTEAMIENTO DEL PROBLEMA

1.1 PROBLEMA

Actualmente, para la Universidad Pontificia Bolivariana, es difícil conocer y cuantificar la información relacionada con la producción científica institucional indexada en la base de datos *Scopus*. Esto se debe a la falta de herramientas confiables que permitan la extracción y almacenamiento de datos de las diferentes fuentes de datos que puedan proporcionar dicha información, tampoco se cuenta con políticas de calidad de datos institucionales que faciliten tener un repositorio de datos, al cual se pueda acceder para consultar información precisa sobre los diferentes entes participantes en la generación de producción científica institucional.

Lo anterior aunado al exigente trabajo manual para la obtención de la información, impide a la institución realizar análisis frecuentes y fiables, a partir de los cuales se llegaría a identificar falencias, generar conocimiento y plantear nuevas estrategias, con el fin de aumentar los niveles de producción científica de la Universidad, mejorando el posicionamiento institucional en los diferentes rankings universitarios y consolidando a la institución dentro de las primeras en el ámbito nacional e internacional.

En vista de lo anterior, este trabajo dota a la Universidad de una herramienta analítica, la cual permitirá conocer y generar indicadores de producción científica basados en la cantidad de productos y citaciones, tipos de productos, lenguaje de publicación, entre otros. Permitiendo a los investigadores conocer sus niveles de transferencia de conocimiento, parte fundamental de las concepciones estipuladas en el Modelo Pedagógico Institucional (Universidad Pontificia Bolivariana, 2015). Otro aspecto importante en este proyecto, tiene que ver con la reducción de la dependencia del profesional encargado de generar los indicadores, pues se pretende que estos sean fáciles de obtener y analizar mediante unas indicaciones básicas.

Para esto se realizó la extracción y transformación de diferentes datos provenientes de múltiples fuentes de información como: *Scopus*, *SCImago* y del sistema Banner, entre otros. Estos datos fueron llevados a la herramienta analítica *Qlik Sense* para ser consolidados y analizados mediante un modelo multidimensional, a través del cual se pueden realizar en forma interactiva múltiples consultas para obtener diferentes resultados, y así lograr dar respuesta a

las necesidades de los usuarios de la aplicación, haciendo uso de tableros de control.

1.2 JUSTIFICACIÓN

La competitividad y la alta producción científica aplicada, son en sí mismas indicadores de desarrollo que posibilitan el crecimiento económico y social de un país, una institución que cuente con buenos índices de producción tiene ventajas en cuanto a la financiación pública y privada, y mayor demanda en el mercado global, lo que es esencial para el crecimiento de las instituciones (Álvarez de Toledo, 2012).

Aunque en la actualidad existen empresas que ofrecen servicios para el análisis y seguimiento de la producción científica de las instituciones, estos son costosos, poco flexibles y como el usuario no tiene control sobre los datos no permiten enriquecer el contenido de la información con datos de interés de la institución, por ejemplo, grupo de investigación, área a la que pertenece (escuela, facultad, programa), nivel de información de los autores, etc. Esto dificulta enormemente una detallada caracterización de los investigadores limitando las opciones de análisis que se pueden realizar. Adicionalmente, al tener el control de los datos, sería posible a futuro realizar análisis de minería de datos buscando patrones sobre la productividad de los investigadores y otros análisis interesantes.

Por lo anterior en este trabajo se implementó un modelo multidimensional propio de la institución, que permite relacionar la producción científica, las escuelas, grupos de investigación, entre otros aspectos relacionados, con el fin de tener un insumo para tomar acciones enfocadas a la consolidación nacional y fortalecimiento internacional, para el mejoramiento de la visibilidad e impacto científico de la Universidad Pontificia Bolivariana, según lo planteado en la meta número 30.01 “Documento de Monitoreo y Vigilancia de la Producción Científica” del Plan de Desarrollo Institucional (Universidad Pontificia Bolivariana, 2013).

2 OBJETIVOS

2.1 OBJETIVO GENERAL

Analizar el comportamiento y desempeño de los datos asociados a las publicaciones científicas de la Universidad Pontificia Bolivariana, mediante la construcción de un modelo multidimensional, con el fin de consolidar la información y conocer el estado actual de la producción científica de la institución.

2.2 OBJETIVOS ESPECÍFICOS

- Realizar el levantamiento de los requisitos necesarios para el análisis de las publicaciones científicas.
- Realizar el diseño de un modelo multidimensional para las publicaciones científicas de la Universidad Pontificia Bolivariana.
- Realizar la implementación de un modelo multidimensional para las publicaciones científicas de la Universidad Pontificia Bolivariana.
- Realizar un análisis de los datos obtenidos del tablero de control.

3 MARCO REFERENCIAL

3.1 MARCO CONTEXTUAL

La Universidad Pontificia Bolivariana ha venido concentrando sus esfuerzos en lograr un modelo universitario de docencia con énfasis en investigación e innovación, como resultado de lo anterior busca un nivel importante de desarrollo de la investigación, que contribuya al incremento del número de proyectos de investigación y publicaciones científicas de alto impacto. Las publicaciones científicas son el resultado fundamental de la actividad investigativa, que permiten que la comunidad científica conozca el nuevo conocimiento y puedan desarrollar aportes a partir de este, lo que conlleva a contribuir con el avance científico (Jiménez, 2006).

Para consolidar este modelo, en 2013 se configura un cargo que depende de la Vicerrectoría académica llamado coordinación de visibilidad académica e impacto científico. Esta coordinación define un modelo de trabajo con tres objetivos (Gómez Vargas, 2014):

- Sistematizar y posicionar los procesos editoriales de las revistas académicas y científicas
- Incrementar y promover la producción académica y científica en la comunidad universitaria
- Asegurar e incrementar la visibilidad de la producción científica y académica

Estos objetivos se ejecutan en metas dentro del plan de desarrollo 2014-2016 (Universidad Pontificia Bolivariana, 2014), asociados a la indexación internacional de las revistas, aumento de la producción científica y mejoras en el posicionamiento en los rankings.

Aunado a lo anterior, el Sistema de Investigaciones, Transferencia e Innovación a través del Programa Fortalecimiento Investigativo fomenta la proyección y visibilidad de los resultados de la investigación por parte de las estructuras de conocimiento en materia de publicaciones y ponencias de alto impacto a través de cuatro grandes dinamizadores a saber: apoyos a movilidad y publicaciones, gestión de grupos de investigación, publicaciones de alta visibilidad y monitoreo y acompañamiento a convocatorias (Centro de Investigación para el Desarrollo y la Innovación (CIDI) , 2014).

3.2 MARCO CONCEPTUAL

A continuación, se presentan algunos términos necesarios para el entendimiento del proyecto:

Base de datos

Las bases de datos son herramientas sistematizadas que soportan de manera eficiente la recolección, almacenamiento, entrega, actualización y mantenimiento de datos, estas permiten automatizar tareas, entregar o recuperar información de manera rápida y flexible, facilitando el almacenamiento, organización y actualización de grandes cantidades de información (Microsoft, 2010).

Modelo de datos relacional

En este modelo los datos están almacenados en varias tablas, cada tabla debe tener un campo único denominado clave o llave, el cual permite que las diferentes tablas se conecten entre sí, es decir, se relacionen.

Los componentes de los modelos de datos relacionales son (Sharma, y otros, 2010):

- Tablas: denominadas relaciones
- Columnas: denominadas campos o atributos, son características de los datos.
- Filas: denominadas registros, son un conjunto ordenado de valores en cierto momento.
- Dominio: conjunto de posibles resultados de un atributo.
- Esquema: es una descripción formal de todas las tablas y las relaciones existentes entre ellas.

Base de datos relacional

Son el tipo de bases de datos más utilizadas, están compuestas por varias tablas basadas en el modelo relacional, y son creadas a través del lenguaje de consulta estructurado, las relaciones entre las diferentes tablas se ejecutan utilizando el comando Join (How stuff works, 2001).

Permiten encontrar información específica fácilmente y la agregación de nuevos datos, es decir, son escalables. A estas bases de datos se les puede aplicar procesos de normalización, el cual simplemente es la aplicación de varias reglas que ayudan a eliminar la ambigüedad y la redundancia en los datos, evitando las

anomalías y pérdida de integridad de datos, asegurando así una estructura óptima de los datos.

Modelo de datos multidimensional

Estos modelos de datos utilizan la idea de un cubo de datos, para representar las diferentes dimensiones de datos almacenados disponibles para los usuarios y son utilizados para optimizar las consultas.

Un cubo se puede definir como una estructura de varias dimensiones, el cual siempre debe tener presente la dimensión tiempo. Sus principales componentes de estos modelos son las dimensiones y las métricas contenidas en la tabla de hechos. Esta estructura permite que los usuarios puedan particionar los datos de diferentes maneras para personalizar los resultados de las métricas (Bazan, 2013)

Los conceptos de un modelo multidimensional son (Carpani, 2000)

- Dimensiones: son los ángulos desde los cuales la organización quiere ver sus datos, cada dimensión debe contener una clave primaria.
- Atributos: se encuentran dentro de las dimensiones, se pueden considerar como columnas, una dimensión puede contener múltiples atributos, que cómo se explicó anteriormente son las características de los datos.
- Jerarquías: los atributos de las dimensiones se pueden estructurar por jerarquías que contienen múltiples niveles de agregación.
- Tabla de hechos: este componente almacena las claves primarias de cada una de las dimensiones y las métricas que se quieren obtener.

Bases de datos multidimensionales

Estas bases de datos están optimizadas para ser utilizadas como parte de un almacén de datos y aplicaciones procesamiento analítico en línea. Brindan un rápido procesamiento de datos y más capacidades de análisis que las bases de datos relacionales. Utilizan el lenguaje de expresiones multidimensionales (Rouse, 2005).

Bodega de datos

Es un depósito de datos consolidado, el cual permite tener una sola versión de los datos y debe ser (George, 2012):

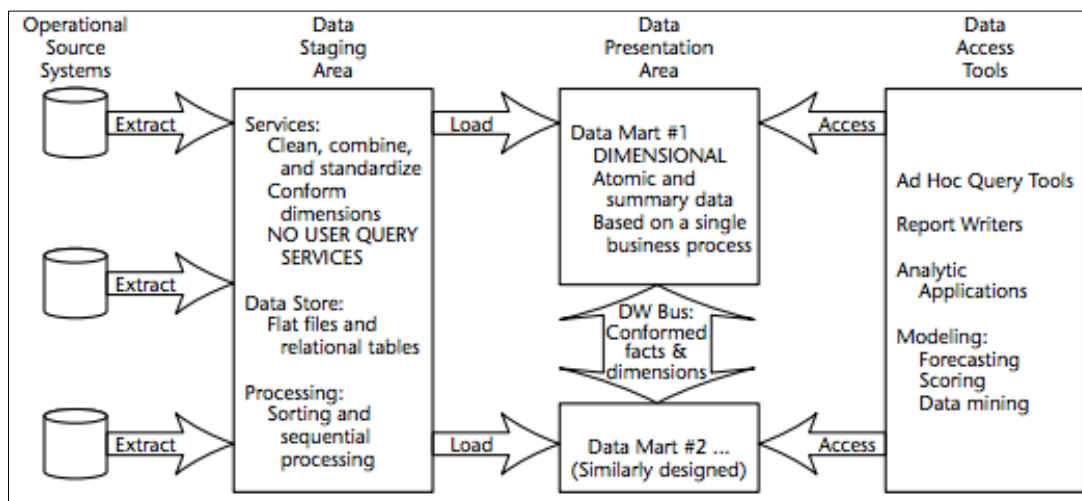
- Orientado: Solo la información relevante para el negocio es almacenada.
- Integrado: Las bodegas de datos deben integrar la información generada por todos los distintos sistemas de la empresa.

- Variable en el tiempo: Los datos que cambien su valor en el tiempo deben ser cargados periódicamente para permitir realizar comparaciones.
- No volátil: La información almacenada en una bodega de datos es estática, solamente puede ser leída y no modificada, es decir, los datos se incrementan, no se actualizan.

En la Figura 1 se presentan los componentes de una bodega de datos (Kimball & Ross, 2002):

- Sistemas operativos fuentes: son los sistemas que capturan todas las transacciones del negocio.
- Área de almacenamiento temporal: en esta área se almacenan los datos temporalmente para luego ser transformados mediante los procesos de Extracción, Transformación y Carga (ETL, por sus siglas en inglés) en salidas que cumplan con las peticiones de los usuarios sin necesidad de que el usuario intervenga. En esta área también se crea el repositorio de metadata, el cual contiene toda la información relacionada a las transformaciones, los flujos de trabajo y demás procesos llevados a cabo.
- Área de presentación de datos: en esta área es donde los datos están organizados, almacenados y disponible para las peticiones de los usuarios, creación de reportes y otras aplicaciones.
- Herramientas de acceso de datos: son todos los componentes o herramientas que permiten tomar los datos disponibles en el área de presentación para iniciar la toma de decisiones, pueden ser herramientas de visualización, minería de datos y creación de reportes entre otras.

Figura 1: Elementos básicos de la bodega de datos



Fuente: (Kimball & Ross, 2002)

ETL

EL proceso de Extracción, Transformación y Carga (*ETL, por sus siglas en inglés*), permite mejorar la calidad de los datos, detectar datos inconsistentes, estandarizar la información y unificar todas las diferentes fuentes de datos.

Extracción: en esta parte del proceso se capturan los datos de las diferentes fuentes.

Transformación: este proceso permite eliminar anomalías y duplicados, descomponer valores y permite la unificación estandarizada de los datos en sus nuevos formatos. Es llevado a cabo en el área de almacenamiento temporal.

Carga: en este proceso, la información que esté en el área de almacenamiento temporal es trasladada finalmente a la bodega de datos (Sinergia e Inteligencia de Negocio, 2010)

Data Mart

Es un subconjunto especializado de la bodega de datos, que une la información necesaria para responder una necesidad de un departamento o área específica del negocio.

Los datos se organizan siguiendo un esquema de estrella o en un cubo (Kimball, Reeves, Ross, & Thornthwaite, 1998)

OLAP-Procesamiento analítico en línea

Son bases de datos que contienen datos con formatos estandarizados, incluyendo información histórica, provenientes de la bodega de datos. Con estos sistemas se pueden realizar tareas analíticas de los datos para tomar decisiones, únicamente se pueden realizar lecturas o consultas (Sinergia e Inteligencia de Negocio, 2010).

OLTP-Procesamiento transaccional en línea

Son bases de datos que contienen los datos operacionales y permiten el procesamiento de transacciones, en estas se pueden insertar, modificar y actualizar los datos. Estos sistemas no contienen mucha información histórica, no son centralizados y los formatos de los datos no son estandarizados (Sinergia e Inteligencia de Negocio, 2010).

Inteligencia de negocios

La inteligencia de negocios se conoce como un conjunto de herramientas, procesos, técnicas y buenas prácticas, que se centran en resaltar el valor estratégico que

representan los datos para cualquier organización sin importar cuál sea su tipo, con el fin de generar conocimiento y obtener ventaja competitiva.

La inteligencia de negocios da soporte a la toma de decisiones rápidas y correctas, mediante la utilización de datos centralizados y verídicos. Las soluciones de inteligencia de negocios deben de estar enfocadas en las necesidades del negocio.

Estas soluciones se componen de sistemas *OLTP*, bodegas de datos, y sistemas *OLAP*, los cuales al final permiten analizar los datos para crear reportes y aplicar técnicas de minería de datos y visualización (Ranjan, 2010).

Tablero de control

Es una herramienta que permite visualizar los datos que son extraídos de la bodega de datos por medio de peticiones realizadas por una herramienta de acceso a datos. Los tableros de control brindan una visualización simplificada de los parámetros más importantes para el negocio o un área específica, es decir, permite hacer seguimiento al desempeño de las métricas definidas y de los indicadores clave de desempeño (*KPI*, por sus siglas en inglés).

Estas herramientas permiten a los usuarios estar permanentemente actualizados sobre la información del negocio. Se sugiere que sean interactivos y permitan realizar filtrado de datos para personalizar las presentaciones según las necesidades de visualización requeridas.

Scopus

Es una base de datos de acceso web, que contiene información bibliográfica, resúmenes y citas de publicaciones de revistas científicas de todo el mundo y de diferentes áreas disciplinares.

CRIS

Los sistemas *CRIS* (*Current Research Information Systems*), son bases de datos que permiten almacenar y gestionar la información relacionada a las investigaciones científicas desarrolladas por una institución.

Estos sistemas presentan información que puede ser utilizada para realizar mediciones y reportes de productividad. Algunos detalles que se pueden obtener de un sistema *CRIS* son: autores, detalle de las publicaciones, colaboraciones y resultados de las investigaciones. Toda esta información permite llevar a cabo comparaciones sobre el desempeño y el progreso de la actividad científica desempeñada en una institución (Bijsterbosch, 2015).

3.3 ESTADO DEL ARTE

A continuación, se presentan trabajos similares realizados en el ámbito local, nacional e internacional.

3.3.1 Análisis de la producción científica de entidades universitarias.

En el estudio realizado por Sandra Raiher en el año 2010, se presenta un análisis de la actividad científica Latinoamérica relacionada con la temática: inequidades en la salud. Este proyecto recopiló los datos existentes desde el año 1990 hasta el año 2008, en las bibliotecas virtuales de la salud LILACS y MEDLINE, y en la base de datos *Web of Science*.

El trabajo permitió comparar la producción científica, a través del uso de visualizaciones creadas por la herramienta *Many Eyes* de IBM. A partir de las comparaciones se pudo definir: cómo había sido la evolución de la producción científica de esta temática a lo largo de los años estudiados, cuáles habían sido los patrones de colaboración entre países e instituciones, entre otros hallazgos. También se pudo identificar cuáles eran los focos de producción, tanto países como revistas y conocer las temáticas desarrolladas asociadas al tema principal (Raiher, 2010).

En el contexto nacional, en (Eslava Schmalbach, Gaitán Duarte, & Escobar Córdoba, 2014), se muestra un estudio que permitió analizar la producción científica de las facultades de Medicina de Colombia en revistas indexadas en la base de datos *Scopus* desde el año 1940 al año 2014.

Para el despliegue de este proyecto se creó una base de datos en Excel que permitió comparar los niveles de producción de varias instituciones de Colombia. A partir de esto se crearon histogramas y figuras de tendencias para conocer el desempeño de las distintas facultades de medicina de Colombia.

La Universidad de Antioquia con el fin de mejorar su proyección nacional e internacional, decidió en 2013 contratar los servicios de *SCImago Research Group*, para realizar un informe de caracterización de la producción científica de la institución para el periodo 2003-2011, utilizando la base de datos *Scopus*. Los análisis del informe permitieron identificar los niveles de producción de la institución, visualizar la distribución de la producción por área temática, identificar las áreas emergentes y de mayor crecimiento, y conocer los niveles de colaboración internacional. En los resultados obtenidos se identificó que aproximadamente un 60% de la producción de la universidad era en inglés, por lo cual se sugirió crear

estrategias para el aumento a los incentivos por generar publicaciones en inglés (Cortés Rodas, 2013)

De manera similar, la Universidad Nacional sede Medellín realizó en el año 2009, la caracterización de la producción científica de la institución, a partir de información recopilada de la base de datos *Web of Science*, abarcando desde el año 1990 hasta el año 2007.

Este estudio se realizó para conocer el comportamiento de la actividad y medir la visibilidad e impacto de las producciones. Se logró a través de técnicas estadísticas y herramientas de visualización, identificar patrones de colaboración entre instituciones y docentes, conocer la producción total, el tipo de documento, los idiomas de publicación y realizar proyecciones (**Ospina Rúa, 2009**).

3.3.2 Soluciones de inteligencia de negocios para instituciones académicas.

La universidad McMaster en Canadá muestra en (**Weisensee, Matthews, & McInnis, 2005**), cómo se implementó un *Data Mart* orientado al financiamiento de proyectos de investigación, como parte inicial de la creación incremental de una bodega de datos para toda la universidad.

Este proyecto se hizo con el objetivo de crear una solución integrada entre la base de datos de recaudación de fondos y la financiera para analizar, crear reportes, evaluar y proyectar el desempeño a nivel institucional, nacional e internacional. Con esto se proporcionó a la comunidad del campus una mayor capacidad de análisis de datos y mayor interpretación para apoyar la toma de decisiones y facilitar la rendición de cuentas. Para el desarrollo de esta solución se crearon cubos *OLAP*, se utilizó *ETL Studio* para el proceso de *ETL* y otras soluciones SAS para crear reportes estáticos y dinámicos, que permitieran medir el desempeño de los *KPIs*.

Con un objetivo similar al del proyecto anterior de mejorar la toma de decisiones en instituciones, el trabajo realizado por Mario Roberto Reyes y Pablo Augusto Rosales para la facultad de ingeniería de la Universidad de San Carlos de Guatemala, logró implementar una solución de inteligencia de negocios para dar apoyo a la toma de decisiones tácticas y estratégicas de la organización. Esta solución se basó en el desarrollo de un *Data Mart* orientado a información académica de la escuela de Ciencias y Sistemas, y permitió conocer los indicadores académicos y conocer los factores que influyen en el desempeño

académico, a través de la generación de reportes que permitieron conocer factores como: demografía de estudiantes, historial de cursos por estudiante carga académica por catedrático, ranking de estudiantes por curso y cantidad de graduados de la escuela **(Reyes Marroquín & Rosales Tejada, 2007)**.

Continuando con las herramientas para soporte y mejora de los sistemas de gestión académica, en el documento presentado por el Laboratorio de Investigación en Nuevas Tecnologías Informáticas de la Facultad de Informática de la Universidad Nacional de La Plata, se dan a conocer los acercamientos que ha tenido esta institución con las temáticas relacionadas a inteligencia de negocios, todo con la finalidad de generar conocimiento a partir de los datos almacenados en los diferentes sistemas académicos que se utilizan con frecuencia en la organizaciones, entre estos están; el anuarios estadístico, la plataforma de conocimiento *Moodle*, cuentas de *Facebook* de difusión académica vinculadas a la institución y el software Meran de uso bibliotecario, entre otros.

Esta solución se realizó a través de la utilización de herramientas de inteligencia de negocios como *IdeasoftO3*, *Pentaho* y *Spago BI*. Entre los resultados logrados hasta el año 2013, se encontraba la implementación de un cubo, que permite conocer los niveles de deserción de la Facultad y plantear nuevas estrategias y políticas para abordar la problemática. También se logró, conocer el desempeño de algunas materias, cohortes, cursos y profesores asociados a la Facultad y a partir de estos análisis desarrollar una nueva estrategia para el periodo 2012-2016. Esta solución también permitió conocer los perfiles de los estudiantes y egresados, y caracterizar a los nuevos aspirantes.

El objetivo de este proyecto fue consolidar la información proveniente de todos los sistemas operacionales académicos y transformarla para dar soporte a la toma de decisiones, resolución de situaciones y problemáticas, partiendo de datos confiables y estándares **(Díaz, Osorio, Amadeo, & Romero, 2013)**.

En el informe presentado por Miguel Rodríguez Sanz, se presenta la implementación de un modelo dimensional de datos para configurar un *Data Mart* siguiendo la metodología de Ralph Kimball **(Kimball & Ross, The Data Warehouse Toolkit, 2002)**, para conocer cuáles son los factores relacionados al proceso académico de los estudiantes, incluyendo los niveles de deserción y de éxito, carga académica, nivel de dificultad de las asignaturas, porcentaje de aprobados, cambios de titulación y demografía de los alumnos. Este proyecto permitió ampliar la visión de los hechos en el ámbito universitario y apoyar la toma de decisiones, mejorando la satisfacción de los alumnos y la calidad de los servicios de los cursos **(Sanz & Miguel, 2010)**.

Una propuesta similar a la anterior se plantea en **(Escuela Colombiana de Carreras Industriales, 2013)**, dónde se muestra cómo fue el diseño e implementación de un *Data Mart* siguiendo la metodología de *Kimball*, para consolidar los datos de diferentes fuentes y a partir de estos realizar estrategias que logran disminuir los niveles de deserción e identificar factores riesgo para plantear estrategias de acompañamiento.

Para mejorar los procesos que se llevaban a cabo en el departamento de admisiones y registro de la Universidad de Tarapacá en Chile, en **(Fuentes Tapia & Valdivia Pinto, 2010)** se muestra la inclusión de herramientas y técnicas de inteligencia de negocio, para implementar un *Data Mart*, a partir de un modelo dimensional siguiendo la metodología de Kimball.

Para complementar el proyecto se desplegó un tablero de control dinámico, para la visualización de los datos. Este proyecto permitió establecer una fuente única y consolidada de datos, que logró resolver los problemas del área de admisiones de la institución.

En Colombia también es posible encontrar trabajos relacionados con implementaciones de soluciones de inteligencia de negocios en ambientes institucionales, como las que se explican a continuación:

La solución de inteligencia de negocios implementada en la Universidad del Magdalena, se realizó con el objetivo de optimizar la gestión de recursos académicos y espacios físicos. Permitiendo la creación de reportes con datos actuales e históricos de procesos y apoyar la toma de decisiones de compra de nuevos recursos y predicción de su uso para mejorar la disponibilidad de estos.

En el desarrollo de esta implementación se utilizó el *software SQL Server 2008 R2*, a partir del cual se creó la bodega de datos, estructurada como un esquema de estrella y luego para acceder a la información y crear reportes se utilizaron los programas *MS Excel 2010* y *MS Reporting Services 2008 R2*.

El despliegue de este proyecto permitió crear reportes, que aseguraron y mejoraron la eficiencia y eficacia de la gestión de recursos de la institución, aumentando la satisfacción de los usuarios del sistema **(Narváez Triana, Monsalve Hernández, Bustamante Martínez, Galvis Lista, & Gómez Flórez, 2013)**.

En el caso puntual de la ciudad de Medellín, la propuesta desarrollada por investigadores de la universidad del CES, presenta la aplicación denominada

ELOGIM desarrollada por el personal de la biblioteca. La implementación de esta aplicación basada en técnicas de minería de datos, permitió elaborar informes, que brindaran información sobre el uso de los recursos electrónicos y así dar apoyo al proceso de toma de decisiones de la unidad de información.

Esta aplicación permitió conocer el comportamiento informacional de los usuarios de la biblioteca y los recursos implicados, arrojando resultados que permitieron reducir los presupuestos asociados a la adquisición de nuevos recursos y mejorar la identificación y niveles de satisfacción de los mismos **(Quiroz Gil & Valencia, 2012)**.

3.3.3 Aplicación de técnicas y herramientas de inteligencia de negocios para análisis de la producción científica

En el trabajo experimental realizado por Roberto López Martínez, se identifican los patrones presentes en la relación academia-industria mediante la aplicación de técnicas de minería de datos y análisis semánticos, aplicados a un conjunto que contiene datos de la producción científica mexicana y las patentes existentes. Este estudio se realizó con el fin de poder obtener conocimiento que permitiera tomar decisiones adecuadas para estructuras políticas en el ámbito de ciencia y tecnología mexicano **(López Martínez, 2011)**.

Otro ejemplo de análisis de producción científica con herramientas de inteligencia de negocios, se muestra en el estudio realizado por Oscar Saavedra, donde se analiza la producción científica del campo agrícola de América Latina y el Caribe. Este estudio se realizó a través del uso de 22 indicadores de ciencia y tecnología, con el fin de conocer la problemática que las ciencias agrícolas presentan en estas zonas, y de tal manera lograr obtener un mejor conocimiento de la realidad regional. El análisis de los datos se realizó con la ayuda de herramientas de minería de datos. Dichas herramientas de minería permitieron el análisis exploratorio de los datos y la formación de clústeres o grupos conformados por países y otros componentes, los cuales se relacionaron a través del uso de redes neuronales. Este estudio permitió conocer los países con mayor actividad de edición y producción de revistas **(Saavedra Fernández, Sotolongo Aguilar, & Guzmán Sánchez, 2002)**.

Por último y enfatizando en la utilización de técnicas de inteligencia de negocios para análisis de producción científica, Ariagna Martínez Cuesta en compañía de otros investigadores, presentan un estudio en el cual se busca gestionar el

conocimiento generado por el ICIMAF (Instituto de Cibernética, Matemática y Física) de Cuba, y conocer el impacto de su producción científica en el Programa Nacional de Ciencias Básicas. En este trabajo se propone la creación de una metodología basada en bibliominería y bibliometría, la cual permita valorar la calidad de la producción científica usando indicadores bibliométricos y agrupando los documentos que están relacionados entre sí, facilitando la creación de clúster. Para tal fin se utilizan algoritmos de indexación y el algoritmo de agrupamiento *K-medias*, luego de definir estos algoritmos se recurre a la herramienta minería de datos WEKA para analizar los datos almacenados en la base de datos *JabRef* (Martínez Cuesta & Plasencia Salgueiro, 2010).

3.3.4 Análisis del estado del arte

Los diferentes trabajos presentados, permiten reconocer inicialmente que las actividades de análisis y medición de la producción científica en las instituciones académicas, son procedimientos que están siendo aplicados en diferentes universidades tanto a nivel internacional como nacional, ya que se transforman en ventaja competitiva y permiten mejorar los niveles en los rankings a partir de las estrategias planteadas en el proceso. Estos estudios sirvieron como guía para el desarrollo de este proyecto, ya que permiten conocer los procedimientos necesarios llevados a cabo para adquirir conocimiento e identificar cómo ha sido el desempeño de la actividad investigativa en la institución.

Los trabajos asociados a esta temática, muestran las ventajas obtenidas a partir de la implementación de las soluciones de inteligencia de negocios y de igual manera detallan paso a paso cómo fue el despliegue de toda la solución, incluyendo cuales fueron las herramientas, métodos y procesos utilizados para lograr resultados los obtenidos. Estos trabajos fueron base teórica para el desarrollo de este proyecto, ya que a partir de estos se pudo conocer los autores y trabajos más destacados. Puede verse que las soluciones de inteligencia de negocios, son ampliamente aceptadas en los ámbitos universitarios, enfatizando principalmente en la construcción de bodegas de datos y *Data Marts* siguiendo la metodología Kimball.

La revisión del estado del arte, también permite concluir que aunque existen empresas que ofrecen servicios para el análisis y seguimiento de la producción científica de las instituciones, varias instituciones optan por realizar un desarrollo propio ya que al comprar los servicios, la institución no tiene control sobre los datos y no es posible enriquecer el contenido de la información con datos de

interés de la propia institución, por ejemplo, grupo de investigación, área a la que pertenece (escuela, facultad, programa), nivel de información de los autores, etc. Adicionalmente, al tener el control de los datos, sería posible a futuro realizar análisis de minería de datos buscando patrones sobre la productividad de los investigadores y otros análisis interesantes.

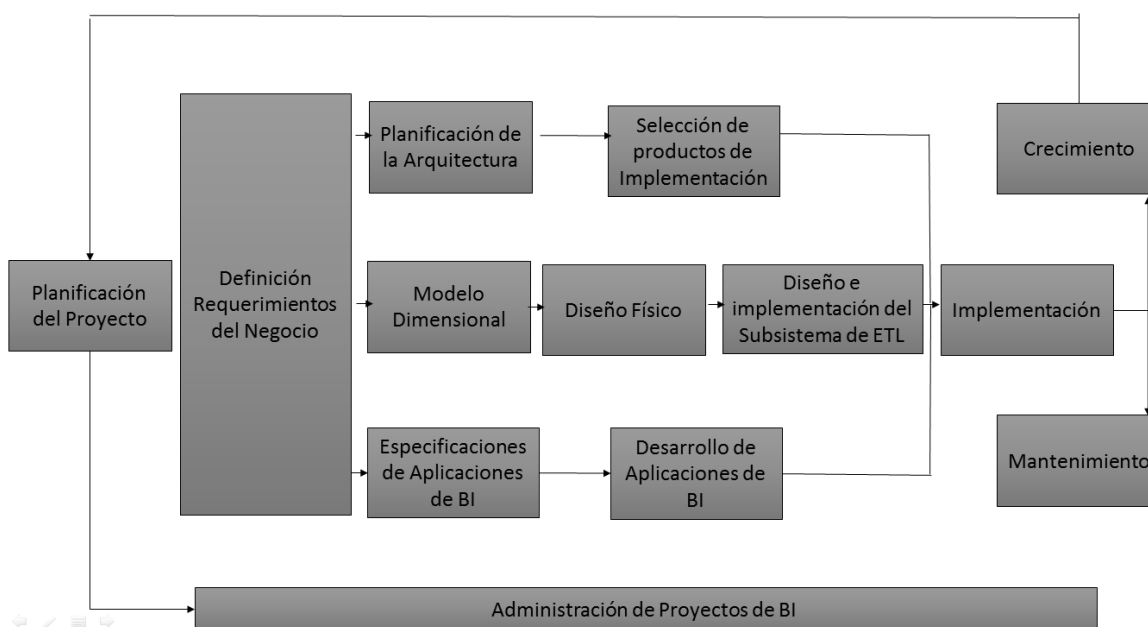
4 METODOLOGÍA

A continuación, se describe la metodología que se trabajó en el presente proyecto.

Para este proyecto se optó por la metodología propuesta por Ralph Kimball, ya que esta es más adecuada en proyectos similares al que se ejecutó, dónde se parte de la construcción de data marts enfocados a necesidades de departamentos específicos de la organización, para luego a partir de estos brindar la oportunidad de construir un data warehouse institucional (1keydata, 2015).

En la Figura 2, se presenta la metodología Kimball, en la cual se basó la ejecución de este proyecto.

Figura 2: Adaptación de la metodología de Kimball para el análisis y visualización de las publicaciones científicas de la Universidad Pontificia Bolivariana



Fuente: (Kimball & Margy, 2010)

A continuación, se explica la aplicación de la metodología al caso particular desarrollado.

Fase 1 Planificación del Proyecto: en esta etapa, se determinó el propósito del proyecto, definiendo de manera clara porqué es necesario y cuáles beneficios obtendrá la institución con el desarrollo de este trabajo de grado. También se definió el alcance y se diseñó el cronograma para alcanzar los objetivos propuestos.

Como parte de esta fase, se realizó una búsqueda de trabajos relacionados con la temática a tratar para conocer más a fondo cómo fueron desarrollados y los resultados obtenidos.

Fase 2 Definición de los requerimientos del negocio: en esta fase se realizó el levantamiento de los requerimientos y necesidades de los futuros usuarios de nuestra solución.

Para el levantamiento de la información, inicialmente se identificaron los usuarios claves para encontrar las necesidades y oportunidades de mejoramiento, luego se entrevistó al personal necesario y finalmente se documentó y analizó los datos recolectados.

Para complementar esta etapa, fue necesario investigar qué datos se podían obtener desde las diferentes bases de datos académicas e instituciones y cómo será el proceso de extracción requerido para robustecer nuestra base de datos.

Fase 3 Modelado multidimensional: según los requerimientos documentados en la fase anterior, se analizaron datos reales para definir así el nivel de granularidad, las dimensiones y los atributos de las mismas, junto con las métricas y tablas de hechos a utilizar. Igualmente, en esta fase se construyeron los *Star nets* necesarios para proporcionar un mayor entendimiento del modelo de cara al negocio.

Fase 4 Diseño Físico: en esta fase, se definieron los requerimientos técnicos necesarios que cada uno de los miembros del proyecto debía tener para realizar el montaje del modelo multidimensional diseñado en la etapa anterior en la herramienta *Qlik Sense*.

Fase 5 Diseño e Implementación del Subsistema de ETL: en esta fase se realizó el proceso de *ETL* para cargar los datos en el modelo definido en la etapa anterior, de acuerdo con las relaciones y estructuras definidas en la fase 3.

Fase 6 Especificaciones de las Aplicaciones de BI: en esta fase se definió cual sería la aplicación de Inteligencia de Negocios a utilizar, con el fin de permitir a los usuarios un acceso más fácil a los datos almacenados en el modelo multidimensional.

Fase 7 Desarrollo de Aplicaciones de BI: en esta fase se diseñó el tablero de control para plasmar los resultados de las consultas hechas al modelo. El diseño de este elemento se realizó con base en la metodología *DAR (Dashboard, Analysis, Reporting)* sugerida por Qlik, la cual consiste en el desarrollo de un tablero de control, páginas de análisis y páginas de reportes.

Fase 8 Implementación: en esta fase se implementó el tablero de control, desde el cual se obtuvieron las métricas y se analizaron de acuerdo con las dimensiones establecidas en las fases anteriores. Todo esto con el fin de visualizar la información de diversas maneras y de acuerdo con las necesidades de los diferentes usuarios.

5 PRESENTACIÓN Y ANÁLISIS DE RESULTADOS

5.1 IMPLEMENTACIÓN DE LA SOLUCIÓN

5.1.1 Levantamiento de requisitos

Objetivo 1: realizar el levantamiento de requisitos para el análisis de las publicaciones científicas

Para dar inicio a la solución del problema planteado anteriormente en este documento, se procedió a dar cumplimiento al primer objetivo específico, el cual proponía realizar el levantamiento de requisitos para el análisis de las publicaciones científicas. Por ende, con el fin de cumplir las necesidades de información, fue fundamental identificar los requisitos necesarios para el desarrollo e implementación del modelo de datos, que conllevó a la creación de visualizaciones para un posterior análisis de las publicaciones científicas indexadas en la base de datos de *Scopus*.

Inicialmente se realizaron encuestas a personal de Vicerrectoría Académica, del Centro de Investigación para el Desarrollo y la Innovación (CIDI) y de la Coordinación de Investigaciones de Ciencias de la Salud, mediante encuestas estructuradas que permitieron identificar los puntos claves para el correcto análisis de las publicaciones científicas.

En cada una de las encuestas se realizó en siguiente procedimiento:

Primero se planteó la siguiente introducción, que se basaba en presentar el proyecto de investigación.

- Introducción: el programa de gestión al día tiene como reto garantizar la productividad, competitividad y servicio al cliente del sistema de investigación y transferencia e innovación de la UPB.

Luego se presentó la estructura de cada encuesta de la siguiente manera:

- Se realizaron encuestas con cada una de las personas seleccionadas: En la primera se buscaba recopilar la información y en la segunda se pretendía presentar un prototipo de los requisitos para conocer la opinión del usuario entrevistado

Finalmente, la encuesta realizada fue la siguiente:

- ¿Cuáles son sus indicadores claves para la producción científica?
- ¿Qué reportes genera a partir de la información de la base de datos *Scopus*?
- ¿Qué datos son los más importantes del reporte de la producción científica de alto impacto indexada en *Scopus*?
- ¿Para qué utiliza esta información? ¿Cuál son los usos?
- ¿Hay cuellos de botella para obtener la información?
- ¿Cuánta información histórica es requerida?
- ¿Qué oportunidades de mejora existen en lo referente al análisis de información para toma de decisiones?
- ¿Cuáles son los aspectos que encuentran más difíciles en el proceso actual y qué considera que puede ser cambiado para una mejoría de estos?

Los usuarios encuestados se presentan a continuación en la Tabla 1 y las encuestas se encuentran en el ANEXO 1.

Tabla 1: Usuarios encuestados
Elaborado por los Autores

Dependencia	Cargo	Nombre
Dirección de Investigación e Innovación-CIDI	Coordinadora de Investigaciones	Marcela Restrepo Rodas
Vicerrectoría Académica	Coordinadora de visibilidad académica e impacto científico	Maricela Gómez Vargas
Escuela de Ciencias de la Salud	Coordinadora de Investigación Salud	Zulma Vanessa Rueda

A partir de la información capturada en las diferentes encuestas se identificaron los siguientes requisitos funcionales y no funcionales presentados en la Tabla 2 y en la Tabla 3:

Tabla 2: Requisitos Funcionales del sistema
Elaborado por los Autores

Requisitos Funcionales	Identificador
El sistema deberá permitir conocer el número de productos indexados vinculados a la institución	1
El sistema deberá permitir conocer el número de productos indexados por cada investigador	2
El sistema deberá permitir conocer el número de citaciones de los productos científicos indexados	3
El sistema deberá permitir conocer el número de investigadores de la institución con publicaciones indexadas	4
El sistema deberá permitir conocer las fuentes de publicación de los productos indexados	5
El sistema deberá permitir conocer el número de productos según nivel profesional de los investigadores	6
El sistema deberá permitir conocer el número de productos según la vinculación del investigador	7
El sistema deberá permitir conocer cómo ha sido la evolución de las publicaciones científicas a través de los años 2009-2015.	8
El sistema deberá permitir conocer los cuartiles, categorías y países asociados a las fuentes de publicación	9
El sistema deberá permitir conocer los niveles de producción científica de los grupos de investigación asociados a los investigadores	10
El sistema deberá permitir conocer los diferentes tipos de documentos publicados en las base de datos	11
El sistema deberá permitir conocer en que idiomas se ha realizado la producción científica	12

Tabla 3: Requisitos No Funcionales del sistema
Elaborado por los Autores

Requisitos No Funcionales	Identificador
La información de las publicaciones indexadas será extraída de la base de datos <i>Scopus</i> entre los años 2009-2015	13
La información de los docentes y estudiantes se complementará con datos del sistema Banner	14
La información faltante de los docentes y estudiantes se complementará a través de la plataforma CvLAC de Colciencias y correos internos	15
Para la creación de los tableros de control se utilizará el software analítico de Qlik Sense	16
Para el proceso de ETL se utilizará el software analítico de Qlik Sense	17
Para complementar la información de los grupos de investigación se extraerá información de la plataforma ScienTi	18

Requisitos No Funcionales	Identificador
de Colciencias	
Para complementar los datos asociados a los cuartiles de las fuente de publicación se extraerá información de la plataforma SCImago	19

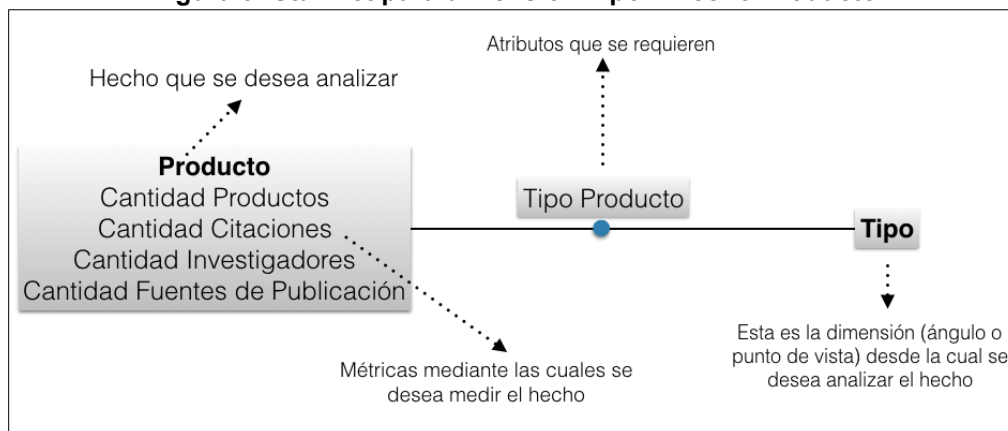
5.2 DISEÑO DEL MODELO MULTIDIMENSIONAL DE DATOS

Objetivo 2: realizar el diseño del modelo multidimensional para las publicaciones científicas de la Universidad Pontificia Bolivariana.

Con base en lo anterior y con el fin de visualizar de una manera más sencilla el modelo lógico planteado, se construye un *Star Net*, el cual es un diagrama que permite definir los requerimientos, identificar las métricas y las variables del negocio (Bermúdez García, 2014).

De la Figura 3 a la Figura 8, presentan los *Star Nets* realizados para el “Hecho” que se desea analizar con sus dimensiones y atributos respectivos.

Figura 3: *Star Net* para dimensión Tipo – Hecho Producto



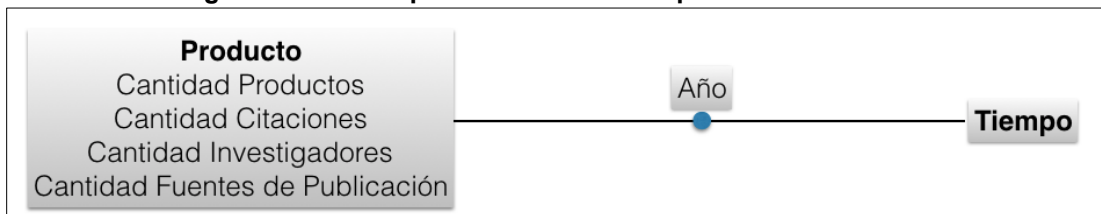
Fuente: Elaborado por los Autores

Figura 4: Star Net para dimensión Idioma– Hecho Producto



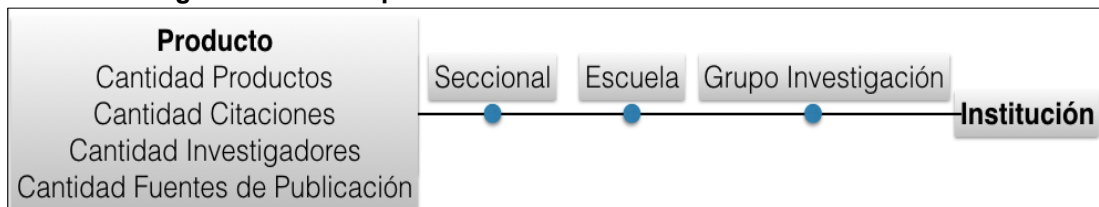
Fuente: Elaborado por los Autores

Figura 5: Star Net para dimensión Tiempo– Hecho Producto



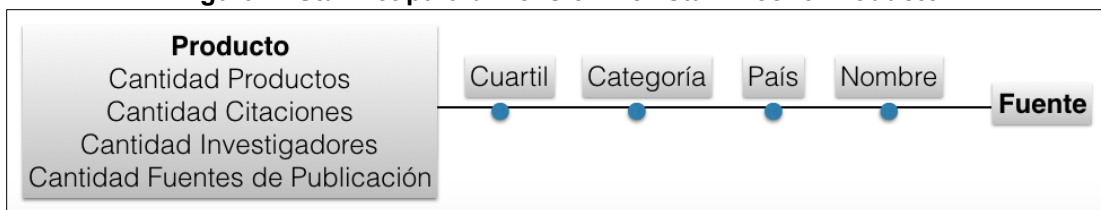
Fuente: Elaborado por los Autores

Figura 6: Star Net para dimensión Institución – Hecho Producto



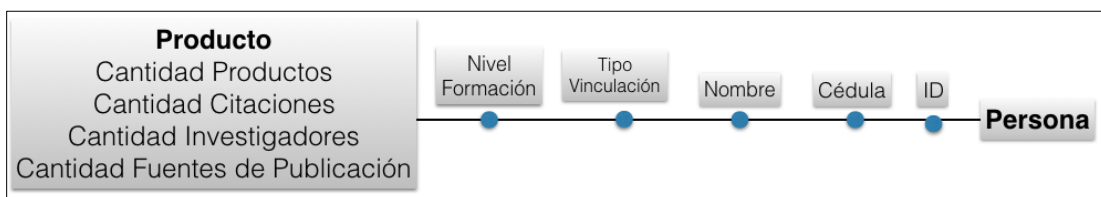
Fuente: Elaborado por los Autores

Figura 7: Star Net para dimensión Revista – Hecho Producto



Fuente: Elaborado por los Autores

Figura 8: Star Net para dimensión Persona – Hecho Producto



Fuente: Elaborado por los Autores

Luego para el diseño del modelo multidimensional físico, se hace uso de la metodología Kimball en la cual se basa este proyecto y cuyo proceso consta de los siguientes pasos:

5.2.1 Definir el proceso de negocio

En el caso de la Universidad Pontificia Bolivariana, no se cuenta con un sistema que permita conocer y cuantificar correctamente las publicaciones científicas indexadas en la base de datos *Scopus*. En este proyecto se busca recolectar toda la información existente, procesarla y utilizarla para la creación de un modelo que permita a los usuarios conocer e interactuar con la información referente al desempeño de las publicaciones científicas de alto impacto de la institución. En general el modelo multidimensional planteado permitirá interactuar con la información asociada a las publicaciones científicas de alto impacto desde diferentes dimensiones estructuradas, todo esto con el fin de generar conocimiento para poder comprender la situación actual de la actividad y poder plantear estrategias de mejoramiento.

5.2.2 Seleccionar el nivel de granularidad

Para realizar este proyecto dando cumplimiento a las necesidades del negocio, se define que el menor nivel de granularidad (mayor detalle) será dado por producto indexado, a partir de la definición de este “hecho” se adquieren mayores posibilidades analíticas desde cada una de las dimensiones asociadas.

5.2.3 Definir las dimensiones

Teniendo en cuenta la granularidad definida en el punto anterior, se establecen que las dimensiones desde las cuales podrá ser analizado, filtrado y agrupado el “hecho” establecido serán:

- Investigador (Dim_Investigador): esta dimensión permitirá conocer los detalles del investigador.

Atributos: nombre, cédula, vinculación, nivel de formación, seccional, escuela y grupo de investigación.

- Producto (Dim_Producto): esta dimensión permitirá conocer el nombre y el tipo de documento indexado; Artículo, Documento de Conferencia, Reseña, Carta, Artículo en prensa, Editorial, Capítulo de libro, Nota y Encuesta corta.

Atributos: nombre, citas y tipo del producto

- Revista (Dim_Fuente): esta dimensión permite conocer el nombre y el país asociado a la fuente de publicación del producto indexado.

Atributos: nombre de la fuente de publicación, país, categoría y cuartil

- Idioma (Dim_Idioma): esta dimensión permitirá analizar el hecho según el idioma en el cual fue publicado; español, inglés, francés entre otros.

Atributos: idiomas del producto indexado.

- Tiempo (Dim_Tiempo): esta dimensión proporcionará los valores temporales desde los cuales será analizado el hecho.

Atributos: años de publicación del producto indexado.

5.2.4 Identificar las métricas

El hecho Productos se medirá mediante las métricas:

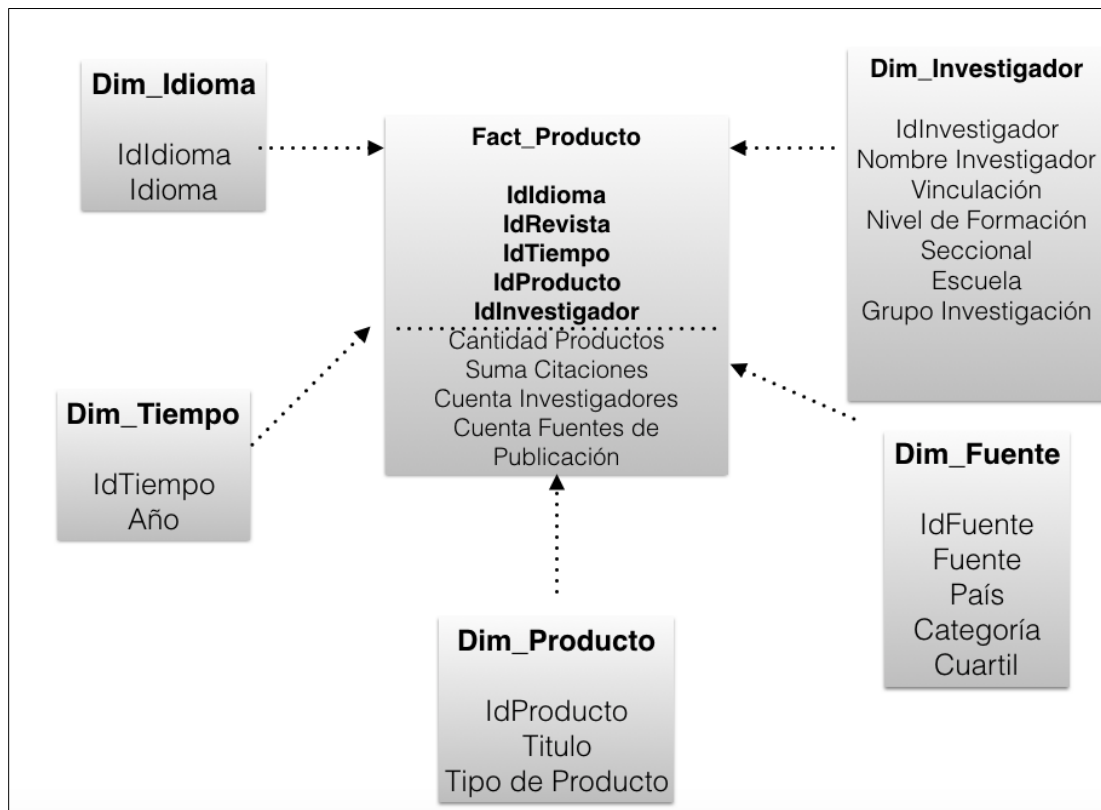
- Cantidad de productos: para conocer permanentemente la cantidad de productos indexados.
- Número de citas: para conocer permanentemente la cantidad de citas obtenidas por cada producto.
- Número de Investigadores: para conocer el número de investigadores con productos indexados.
- Número de Fuentes de Publicación: para conocer la cantidad de fuentes de publicación de los productos indexados.

5.2.5 Definir la tabla de Hechos (Fact_Producto)

Esta tabla consolida las diferentes llaves para establecer las relaciones entre las dimensiones y calcular las métricas definidas anteriormente

La Figura 9, presenta el modelo de datos multidimensional diseñado.

Figura 9: Modelo de datos multidimensional diseñado



Fuente: Elaborado por los Autores

Para estructurar el diseño del proceso de Extracción, Transformación y Carga, se parte de las siguientes actividades fundamentales:

1. Definir cuáles van a ser las herramientas y plataformas que van a hacer parte de la solución de inteligencia de negocios, para este trabajo se utilizará:
 - Microsoft Excel: para la consolidación de los datos, a partir de los cuales se crearon las tablas que fueron usadas como fuentes de datos para la herramienta de inteligencia de negocios.
 - Qlik Sense: para la implementación de los procesos de *ETL* y generación del tablero de control.
2. Identificar las fuentes de datos iniciales de donde van a ser capturados los datos necesarios y definir la manera en que estos serán consolidados.

Para realizar dicha actividad se utilizó la base de datos bibliográfica *Scopus*, mediante la cual se realizó una búsqueda por afiliación a la Universidad Pontificia Bolivariana. Esta consulta permitió extraer un archivo en formato separado por comas (CSV, por sus siglas en inglés) con 1046 registros por

producto indexado para un periodo de búsqueda desde el año 2009 hasta el año 2015, este archivo contenía entre otras, las siguientes variables:

- Firma de los Autores
- Título
- Año
- Fuente de Publicación
- Citaciones
- Afiliaciones de los Autores
- Idioma del Documento
- Tipo de Documento
- Identificador Digital del Producto (EID)

Para la identificación de los investigadores mediante el campo Afiliaciones de los Autores, fue necesario duplicar el registro de cada producto tantas veces como firmas de autores asociados a la Universidad Pontificia Bolivariana tuviese, para así lograr de manera básica identificar la escuela, seccional y firma del autor, en los registros en los que fuera posible.

A partir de la identificación anterior, se procede de manera manual para obtener el nombre completo del investigador, realizando búsquedas exhaustivas de los diferentes productos en la misma base de datos *Scopus* y páginas de internet. Se debió consultar en la misma base de datos ya que cuando se realiza la extracción de los datos, la plataforma no ofrece la posibilidad de obtener el nombre completo del investigador o el número identificador del autor utilizado por *Scopus*. Esta actividad fue la que más tiempo consumió ya que fue necesaria la ayuda de docentes y empleados administrativos de la institución, para así lograr identificar por nombre completo a los investigadores.

Luego de tener el nombre completo de cada investigador, se procede a realizar consultas en el sistema Banner para así enriquecer los datos iniciales con datos como; el nivel de formación, cédula y tipo de vinculación a la institución, se debe resaltar que para algunos investigadores fue necesario realizar más procesos de búsqueda a la plataforma ScienTI, a través del aplicativo CvLAC, para lograr así completar todos los campos necesarios como seccional y escuela. Adicionalmente fue necesario depurar los datos provenientes del

sistema Banner, ya que se encontraron duplicados debido a los diferentes cambios de vinculación de un investigador con la institución.

Para completar la información de los grupos de investigación a los que pertenecían los investigadores, se realizaron nuevamente búsquedas manuales a la plataforma ScienTI, para recopilar la información sobre los grupos de investigación pertenecientes a la Universidad Pontificia Bolivariana.

Después, para consolidar la información necesaria de las fuentes de publicación, se realizaron consultas a la plataforma SCImago para conocer los cuartiles por año asociados a las diferentes fuentes, las categorías de dichas fuentes de información y los países. La extracción de los campos anteriores se efectuó de manera manual al igual que la consolidación con los datos existentes.

Finalmente, para consolidar todos los datos necesarios y verificar su integridad mediante procesos de calidad de datos, se procedió a eliminar ciertos registros que tenían las afiliaciones equivocadas, es decir, sus autores en realidad no eran de la institución.

Toda la información recopilada fue estructurada a través de Excel mediante la creación de 8 tablas (Tabla 4 a 11), las cuales se utilizaron como fuentes de datos para el modelo multidimensional.

Tabla 4: Integrantes de Grupos de Investigación
Elaborado por los Autores

Fuente: Int Grupos.xlsx	
Escuela	Escuela a la cual pertenece el grupo de investigación
Grupo de Investigación	Nombre del grupo de investigación
Cédula	Cédula del investigador

Tabla 5: Investigadores
Elaborado por los Autores

Fuente: Personas.xlsx	
NIT	Cédula del investigador
ID	ID institucional del investigador
PRIMER_APELLIDO	Apellidos del investigador
NOMBRE	Nombres del investigador
TIPO_EMPLEADO	Tipo de vinculación con la Universidad
Nivel Formacion	Descripción del nivel de formación

Tabla 6: Artículos
Elaborado por los Autores

Fuente: Articulos.xlsx	
Seccional	Seccional a la cual pertenece el investigador
Escuela	Escuela a la cual pertenece el investigador
Título	Título del Producto
Cedula	Cédula del Investigador
Apellidos	Apellidos del Investigador
Nombre	Nombre del Investigador
Year	Año de publicación del producto
Source title	Nombre de la fuente de publicación de producto
Cited by	Número de citaciones del producto
Lenguaje of original document	Idiomas de publicación
IdCategoria	Identificador de la categoría del idioma
Doument Type	Tipo de documento
EID	Identificador digital del producto

Tabla 7: Países de la Fuente de Publicación
Elaborado por los Autores

Fuente: PaisesRevista.xlsx	
Revista	Nombre de la fuente de publicación
País	País de la fuente de publicación

Tabla 8: Descripción de los Niveles de Formación de los Investigadores
Elaborado por los Autores

Fuente: NivelFormación.xlsx	
Código	Código del nivel de formación
Descripción	Descripción del nivel de formación

Tabla 9: Cuartiles de las Fuentes de Publicación
Elaborado por los Autores

Fuente: Cuartiles.xlsx	
Revista	Nombre de la fuentes de publicación
2009	Número de cuartil para el año específico
2010	Número de cuartil para el año específico
2011	Número de cuartil para el año específico
2012	Número de cuartil para el año específico
2013	Número de cuartil para el año específico
2014	Número de cuartil para el año específico
2015	Número de cuartil para el año específico

Tabla 10: Categorías de las Fuentes de Publicación
Elaborado por los Autores

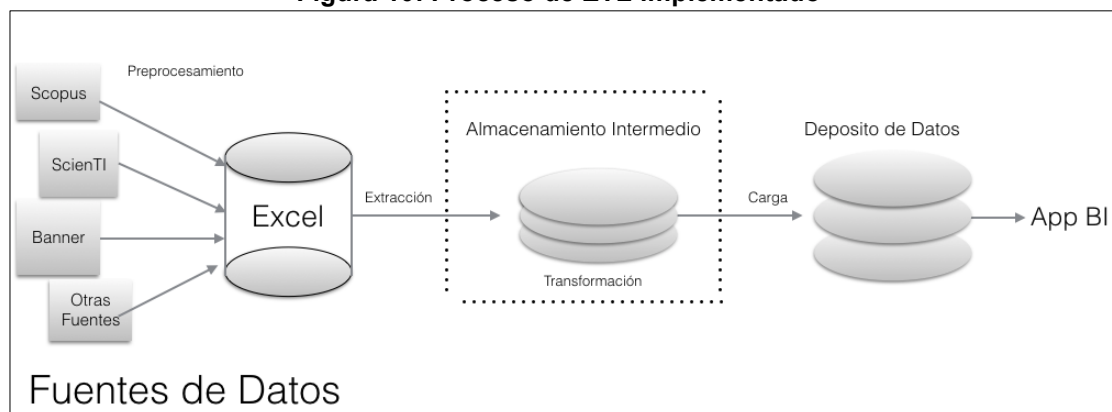
Fuente: Categorías.xlsx	
Revista	Nombre de la fuentes de publicación
Categoría	Categorías de las fuentes de publicación

Tabla 11: Idiomas de Publicación de los Productos
Elaborado por los Autores

Fuente: Idioma.xlsx	
IdCategoria	Identificador de la categoría del idioma
Descripción	Descripción de los idiomas de publicación

La Figura 10, resume el proceso realizado para la ejecución del proceso de ETL que se implementó.

Figura 10: Proceso de ETL implementado



Fuente: Elaborado por los Autores

5.3 IMPLEMENTACIÓN DEL MODELO MULTIDIMENSIONAL DE DATOS

Objetivo 3: realizar la implementación del modelo multidimensional para las publicaciones científicas de la Universidad Pontificia Bolivariana.

Partiendo de la consolidación de datos lograda en el punto anterior, se procede a la creación de capas a través de aplicaciones en la herramienta *Qlik Sense* para iniciar el proceso de *ETL* y finalizar con la carga de datos al modelo multidimensional, el código implementado se presenta en el

ANEXO 2.

A continuación, se explica el procedimiento realizado en cada una de las capas:

5.3.1 Proceso de extracción

El objetivo de esta capa es iniciar el proceso de ETL, mediante la creación en *Qlik Sense* de tablas equivalentes a las tablas que son fuentes de datos (Tabla 12 a la

Tabla 19), para este caso los archivos son de tipo *xlsx*. De esta manera se crea un área de almacenamiento intermedio con el fin de minimizar riesgos a la hora de hacer la carga de datos.

Para realizar este proceso se crean 2 conexiones tipo carpeta, estas conexiones definen de cuál fichero se van a cargar los datos a la tabla temporal y en cuál fichero se van a guardar los datos una vez sean cargados a la tabla.

Una vez se hayan cargado los datos a la tabla temporal, serán almacenados en un archivo tipo *qvd*, el cual permite consolidar un único formato para los datos y así poder utilizarlos en las diferentes aplicaciones que los requieran.

Este proceso de carga se realizó para cada uno de los 8 archivos fuentes.

Tabla 12: Tabla equivalente para Integrantes de Grupos de Investigación
Elaborado por los Autores

Grupos		
Fuente	Atributos de Extracción	Destino Extracción
Int Grupos.xlsx	Escuela	Grupos.qvd
	Grupo de Investigación	
	Cédula	

Tabla 13: Tabla equivalente para Investigadores
Elaborado por los Autores

Personas		
Fuente	Atributos de Extracción	Destino Extracción
Personas.xlsx	ID	Personas.qvd
	NIT	
	PRIMER_APELLIDO	
	NOMBRE	
	TIPO_EMPLEADO	
	Nivel Formación	

Tabla 14: Tabla equivalente para Artículos
Elaborado por los Autores

Artículos		
Fuente	Atributos de Extracción	Destino Extracción
Articulos.xlsx	Seccional	Articulos.qvd
	Escuela	
	Titulo	
	Cedula	
	Apellidos	
	Nombre	
	Year	
	Source Tittle	
	Cited by	
	Lenguaje of original document	
	IdCategoria	
	Doument Type	
	EID	

Tabla 15: Tabla equivalente para Países de la Fuente de Publicación
Elaborado por los Autores

Países Revista		
Fuente	Atributos de Extracción	Destino Extracción
PaisesRevista.xlsx	Revista	PaisesRevistas.qvd
	País	

Tabla 16: Tabla equivalente para Niveles de Formación de los Investigadores
Elaborado por los Autores

Niveles de Formación		
Fuente	Atributos de Extracción	Destino Extracción
Nivel Formacion.xlsx	Codigo	NivelFormacion.qvd
	Descripcion	

**Tabla 17: Tabla equivalente para Cuartiles de las Fuentes de Publicación
Elaborado por los Autores**

Cuartiles		
Fuente	Atributos de Extracción	Destino Extracción
Cuartiles.xlsx	Revista	Cuartiles.qvd
	Años	
	Cuartiles	

**Tabla 18: Tabla equivalente para Categorías de las Fuentes de Publicación
Elaborado por los Autores**

Categorías		
Fuente	Atributos de Extracción	Destino Extracción
Categorías.xlsx	Revista	Categoria.qvd
	Categoria	

**Tabla 19: Tabla equivalente para Idiomas de Publicación de los Productos
Elaborado por los Autores**

Idioma		
Fuente	Atributos de Extracción	Destino Extracción
Idioma.xlsx	IdCategoria	Idioma.qvd
	Descripción	

5.3.2 Proceso de transformación

El objetivo de esta capa es iniciar la estructuración del modelo de datos a partir de los datos extraídos en la capa anterior. En esta capa se da inicio a la creación de las dimensiones, atributos y tablas de hechos requeridas para definir las reglas de negocio y cumplir con las necesidades de información.

En esta capa también se crean dos conexiones similares a las anteriores las cuales especificarán de cuál ruta serán extraídos los datos cargados previamente y en cuál ruta se almacenarán al terminar la configuración de esta capa, de igual manera se utilizan tablas temporales para la carga de los datos.

Para dar inicio al proceso de transformación, se comienzan a cargar las tablas que serán utilizadas como dimensiones y tablas de hechos, en cada una de estas, se cargarán datos provenientes de los archivos tipo qvd. Estas nuevas tablas de dimensiones contarán con un identificador único y una serie de atributos necesarios para establecer las relaciones con las demás dimensiones.

Las dimensiones creadas fueron (Tabla 20 a la Tabla 28):

Tabla 20: Proceso de Transformación para la Dimensión Fuentes
Elaborado por los Autores

DimFuentes				
Origen Extracción	Atributos de Extracción	Transformación Realizada	Atributos de Transformación	Destino Transformación
PaisesRevista.qvd	Revista	Ninguna	%KeyRevista	DimFuentes.qvd
	País		Pais	

Tabla 21: Proceso de Transformación para la Dimensión Categoría
Elaborado por los Autores

DimCategoría				
Origen Extracción	Atributos de Extracción	Transformación Realizada	Atributos de Transformación	Destino Transformación
Categoría.qvd	Revista	Ninguna	%KeyCategoría	DimCategoríaFuente.qvd
	Categoría		Categoría	

Tabla 22: Proceso de Transformación para la Dimensión Cuartiles
Elaborado por los Autores

DimCuartiles				
Origen Extracción	Atributos de Extracción	Transformación Realizada	Atributos de Transformación	Destino Transformación
Cuartiles.qvd	Año	Concatenación	%KeyCuartil	DimCuartiles.qvd
	Revista			
	Cuartil	Ninguna	Cuartil	

Tabla 23: Proceso de Transformación para la Dimensión Idioma
Elaborado por los Autores

DimIdioma				
Origen Extracción	Atributos de Extracción	Transformación Realizada	Atributos de Transformación	Destino Transformación
Idioma.qvd	IdCategoría	Ninguna	%KeyIdioma	DimIdioma.qvd
	Descripción		Lenguaje	

Tabla 24: Proceso de Transformación para la Dimensión Tipo Producto
Elaborado por los Autores

DimTipoProducto				
Origen Extracción	Atributos de Extracción	Transformación Realizada	Atributos de Transformación	Destino Transformación
Articulos.qvd	Document Type	Distinct	%KeyTipoProducto	DimTipoProducto.qvd

Tabla 25: Proceso de Transformación para la Dimensión Tiempo
Elaborado por los Autores

DimTiempo				
Origen Extracción	Atributos de Extracción	Transformación Realizada	Atributos de Transformación	Destino Transformación
Articulos.qvd	Year	Generación de rango de años	%KeyTiempo	DimTiempo.qvd
			Año	

En nuestro caso la tabla de hechos se conformará por una tabla de hechos y 2 tablas complementarias: Producto e Investigadores

Tabla 26: Proceso de Transformación para la Tabla Complementaria Producto
Elaborado por los Autores

Producto				
Origen Extracción	Atributos de Extracción	Transformación Realizada	Atributos de Transformación	Destino Transformación
Articulos.qvd	Titulo	Distinct	TituloProducto	Producto.qvd
	EID	Ninguna	EID	
	Cited by		Citacion	

Tabla 27: Proceso de Transformación para la Tabla Complementaria Investigadores
Elaborado por los Autores

Elaborado por los Autores					
Investigadores					
Origen Extracción	Atributos de Extracción	Tabla Auxiliar	Transformación Realizada	Atributos de Transformación	Destino Transformación
Personas.qvd	ID	Investigadores	Ninguna	ID	
	NIT		Ninguna	Cedula	
	PRIMER_APELLIDO		Concatenación	NombreCompleto	
	NOMBRE				

Investigadores					
Origen Extracción	Atributos de Extracción	Tabla Auxiliar	Transformación Realizada	Atributos de Transformación	Destino Transformación
	TIPO_EMPLEADO		Ninguna	TipoPersona	Investigadores.qvd
	NivelFormación		Apply Map entre las Tablas Auxiliares Investigadores y TablaNiveles	NivelFormación	
NivelFormación.qvd	Codigo	TablaNiveles			
	Descripcion				
Articulos.qvd	Titulo		Join con Tabla Auxiliar Investigadores	TituloProducto	
	Seccional			Seccional	
	Escuela			EscuelaInvestigador	
	Cedula			Cedula	
Grupos.qvd	Escuela		Join con Tabla Auxiliar Investigadores	EscuelaGrupo	
	Grupo de Investigación			Grupo de Investigación	
	Cédula			Cedula	

Tabla 28: Proceso de Transformación para la Tabla de Hechos
Elaborado por los Autores

FactProducto				
Origen Extracción	Atributos de Extracción	Transformación Realizada	Atributos de Transformación	Destino Transformación
Articulos.qvd	Titulo	Ninguna	TituloProducto	FactProducto.qvd
	YEAR	Ninguna	%KeyTiempo	
	Source Tittle	Ninguna	%KeyRevista	
	Document Type	Ninguna	%KeyDimTipoProducto	
	YEAR	Concatenación	%KeyCuartil	
	Source Tittle			
	Source Tittle	Ninguna	%KeyCategoria	
	IdCategoria	Ninguna	%KeyIdioma	

5.3.3 Proceso de carga

En esta capa finalmente se carga el resultado de la capa de transformación, es decir, las tablas de hechos y dimensiones con sus respectivos atributos. En esta capa quedan los archivos qvd con los que se van a alimentar las consultas del tablero.

De manera similar a los procesos anteriores, en esta capa se crean dos conexiones similares a las anteriores las cuales especificarán de que ruta serán extraídos los datos cargados previamente en la capa de transformación y en que ruta se almacenarán los datos al terminar la configuración de esta capa para ser utilizados finalmente por la aplicación de inteligencia de negocios.

Finalmente las dimensiones utilizadas son (Tabla 29 a la Tabla 37):

- Dimensión Categoría: contiene el nombre de las fuentes de publicación y sus respectivas categorías.

Tabla 29: Dimensión Categoría
Elaborado por los Autores

DimCategoría	
Atributos	Destino
%KeyCategoría	DimCategoríaFuente.qvd
Categoría	

- Dimensión Cuartiles: contiene los cuartiles de cada fuente de publicación por cada uno de los años entre 2009 y 2015.

Tabla 30: Dimensión Cuartil
Elaborado por los Autores

DimCuartiles	
Atributos	Destino
%KeyCuartil	DimCuartiles.qvd
Cuartil	

- Dimensión Idioma: contiene los idiomas en que los productos han sido publicados.

Tabla 31: Dimensión Idioma
Elaborado por los Autores

DimIdioma	
Atributos	Destino
%KeyIdioma	DimIdioma.qvd
Lenguaje	

- Dimensión Fuentes: contiene los nombres de las fuentes de publicación y sus respectivos países.

Tabla 32: Dimensión Fuentes
Elaborado por los Autores

DimFuentes	
Atributos	Destino
%KeyRevsita	DimFuentes.qvd
Pais	

- Dimensión Tiempo: contiene el rango de años de publicación de los productos.

Tabla 33: Dimensión Tiempo
Elaborado por los Autores

DimTiempo	
Atributos	Destino
%KeyTiempo	DimTiempo.qvd
Año	

- Dimensión Tipo de Producto: contiene el tipo de producto publicado.

Tabla 34: Dimensión Tipo Producto
Elaborado por los Autores

DimTipoProducto	
Atributos	Destino
%KeyDimTipoProducto	DimTipoProducto.qvd

- Tabla Complementaria Producto: esta tabla contiene el título de los productos sin duplicados con su respectivo EID, junto con las citaciones recibidas por cada producto a mayo de 2016. Estos atributos permitirán establecer las métricas de cantidad de productos y número de citaciones.

Tabla 35: Tabla Complementaria Producto
Elaborado por los Autores

Producto	
Atributos	Destino
TituloProducto	Producto.qvd
Citacion	
EID	

- Tabla Complementaria Investigadores: esta tabla contiene atributos asociados a los investigadores, los cuales permitirán construir la métrica de número de investigadores:

**Tabla 36: Tabla Complementaria Investigadores
Elaborado por los Autores**

Investigadores	
Atributos	Destino
TituloProducto	Investigadores.qvd
ID	
NombreCompleto	
TipoPersona	
NivelFormación	
Cedula	
Seccional	
EscuelaInvestigador	
EscuelaGrupo	
Grupo de Investigación	

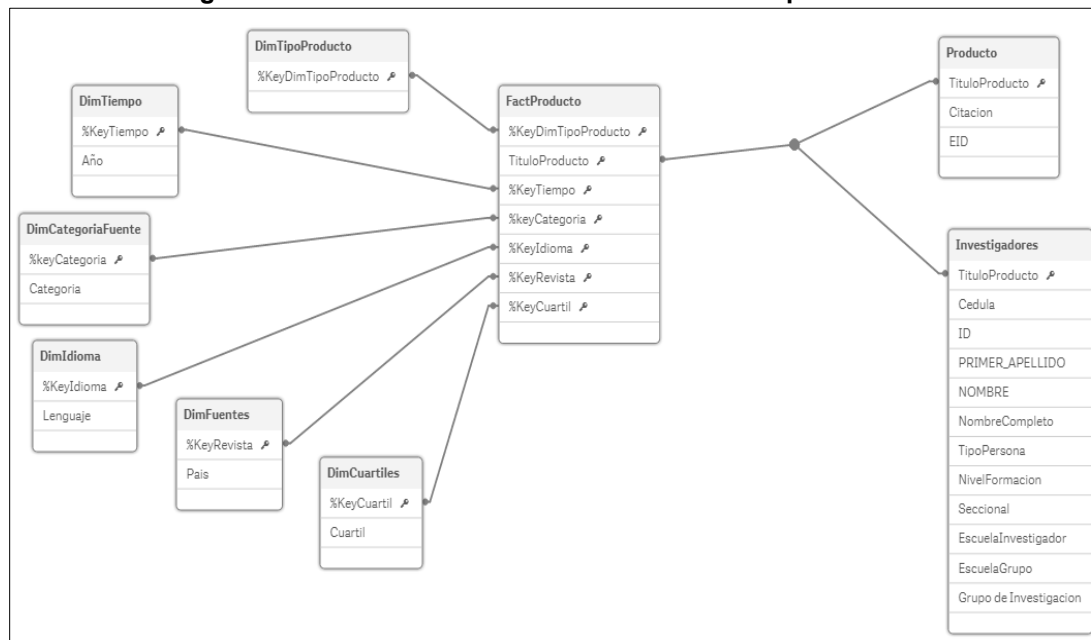
- Tabla de Hechos: esta tabla contiene el título del producto y las diferentes llaves necesarias para realizar las consultas a las dimensiones y tablas complementarias definidas. También permitirá establecer la métrica; número de fuentes de publicación.

**Tabla 37: Tabla Complementaria Investigadores
Elaborado por los Autores**

FactProducto	
Atributos	Destino
TituloProducto	FactProducto.qvd
%KeyTiempo	
%KeyRevista	
%KeyDimTipoProducto	
%KeyCuartil	
%KeyCategoria	
%KeyIdioma	

A continuación, en la Figura 11, se puede observar el modelo de datos multidimensional finalmente implementado, el cual es utilizado para la creación de reportes que darán cumplimiento a las necesidades de información de usuario final.

Figura 11: Modelo de datos multidimensional implementado



Fuente: Elaborado por los Autores

5.4 PRESENTACIÓN DE RESULTADOS

En el presente apartado se realiza una descripción gráfica de los resultados obtenidos a partir de la información extraída. Debido a la alta cantidad de datos a procesar y a la alta diversidad de fuentes de información, se optó por elegir una herramienta especializada en Inteligencia de Negocios (*BI, Business Intelligence* por sus siglas en inglés), conocida como *Qlik Sense*, para agrupar y procesar de una manera clara, ágil y detallada la información obtenida de las bases de datos señaladas anteriormente.

Dentro de la metodología empleada para la extracción, transformación y carga, se destaca que los análisis realizados se efectuaron entre los años 2009 a octubre de 2015, no obstante, se aclara que el proceso debe ser dinámico y actualizado en el tiempo, lo cual se espera que se pueda lograr más adelante de manera sistematizada, con propuestas de trabajos futuros que sean alineados con los resultados plasmados en este trabajo de grado.

A continuación, se explican los resultados, y aún más importante la forma de interpretar y comprender su lectura a partir del entorno gráfico presentado por la herramienta de BI empleada.

La explicación de la metodología, se efectuará mostrando el modelo en sus tres etapas, la primera se conoce como Tablero de control (*Dashboard*), la segunda Análisis (*Analysis*) y la tercera y última Reporte (*Reporting*) (*DAR*, por sus siglas en inglés).

Para hacer más clara la lectura de cada parte del modelo, se realizará una explicación genérica, y el detalle se desglosará en la sección de análisis.

Una vez se presente de manera general los componentes de cada etapa del modelo, se procederá a analizar los resultados más relevantes.

5.4.1 Etapa 1: tablero de control

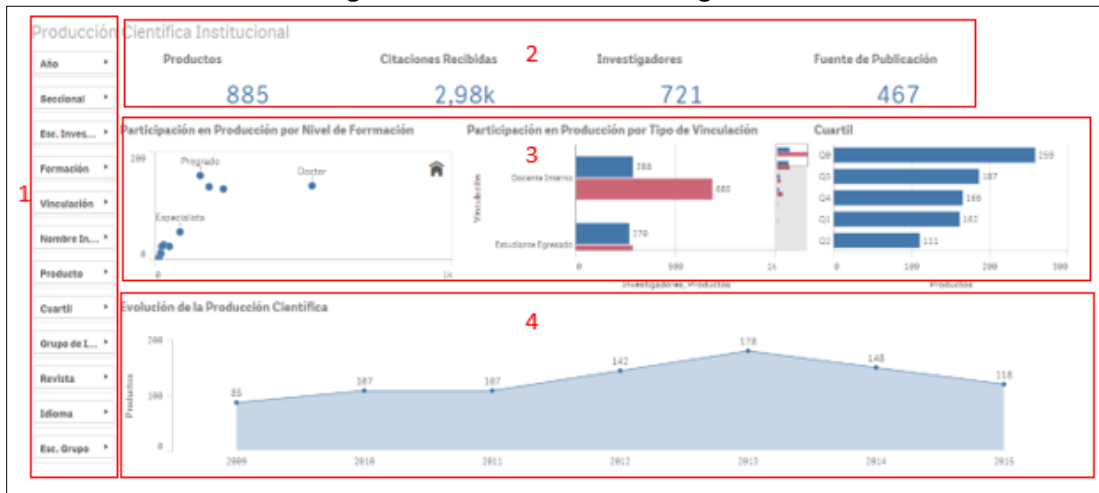
El Tablero de control general se denomina “Producción Científica Institucional”, siendo éste el más universal de todos, ya que permite hacer un escaneo completo del sistema.

La Figura 12 muestra el tablero de control general que incluye todos los elementos gráficos y el análisis de los *Key Performance Indicators* (*KPI* por sus siglas en inglés).

Las secciones son:

- 1 Filtros.
- 2 Indicadores globales (*KPI*)
- 3 Gráficos estadísticos (barras y dispersión).
- 4 Gráfico estadístico de área tendencial.

Figura 12: Tablero de control general



Fuente: Elaborado por los autores a través de la herramienta Qlik Sense

Por ser esta la etapa más representativa, se realizará una descripción detallada de cada sección, de tal forma que pueda ser fácilmente interpretada por el lector.

La sección número 1, contiene todos los tipos de filtros que se pueden lograr, dependiendo del tipo de análisis o lectura que se requiera realizar. Este permite al usuario conocer de manera ágil el estado actual de la información procesada.

La sección número 2, presenta los indicadores globales o también conocidos como *KPI*, estos indican de izquierda a derecha la cantidad de publicaciones realizadas durante el periodo de estudio, el estado actual de las citaciones a mayo de 2016, la cantidad de investigadores que han participado en las diferentes publicaciones (estudiantes docentes, egresados, entre otros) y el último señala el número de fuentes de publicación utilizadas.

Por último, las secciones 3 y 4, presentan las gráficas estadísticas de la información procesada, estas se dan en formatos de dispersión, barras y área.

La elección de estos tipos de gráficos se hizo con base en el tipo de información a visualizar según las recomendaciones de la teoría sobre visualización.

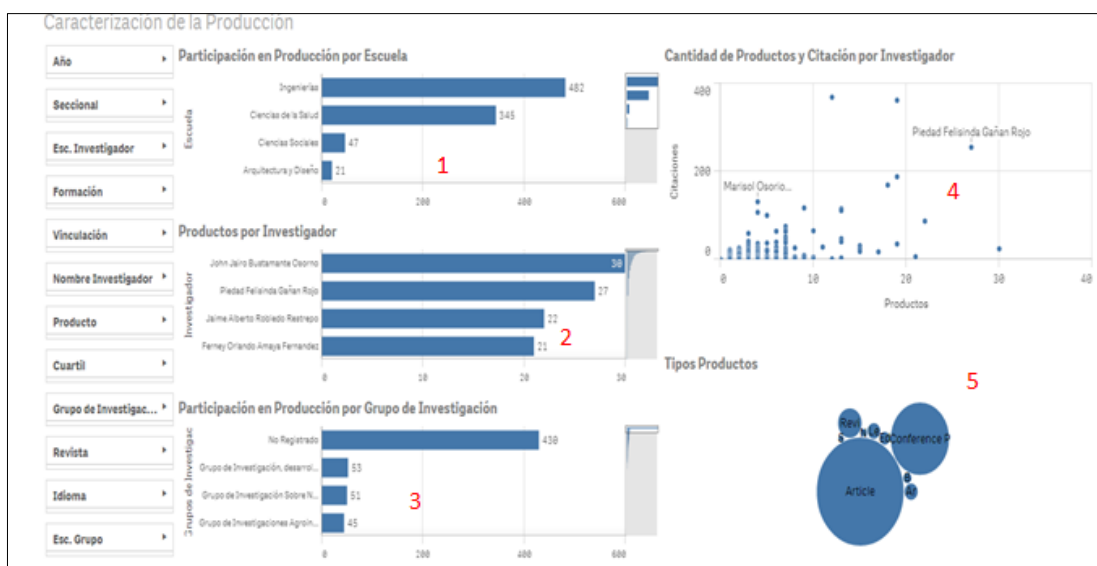
5.4.2 Etapa 2: análisis

La Figura 13 y

Figura 14 muestran la caracterización de la producción. Se pueden apreciar las relaciones que fueron consideradas más importantes en función de cada investigador de la UPB con su producción. Se asignaron números para explicar el tipo de gráfico y su significado.

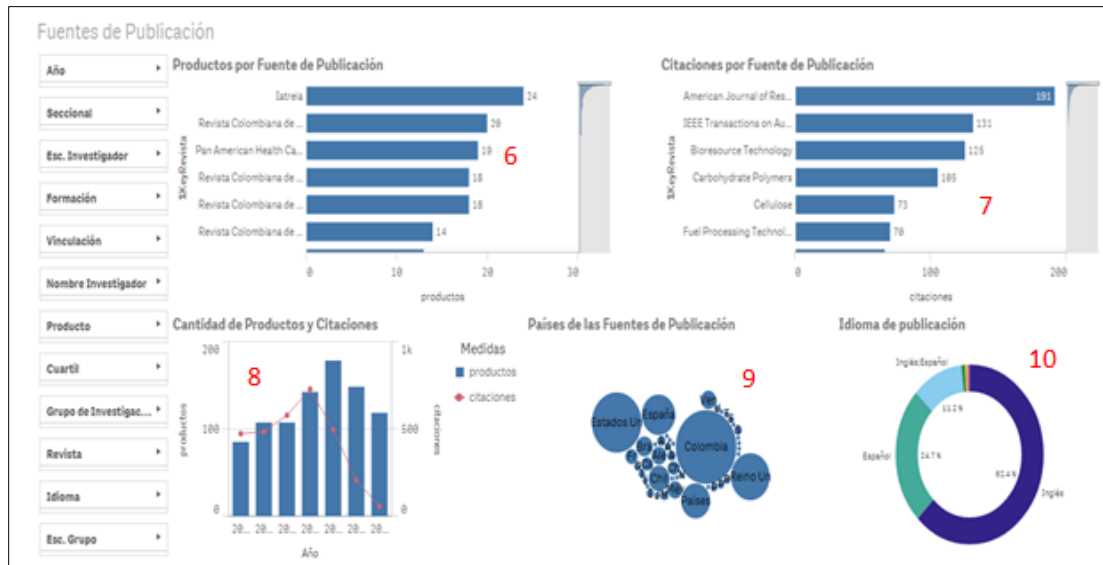
- 1 Cantidad de productos que posee cada una de las escuelas de la Universidad.
- 2 Productos por investigador.
- 3 Productos por grupo de investigación.
- 4 Cantidad de productos y citación por investigador.
- 5 Tipos de productos.
- 6 Productos por fuente de publicación.
- 7 Citaciones por fuente de publicación.
- 8 Cantidad de productos y citaciones en el periodo de tiempo analizado.
- 9 Países de las fuentes de publicación.
- 10 Idioma de publicación.

Figura 13: Caracterización de la producción



Fuente: Elaborado por los autores a través de la herramienta Qlik Sense

Figura 14: Fuentes de publicación



Fuente: Elaborado por los autores a través de la herramienta Qlik Sense

5.4.3 Etapa 3: reporte

La Figura 15 y Figura 16, muestran la información más granular y permite a los usuarios obtener datos detallados.

La Figura 15 contiene el número de productos y citaciones por año, seccional, escuela, nivel de formación, nombre completo del investigador y producto.

Figura 15: Detalle producto

Detalle de Productos			
Año ▾ Seccional ▾ Escuela ▾ Investigador ▾ Producto ▾ Nivel de Formación ▾		Medidas	
		Productos	Citaciones
2009		85	468
2010		107	476
2011		107	572
2012		142	721
2013		178	490
2014		148	202
2015		118	53

Fuente: Elaborado por los autores a través de la herramienta Qlik Sense

La Figura 16 contiene la información de la cantidad de productos y citas de las fuentes de publicación y las categorías de las revistas.

Figura 16: Categorías de las fuentes de publicación

Categorías de las Fuentes de Publicación

			Medidas	
Fuente	Categoría	TítuloProducto	produc tos	citacio nes
Actas Dermo-Sifiliograficas			1	0
Actas Espanolas de Psiquiatria			4	20
Adsorption		Chemical Engineering (miscellaneous)	2	26
		Chemistry (miscellaneous)	2	26
		Surfaces and Interfaces	2	26
Advanced Materials Research			2	0
Advances in Applied Mathematics and Mechanics			1	4
Advances in Intelligent Systems and Computing			1	0
Aging Research - Methodological Issues			1	0
Agrociencia			2	0
Agronomy for Sustainable Development			1	8
AIChE Annual Meeting, Conference Proceedings			1	0
AIP Conference Proceedings			2	0
Air and Waste Management Association - International Conference on Thermal Treatment Technologies and Hazardous Waste Combustors 2013			1	0
Allergologia et Immunopathologia			1	0
Alzheimer's Research and Therapy			1	6
American Journal of Hospice and Palliative Medicine			1	1

Fuente: Elaborado por los autores a través de la herramienta Qlik Sense

Una vez presentado el entorno que contiene cada tablero (Tablero de control, Análisis y Reporte), se procederá a describir las gráficas e indicadores

relacionados con el desempeño de la producción científica de alto impacto de la Universidad Pontificia Bolivariana desde el 2009 a octubre de 2015.

El análisis se divide en dos etapas con base en indicadores bibliométricos (Bordons & Zulueta, 1999). La primera etapa se hace según la producción o indicadores cuantitativos de la actividad científica, y la segunda etapa se hace según indicadores de impacto basados en las citas y que caracterizan la importancia de dicha producción en función del reconocimiento otorgado por otros investigadores.

5.4.4 Análisis según la producción

Se presentan los análisis de las gráficas más representativas. Algunas de ellas permiten identificar elementos de forma directa, y otras requieren la aplicación de filtros para poder ser analizadas de forma más precisa y detallada.

En cuanto a la primera etapa del modelo, el *Dashboard*, la Figura 17 muestra los *KPI* en relación con los productos, las citaciones, la cantidad de investigadores y las fuentes de publicación. En el rango de años de análisis, fueron publicados 885 productos de alto impacto en 467 tipos de fuentes de publicación (revistas, *conference paper* y libros, entre otros) en los cuales participaron un total de 721 personas con afiliación UPB y dichos productos han recibido alrededor de 2.980 citaciones a mayo de 2016.

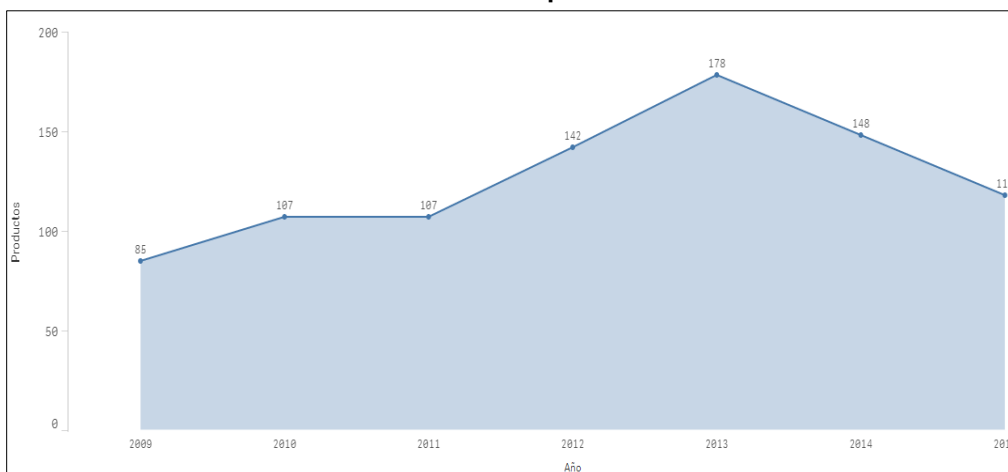
Figura 17: KPI

Productos	Citaciones Recibidas	Investigadores	Fuente de Publicación
885	2,98k	721	467

Elaborado por los autores a través de la herramienta Qlik Sense

En la Gráfica 1 se aprecia la evolución de los niveles de producción científica que ha tenido la UPB en los últimos 7 años. Se observa una disminución en los últimos 3 años 2013-2015.

Gráfica 1: Evolución de la producción científica



Fuente: Elaborado por los autores a través de la herramienta Qlik Sense

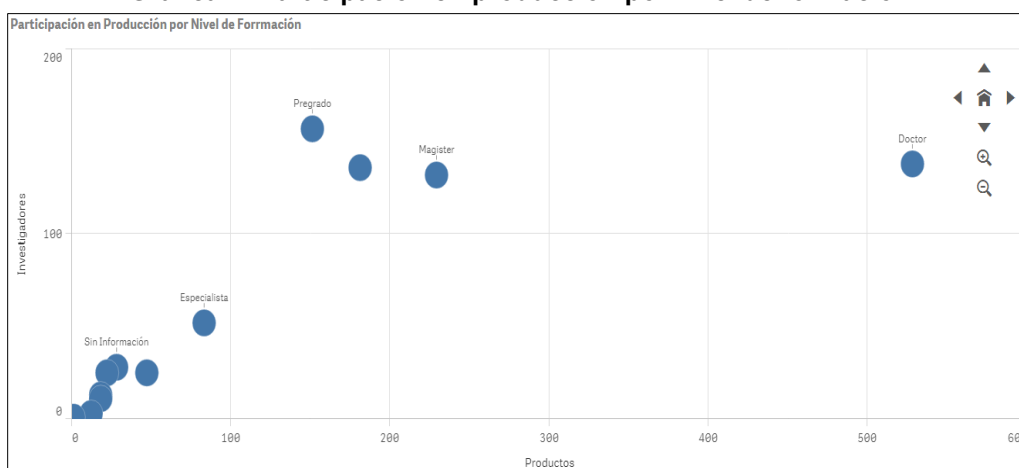
En general se observa que existe una tendencia de producción de aproximadamente 118 productos por año: En el 2013, la producción ascendió a 178 productos, este aumento, al parecer, se da por cuatro eventos que tuvieron lugar ese mismo año, dos de ellos en Medellín, uno en Bucaramanga y otro en Estados Unidos, y en los que hubo una nutrida participación de docentes de UPB con productos de tipo *conference paper*. En total luego de analizar los eventos, se encontró que 29 productos fueron publicados solo en esos 4 eventos. Lo anterior se debió a la cercanía geográfica existente entre la UPB y el lugar de los eventos, lo que facilitó la participación de los docentes.

La tendencia de producción en cada año desde el 2010, le ha permitido a la universidad figurar en la octava posición nacional del ranking *SCImago* (Scimago Lab, 2015), dónde uno de los requisitos es tener 100 artículos indexados en Scopus según el año del ranking. Esto ha permitido que la UPB se mantenga en altas posiciones en Colombia, y como primera institución privada en Antioquia. No obstante, es necesario que la Universidad revise cuales son las metas de producción anual y la calidad de esas publicaciones, para lo cual debe establecer estrategias que permitan incentivar la producción en especial las publicaciones en revistas de alto cuartil.

La Gráfica 2 muestra, los diferentes niveles de formación de los investigadores asociados a la UPB, la cantidad de investigadores pertenecientes a un nivel de formación específico y la cantidad de productos en los cuales han participado.

Para dar mayor claridad, es necesario leer cada eje de la gráfica. En el eje “Y”, figura la cantidad de investigadores y en el eje “X” se encuentra el número de productos. De esa forma, la gráfica indica que, los 138 investigadores registrados con formación doctoral para el año 2016 han participado en la elaboración de 528 productos. De igual manera, aplica para los otros niveles de formación. Se debe aclarar que en un mismo producto pueden participar múltiples investigadores con diferentes niveles de formación; doctores, magísteres, especialistas, estudiantes de pregrado o profesionales. Según lo anterior, se puede inferir que los investigadores que más participan en la producción de artículos son aquellos con nivel de formación doctoral, lo cual tiene sentido en el marco de generación de nuevo conocimiento que implica la formación doctoral.

Gráfica 2: Participación en producción por nivel de formación

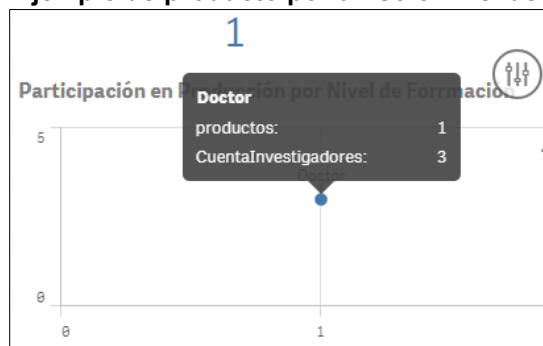


Fuente: Elaborado por los autores a través de la herramienta Qlik Sense

Para un mejor entendimiento de lo anterior, se hace un ejemplo con un solo producto, “*Bacterial cellulose films with controlled microstructure-mechanical property relationships*” tal como se aprecia en la Gráfica 3, se muestra la cantidad de investigadores con formación doctoral que participaron dentro del producto, en este caso el producto solo aplica para el nivel formación doctoral, es decir, en este único producto hubo participación de 3 investigadores con formación doctoral. En la Gráfica 4, por su parte, se aprecia cómo en un solo producto, “*Contingency analysis model of electrical power systems based on central angles from PMUs*”, hay participación de doctores y magísteres, en el ejemplo se aprecian cuatro doctores y, de forma paralela se señala la participación de un magíster. Es importante resaltar el fenómeno de colaboración entre los doctores, lo cual se

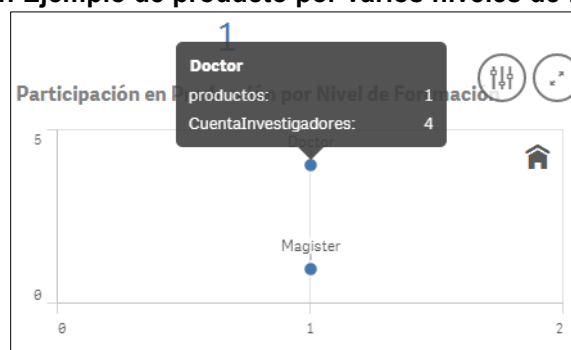
convierte en un posible factor de éxito en la publicación de artículos en revistas de alto impacto.

Gráfica 3: Ejemplo de producto por un solo nivel de formación



Fuente: Elaborado por los autores a través de la herramienta Qlik Sense

Gráfica 4: Ejemplo de producto por varios niveles de formación



Fuente: Elaborado por los autores a través de la herramienta Qlik Sense

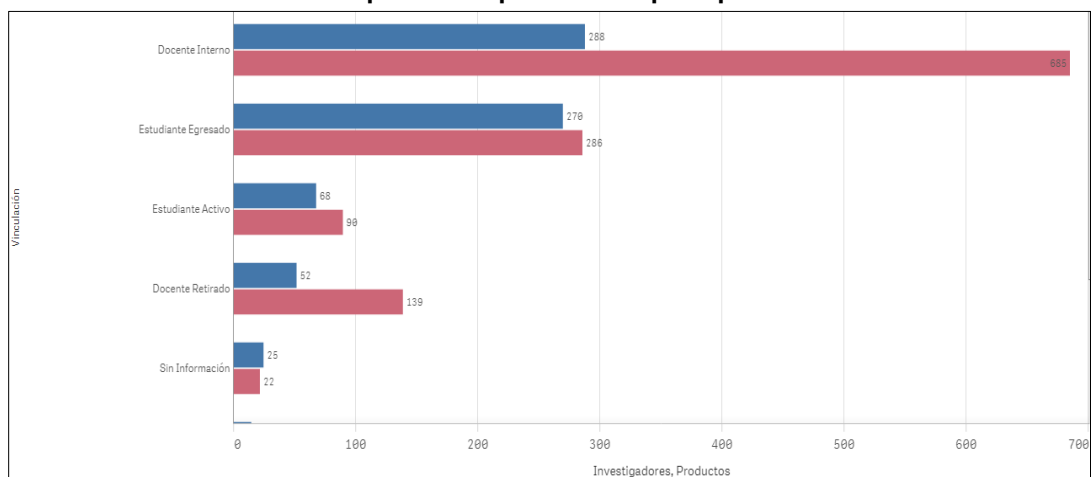
La Gráfica 5, se entiende como un complemento a la anterior, pero en este caso se está analizando el tipo de vinculación que tiene la persona con la UPB y la cantidad de investigadores.

La barra azul indica la cantidad de investigadores por tipo de vinculación y la roja la cantidad de productos por tipo de vinculación.

De forma general, de los 721 investigadores mostrados en el KPI de la Figura 17, 289 son docentes internos, 270 egresados, 68 estudiantes activos, 52 docentes retirados, 14 docentes externos, 2 internos administrativos, 1 docente ad honorem y 25 no tienen información. Para el caso particular de los doctores, aplicando un filtro, se encontró que de los 138 con ese nivel de formación, 108 son docentes internos, 21 docentes retirados, 6 estudiantes egresados, 2 interno administrativos y 1 docente externo. De igual forma se analizan los otros resultados.

De estos resultados se destaca la alta participación de los egresados, esto se da porque muchos de ellos participaron en la elaboración del producto mientras eran estudiantes, lo cual indica que la participación de autores con diferentes perfiles es una práctica exitosa para la publicación y demuestra que las diferentes escuelas están empleando estrategias que fomenta la formación investigativa y relevo generacional.

Gráfica 5: Participación en producción por tipo de vinculación

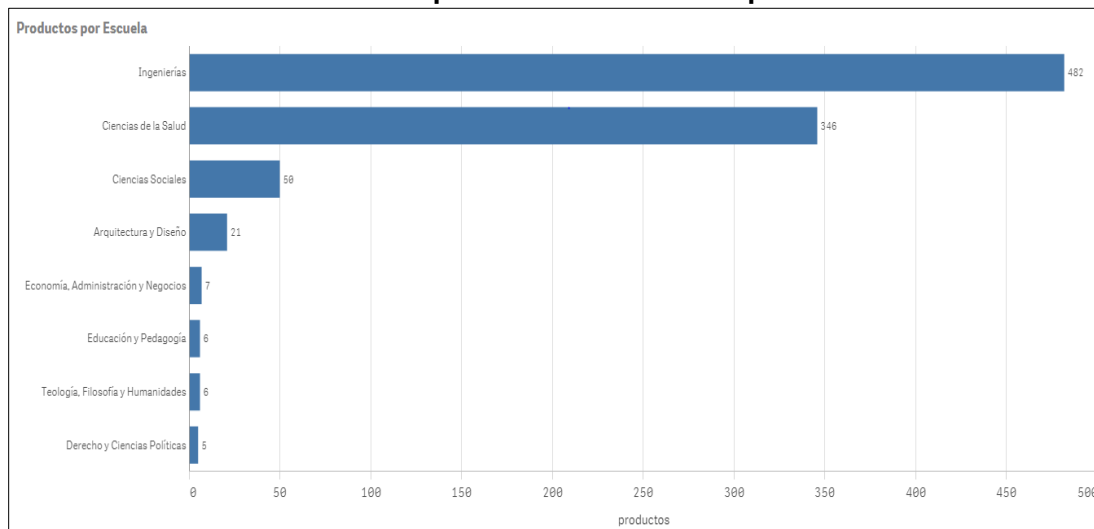


Fuente: Elaborado por los autores a través de la herramienta Qlik Sense

En resumen, la Gráfica 2 y Gráfica 5, muestran que el 40% de los participantes en productos corresponden a personas con vinculación de docentes internos, de los cuales 65% tienen nivel de formación de Doctorado y Maestría.

La Gráfica 6 muestra las veces que una escuela ha tenido participación en un producto. Se puede evidenciar que la mayor participación está concentrada en la Escuela de Ingenierías con 482 veces y en segundo lugar figura la Escuela de Ciencias de la Salud con 346 veces. Cabe aclarar que en un solo producto pueden participar investigadores de ambas escuelas y el producto cuenta para cada una de ellas. Además, es necesario precisar que un 3% de los productos que no se identificaron con vinculación de la UPB, estaban relacionados con temáticas del área de salud, lo que incide en el resultado final para esta escuela.

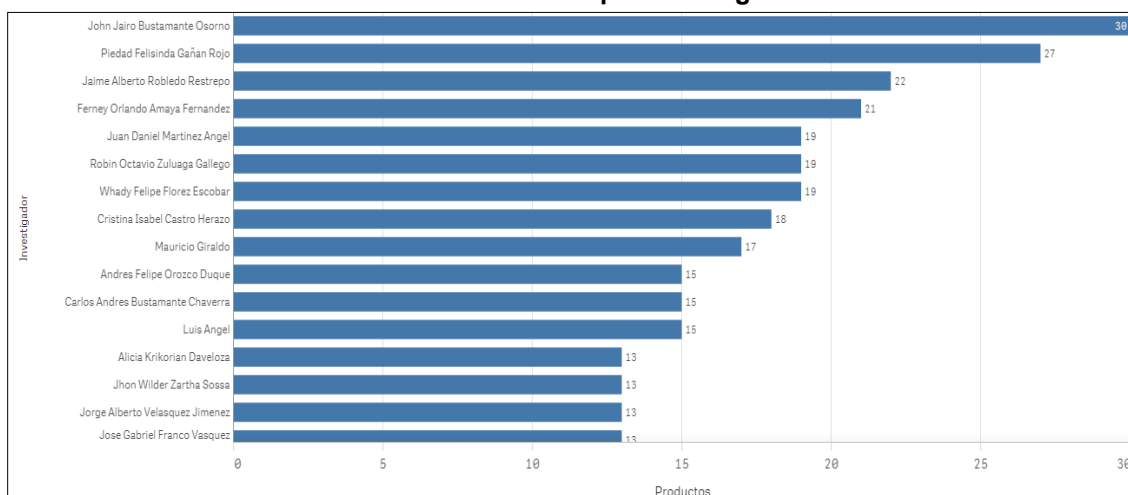
Gráfica 6: Participación en Producción por Escuela



Fuente: Elaborado por los autores a través de la herramienta Qlik Sense

Por su parte, la Gráfica 7 muestra la cantidad de productos realizados por los investigadores. Es de resaltar que los investigadores con mayor producción científica tienen una activa participación en proyectos de investigación y/o transferencia y dirección de tesis de doctorado, pero se hace necesario identificar las buenas prácticas de generación de conocimiento de estos investigadores para replicarlas al resto de la comunidad universitaria.

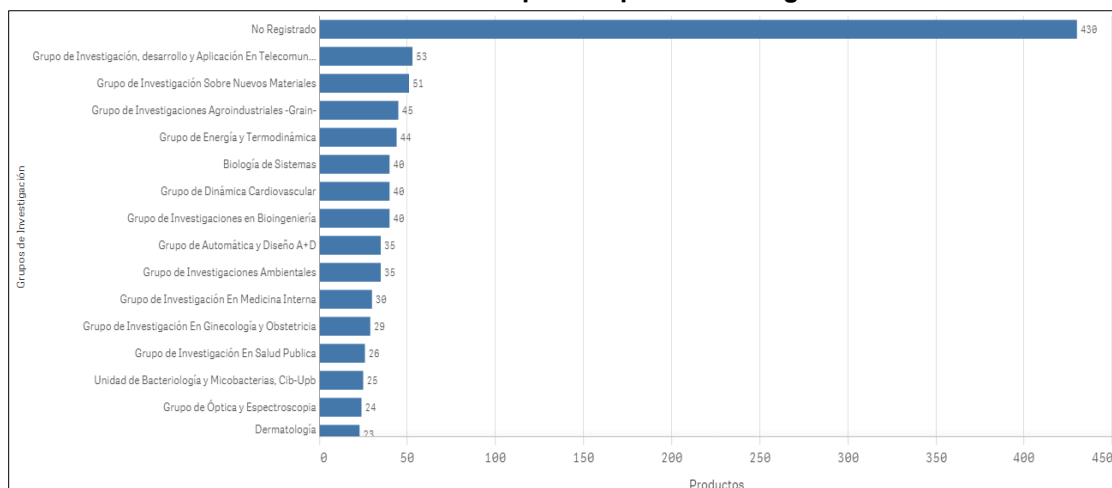
Gráfica 7: Productos por Investigador



Fuente: Elaborado por los autores a través de la herramienta Qlik Sense

En la Gráfica 8 se muestran los productos asociados a grupos de investigación de la Universidad. Se evidencia un alto número de productos sin grupo de investigación registrado, un total de 430. Esto se explica debido a que algunos de los autores participantes de dichos productos, no poseen vinculación alguna con un grupo de investigación específico, sin embargo, se debe aclarar que el mismo producto puede presentar contribución en otro grupo de investigación según las afiliaciones de los investigadores

Gráfica 8: Productos por Grupo de Investigación

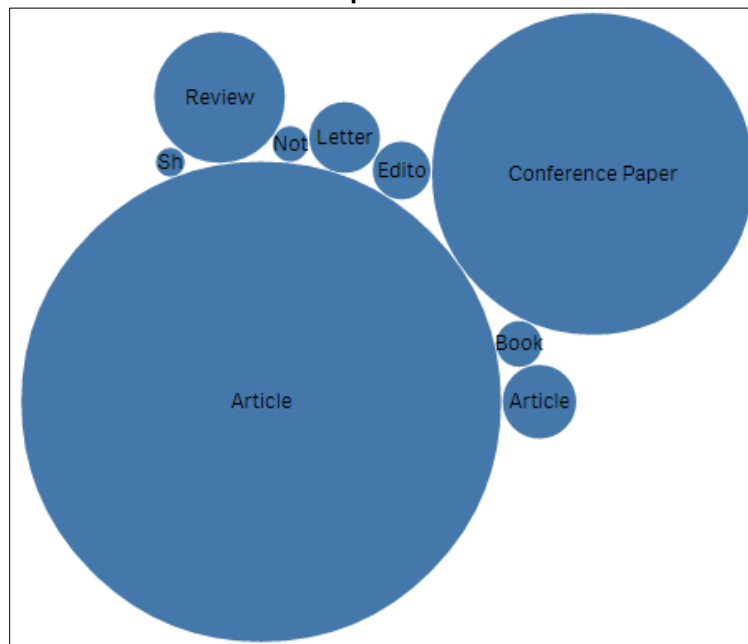


Fuente: Elaborado por los autores a través de la herramienta Qlik Sense

Por último, la

Gráfica 9 muestra los tipos de productos que se encuentran indexados en la base de datos *Scopus*. De los productos publicados por la institución en el rango de estudio, el 62.5% son artículos, el 28% *conference paper*, el 4,6% son *review*, el 1.5% *article in press*, el 1.4% *letter*, 0.9% *editorial*, 0,6% *book*, 0.3% *note* y 0.2% *short survey*.

Gráfica 9: Tipos de Productos



Fuente: Elaborado por los autores a través de la herramienta Qlik Sense

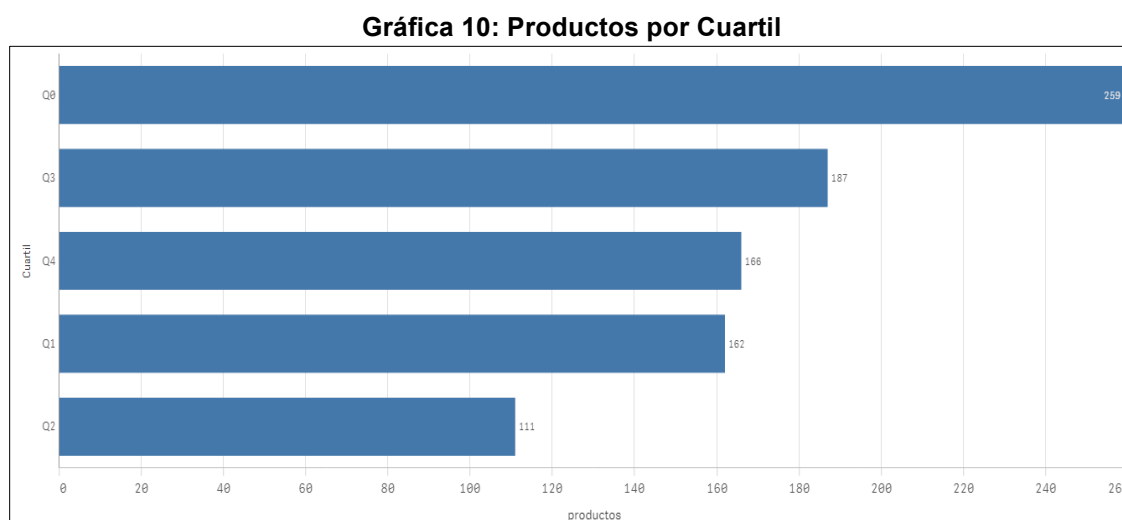
5.4.5 Análisis según el impacto científico

Para complementar los indicadores de producción que son netamente cuantitativos, se describen a continuación los indicadores cualitativos que valoran el impacto de las publicaciones. Según “La valoración del impacto de los trabajos a través de las citas que reciben no es una medida de tipo inmediato, sino que sólo puede aplicarse varios años después de la publicación de los documentos” (Bordons & Zulueta, 1999). Teniendo en cuenta lo anterior se procede a realizar un análisis de los indicadores cualitativos.

El primer indicador se refiere al impacto de la revista, el cual es medido a través de los cuartiles. El cuartil es una indicador o medida de posición de una revista en relación con todas las de su área. Si dividimos en 4 partes iguales un listado de revistas ordenadas de mayor a menor factor de impacto, cada una de estas partes será un cuartil. Las revistas con el factor de impacto más alto estarán en el primer

cuartil, los cuartiles medios serán el segundo y el tercero y el cuartil más bajo será el cuarto. En este sentido, en la Gráfica 10 se presenta el número de productos por cuartil, en el que se destaca que la mayor producción se concentra en fuentes sin cuartil (Q0). Este fenómeno se presenta porque una buena parte de la producción total son productos de tipo *conference paper*, que en su mayoría no tienen cuartil. Esta situación afecta las citaciones, pues al compararlo con el total de citaciones recibidas, se identifica que las publicaciones menos citadas son los *conference paper*.

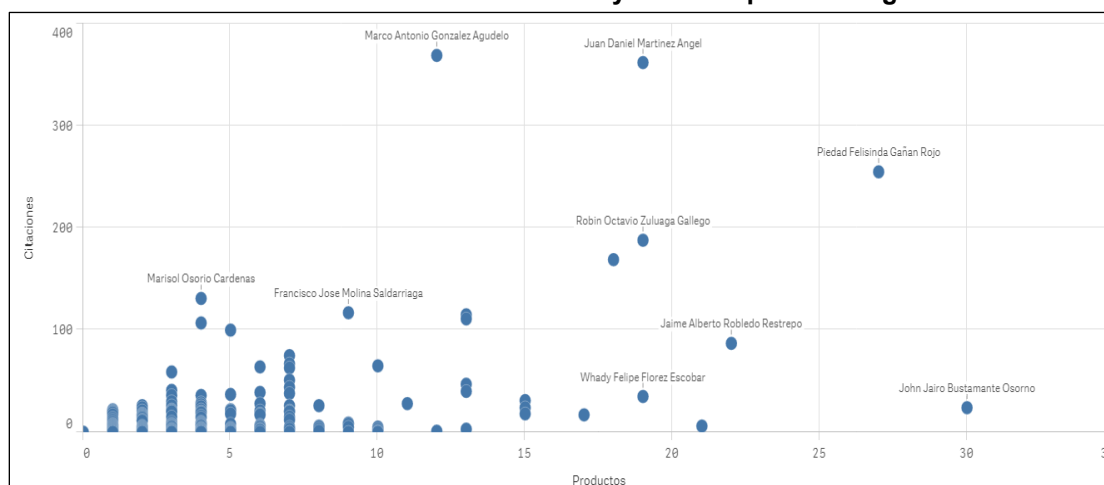
Se aclara que el cuartil de cada revista corresponde al asignado en 2014, y el más alto, dado que una revista puede estar clasificada en varias áreas y en cada una tener cuartiles diferentes. A su vez se evidencia que el 18% de los productos publicados se encuentran en el Cuartil 1, el 13% en Cuartil 2, el 29% no tienen cuartil, el 21% en Cuartil 3, el 19% en Cuartil 4. De acuerdo con este resultado, se puede concluir que el 31% de los artículos tienen gran probabilidad de ser citados (Cuartil 1 y 2).



Fuente: Elaborado por los autores a través de la herramienta Qlik Sense

La Gráfica 11 permite identificar la cantidad de productos que se han publicado por investigador y las citas que han recibido los autores. A su vez muestra que la relación cantidad-citación no es directa, es decir, a mayor cantidad de productos no hay mayor citación, pero esto es un caso muy particular, debido a que la citación se da con mayor frecuencia en artículos de alto impacto, es decir publicados en revistas de cuartil 1; evidenciándose con lo anterior que la mayoría de las publicaciones se dan en revistas de cuartiles 3 y 4, o en su defecto conferencias. Si a lo anterior se suma la tipología de los productos, se evidencia que los productos con mayor citación son los artículos en revistas.

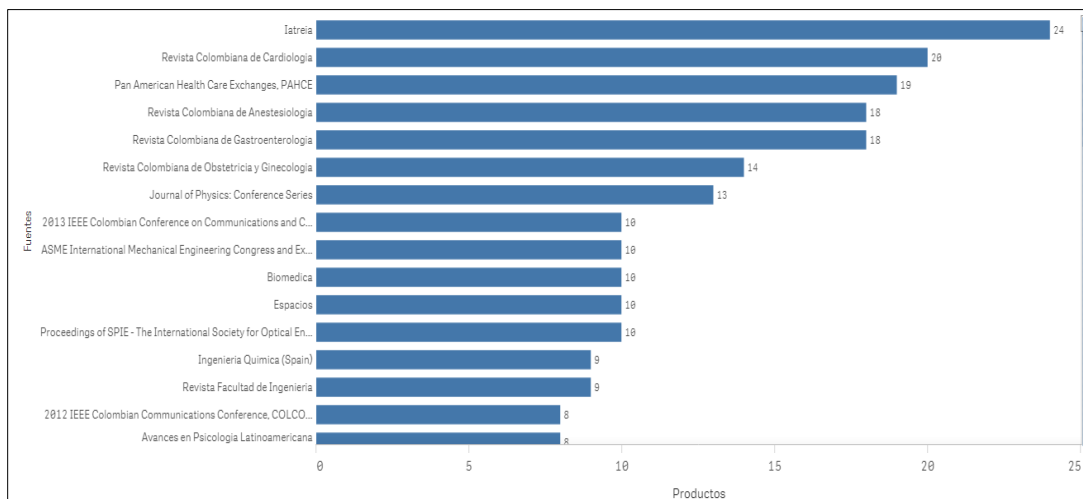
Gráfica 11: Cantidad de Productos y Citación por Investigador



Fuente: Elaborado por los autores a través de la herramienta Qlik Sense

La Gráfica 12 muestra las fuentes de publicación con mayor cantidad de productos. Un análisis detallado filtrando las primeras cinco fuentes, indica que donde más se han realizado publicaciones son de nacionalidad colombiana; las cuales se encuentran en cuartiles 3 y 4 (Q3 y Q4), o incluso sin cuartil. Lo anterior se refleja en el idioma de publicación de la fuente, en este caso, el 41.4% de los productos fueron publicados tanto en español como en inglés, el 39.4% fueron publicados solo en español y el 19,2% fueron publicados en inglés. Este asunto, también es un condicionante para la citación y por ende el impacto de las publicaciones.

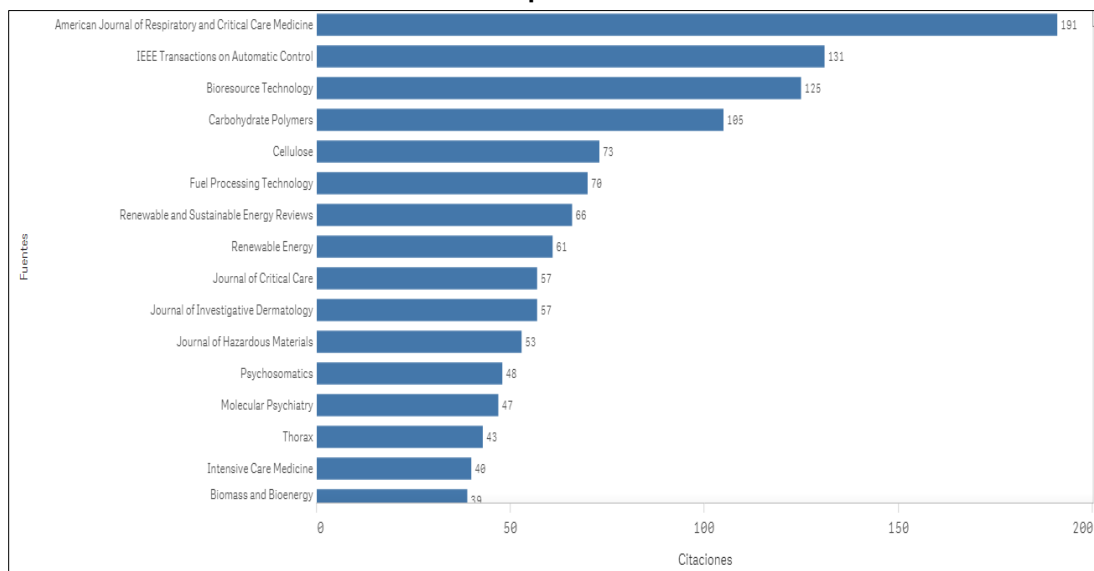
Gráfica 12: Productos por Fuentes de Publicación



Fuente: Elaborado por los autores a través de la herramienta Qlik Sense

En la Gráfica 13 se muestran las citaciones por fuentes de publicación. Un análisis de las primeras cinco fuentes con mayor citación, revela que son publicaciones de artículos en revistas de cuartil 1, donde el idioma de publicación es el inglés.

Gráfica 13: Citaciones por Fuente de Publicación



Fuente: Elaborado por los autores a través de la herramienta Qlik Sense

6 CONCLUSIONES

- De acuerdo con los análisis anteriores, la Universidad debe incentivar la publicación en altos cuartiles, y desestimular la publicación en *proceedings*, o plantear acciones en las que el investigador genere múltiples productos derivados de los *conference paper*. Lo anterior permite tener más citación y prestigio profesional.
- En los resultados se evidencia la participación de estudiantes en los diferentes tipos de productos. Por lo anterior, es necesario continuar estimulando la producción con los estudiantes de pregrado y posgrado para dejar capacidades instaladas y fortalecer la formación investigativa.
- Es pertinente que la institución aproveche las capacidades de los docentes con más producción para replicarlas en el resto de docentes. Permitiendo encontrar las mejores prácticas para publicar.
- Es fundamental que los autores identifiquen muy bien la revista en la cual están publicando los artículos y verifiquen que efectivamente los datos de su firma son correctos, para normalizar la información y facilitar la recuperación del producto por parte de la institución. A su vez para un adecuado análisis de la información, se requiere que a la hora de publicar se respete la afiliación de la persona con la institución a la que pertenece, ya que se encontró la participación de varios autores extranjeros con afiliación de la Universidad.
- Factores exógenos al procedimiento efectuado en este trabajo, como la no actualización de las bases de datos referencia por parte de las entidades

encargadas, ocasionan que algunos resultados no correspondan con la realidad del investigador, esto se vería reflejado en una aparente menor producción a la reportada. Se aclara que este documento no podrá manifestar tales actualizaciones si estas son hechas después de su entrega final. Sin embargo, el modelo podrá ser cargado con la nueva información si es requerida.

- Es relevante realizar una segunda fase, de un estudio en el cual se pueda incluir otro tipo de análisis que indague los tipos de colaboración y el estudio de las citaciones por autor que identifiquen el impacto del docente bolivariano en las comunidades científicas del mundo.
- La realización de este proyecto permitió evidenciar los beneficios que se pueden lograr para cualquier institución u organización a partir de la aplicación de técnicas y herramientas de inteligencia de negocios. Para el caso específico de este proyecto, se logró la consolidación y depuración de múltiples fuentes de información que podrán ser utilizadas para otros proyectos similares, o como avance para la implementación gradual de un almacén de datos institucional que pueda brindar una sola versión de la información y permita una visión de 360° sobre el negocio.
- La creación e implementación de este modelo de datos permitirá a la institución realizar análisis para el futuro desarrollo de estrategias de orientación de producción científica de los investigadores, encaminadas hacia el cumplimiento del alcance proyectado institucionalmente.
- A partir de la realización de este proyecto, se puede concluir que la calidad de los datos es un factor crucial para la implementación de cualquier solución de inteligencia de negocios, ya que puede tener un gran impacto negativo tanto en el tiempo de implementación del proyecto como en los resultados finales obtenidos a partir del modelo. Por lo anterior se propone a la institución implementar políticas institucionales de gobernabilidad de datos.

7 TRABAJOS FUTUROS

El principal trabajo futuro que se recomienda a la Universidad implementar, consiste en algo similar a lo logrado en este trabajo, pero con las siguientes características:

- Que se actualice en forma automática y permanente. Según se ha investigado, esto puede lograr ya que existe un servicio web disponible para extraer información de Scopus. a través del cual se podría extraer periódicamente los nuevos productos publicados y consultar en el modelo información totalmente actualizada.
- Que se extraiga información no únicamente de UPB sino de otras instituciones con las que nos interese compararnos, por ejemplo las universidades acreditadas.

Adicionalmente, de la realización de este proyecto se pueden derivar estudios sobre:

- Implementación de políticas de calidad de datos institucionales.
- Planteamiento de estrategias de aumento de la producción científica institucional.
- Caracterización de la producción científica por medio de tareas de minería de datos.

- Implementación de otras soluciones de inteligencia de negocios para beneficio institucional.

8 REFERENCIAS

- 1keydata. (Junio de 2015). Recuperado el Junio de 2015, de Bill Inmon vs. Ralph Kimball:
<https://www.1keydata.com/datawarehousing/inmon-kimball.html>
- Álvarez de Toledo, M. L. (2012). *Investigación, desarrollo y rankings científicos para América Latina. Universidades TOP*. Recuperado el 26 de Junio de 2015, de Infolatam:
<http://www.infolatam.com/2012/04/01/investigacion-desarrollo-y-rankings-cientificos-para-america-latina-universidades-top/>
- Arencibia, J. R., & de Moya Anegón, F. (Abril de 2008). La evaluación de la investigación científica: una aproximación teórica desde la cienciometría. *ACIMED*, 17(4).
- Bazan, D. (6 de Abril de 2013). *WordPress*. Recuperado el 5 de Noviembre de 2015, de Modelo Dimensional: <https://diegobazan7.wordpress.com/2013/04/06/que-es-un-modelo-dimensional/>
- Bermúdez García, J. (Agosto de 2014). Curso de Bodegas de Datos Universidad Pontificia Bolivariana.
- Bijsterbosch, M. (19 de Septiembre de 2015). *Collaborative organisation for ICT in Dutch higher education and research*. Recuperado el 3 de Julio de 2015, de Research information systems/CRIS: <https://www.surf.nl/en/themes/research/research-information/research-information-systems-cris/index.html>
- Bordons, M., & Zulueta, M. Á. (1999). Evaluación de la actividad científica a través de indicadores bibliométricos. *Revista Española de Cardiología*, 52(10), 790-800.
- Carpani, F. (2000). CMDM: Un Modelo Conceptual para la Especificación de Bases Multidimensionales. *CMDM: Un Modelo Conceptual para la Especificación de Bases Multidimensionales*. Uruguay.

- Centro de Investigación para el Desarrollo y la Innovación (CIDI) . (2014). *Informe Macroprocesos de Investigación*. Medellín.
- Comunidad de Madrid. (2002). *Madri+d*. (C. d. Madrid, Ed.) Recuperado el 30 de 11 de 2015, de https://www.madrimasd.org/informacionidi/biblioteca/publicacion/doc/Libro_12completo.pdf
- Cortés Rodas, F. (Julio de 2013). *Apuntes críticos al estudio del SCImago Research Group*. Recuperado el 4 de Julio de 2015, de <http://educacion.udea.edu.co/notieducacion/sites/default/files/Cr%C7%F0tica%20a%20informe-Scimago.pdf>
- Díaz, F. J., Osorio, M. A., Amadeo, A. P., & Romero, D. (2013). Aplicando estrategias y tecnologías de Inteligencia de Negocio en sistemas de gestión académica. *XV Workshop de Investigadores en Ciencias de la Computación*, (págs. 225-229).
- Escuela Colombiana de Carreras Industriales. (Junio de 2013). *Datamart para apoyo a las estrategias de permanencia estudiantil realizadas por el área de bienestar institucional*. Recuperado el 6 de Julio de 2015, de http://www.colombiaaprende.edu.co/html/micrositios/1752/articles-344993_recurso_3.pdf
- Eslava Schmalbach, J., Gaitán Duarte, H. G., & Escobar Córdoba , F. (2014). Producción científica de las facultades de medicina en Colombia, 1940-2014. *Revista Facultad de Medicina*, 62(3), 363-367.
- Fuentes Tapia, L., & Valdivia Pinto, R. (Diciembre de 2010). Incorporación de elementos de inteligencia de negocios en el proceso de admisión y matrícula de una universidad chilena. *Ingeniare revista chilena de ingeniería*, 18(3), 383-394.
- George, S. (Abril de 2012). *Techtarget*. Recuperado el 20 de Octubre de 2015, de Inmon vs. Kimball: Which approach is suitable for your data warehouse?: <http://www.computerweekly.com/tip/Inmon-vs-Kimball-Which-approach-is-suitable-for-your-data-warehouse>
- Gómez Vargas, M. (2014). *Formulario de Postulación de Buenas Prácticas*.
- How stuff works. (23 de Marzo de 2001). *HowStuffWorks*. Recuperado el 4 de Octubre de 2015, de What are relational databases?: <http://computer.howstuffworks.com/question599.htm>
- Jiménez, W. G. (2006). La formación investigativa y los procesos de investigación científico-tecnológica en la Universidad Católica de Colombia. *Studiositas*, 1(1), 36-43.
- Kimball, R., & Margy, R. (2010). *The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence*. Wiley Publishing.

- Kimball, R., & Ross, M. (2002). *The Data Warehouse Toolkit* (2ª edición ed.). (R. Elliott, Ed.) Estados Unidos de América: Wiley Computer Publishing.
- Kimball, R., Reeves, L., Ross, M., & Thornthwaite, W. (27 de 08 de 1998). *The Data Warehouse Lifecycle Toolkit*. John Wiley & Sons.
- López Martínez, R. E. (Octubre de 2011). Uso de herramientas de minería de datos y de inteligencia tecnológica para identificar patrones de publicación y patentamiento en entornos nacionales, así como sus concordancias. *XVI Congreso Internacional de Contaduría, Administración e Informática*.
- Martínez Cuesta, A., & Plasencia Salgueiro, A. (2010). *Valoración del Impacto de la Producción Científica del ICIMAF dentro del Programa Nacional de Ciencia Básicas*. Recuperado el 10 de Junio de 2015, de <http://www.dynamics.unam.edu/alci/files/docs/seminario2010/2010-P4-JORNADA%20CIENTIFICA%20JUVENIL2010.pdf>
- Microsoft. (2010). *Database basics*. Recuperado el 5 de Octubre de 2015, de <https://support.office.com/en-us/article/Database-basics-a849ac16-07c7-4a31-9948-3c8c94a7c204?>
- Narváez Triana, J. A., Monsalve Hernández, C. A., Bustamante Martínez, A., Galvis Lista, E. A., & Gómez Flórez, L. C. (Septiembre de 2013). Solucion de inteligencia de negocios para la gestión de recursos educativos y espacios físicos en la Universidad del Magdalena. *Avances Investigación en Ingeniería* , 10(1), 7-19.
- Ospina Rúa, D. N. (Febrero de 2009). Caracterización de la producción científica y visibilidad de los investigadores de la Universidad Nacional de Colombia Sede Medellín en la ISI Web of Science (1990-2007. Medellín.
- Quiroz Gil, N. I., & Valencia, C. A. (2012). Aplicación del proceso de KDD en el contexto de bibliominig: el caso Elogim. . *Revista Interamericana de Bibliotecología* , 35(1), 97-108 .
- Raiher, S. (2010). Estudio comparativo entre la producción científica latinoamericana y mundial en inequidades en salud durante los años 1999-2008.
- Ranjan, J. (2010). Business intelligence: concepts, components, techniques and benefits . *Journal of Theoretical and Applied Information Technology*, 9(1).
- Reyes Marroquín, M. R., & Rosales Tejada, P. A. (2007). Desarrollo de un datamart de información académica de estudiantes de la escuela de ciencias y sistemas de la facultad de ingeniería de la USAC . *Trabajo de graduación*.
- Rouse, M. (Septiembre de 2005). *Techtarget*. Recuperado el 26 de Diciembre de 2015, de multidimensional database (MDB): <http://searchoracle.techtarget.com/definition/multidimensional-database>

- Saavedra Fernández, Ó., Sotolongo Aguilar, G., & Guzmán Sánchez, M. V. (2002). Medición de la producción científica en América Latina y el Caribe en el campo agrícola y afines: un estudio bibliométrico. *Revista Española de Documentación Científica*, 25(2), 151-161.
- Sanz, R., & Miguel. (22 de Julio de 2010). Análisis y diseño de un data mart para el seguimiento académico de alumnos en un entorno universitario.
- Scimago Lab. (2015). *SIR Iber Colombia 2015*. Recuperado el Abril de 2016, de http://www.scimagoir.com/pdf/iber_new/SIR%20Iber%20COL%202015%20HE.pdf
- Sharma, N., Perniu, L., Chong, R. F., Iyer, A., Nandan, C., Mitea, A.-C., y otros. (2010). *Database Fundamentals* (1ª edición ed.). Markham, Canadá: IBM.
- Sinergia e Inteligencia de Negocio. (2010). *Sinnexus*. Recuperado el 20 de Octubre de 2015, de Datawarehouse: http://www.sinnexus.com/business_intelligence/datawarehouse.aspx
- Sinergia e Inteligencia de Negocio. (2010). *Sinnexus*. Recuperado el 28 de Diciembre de 2015, de Bases de datos OLTP y OLAP OLTP - On-Line Transactional Processing: http://www.sinnexus.com/business_intelligence/olap_vs_oltp.aspx
- Universidad Pontificia Bolivariana. (2013). PLAN DE DESARROLLO UNIVERSITARIO 2014 – 2016. Medellín: Universidad Pontificia Bolivariana.
- Universidad Pontificia Bolivariana. (2014). *PLAN DE DESARROLLO UNIVERSITARIO 2014 – 2016*. Recuperado el Abril de 2016, de http://www.upb.edu.co/pls/portal/docs/PAGE/GPV2_UPB_MEDELLIN/PGV2_M065_PLAN EACION/PGV2_M065010_PLANDLLO/PGV2_M065030010010_PLANDLLO/PDI%202014%202016%20AGOSTO%2019%20DE%202014%20VERSI%D3N%20CENTRAL.PDF
- Universidad Pontificia Bolivariana. (2015). Modelo Pedagógico Integrado. Medellín, Antioquia, Colombia: Universidad Pontificia Bolivariana.
- Weisensee, D., Matthews, E., & McInnis, A. (2005). Implementing Data Warehousing and Business Intelligence at McMaster University Using the SAS® Intelligence Value Chain. *SUGI 30*. San Francisco.

ANEXO 1
ENCUESTAS REALIZADAS

Encuesta realizada a Marcela Restrepo Rodas

1. ¿Cuáles son sus indicadores claves para la producción científica?
- Tipo de PU
 - Campos de PU
 - Colaboraciones
 - Al encadenar de la PU (índice h)
 - Redes? Tipo de colaboración (con quien nos asociamos)
 - Nivel de desempeño x oportunidades (artículos/documentos) por tipo de
 - PU resultado procesos de I+D+i
2. ¿Qué reportes genera a partir de la información de la base de datos scopus?
- ¿Qué datos son los más importantes del reporte de la producción científica de alto impacto indexada en scopus?
- PU x investigadores UPIB
 - PU ciudad
3. ¿Para qué utiliza esta información? ¿Cuál son los usos?
- Mejoramiento para analizar, evaluar, mejorar, orientar o reorientar resultados de la actividad de I+D+i en la Unidad
4. ¿Hay cuellos de botella para obtener la información?
- No se cuenta con los reportes detallados para
 - Para normalización de la información
5. ¿Cuánta información histórica es requerida?
- Manual
6. ¿Qué oportunidades de mejora existen en lo referente al análisis de información para toma de decisiones?
- Concentración de las capacidades institucionales
 - Por investigador, por estructura de alto, general
 - Tipos de relacionamiento de la U con los generadores de alto nivel mundial x
7. ¿Cuáles son los aspectos que encuentran más difíciles en el proceso actual y que considera que puede ser cambiada para mejor?
- * Información dispersa, Manual, y

Xana Hilda Riquelme
FIRMA

Encuesta realizada a Maricela Gómez Vargas

Natalia Henao Saldarriaga

De: Maricela Gomez Vargas
Enviado el: miércoles, 21 de octubre de 2015 8:02 a. m.
Para: Natalia Henao Saldarriaga
Asunto: RE: Preguntas

Hola

Las respuestas son:

1. ¿Cuáles son sus indicadores claves para la producción científica?
No. de documentos x año
Fuentes de publicación
Autores
Idioma de publicación
Colaboración países (afiliación autores)
Colaboración Instituciones (afiliación de los autores)
Tipo de documento publicado
Área de conocimiento (tomar del área en que está clasificada en Scimago)
Impacto de las revistas (cuartil)
No. De citas obtenidas
Documentos no citados (los que tienen 0 citas)
Liderazgo (cuantos firman como autor de correspondencia)
quienes citan a nuestros autores- como se generan comunidades científicas (a partir de las referencias bibliográficas) descargar
2. ¿Qué reportes genera a partir de la información de la base de datos scopus? ¿Qué datos son los más importantes del reporte de la producción científica de alto impacto indexada en scopus?

producción por área

Colaboración internacional o nacional, internacional y nacional, sin colaboración (solo UPB) Países e instituciones

Colaboración por áreas temáticas

Eficiencia de producción (artículos por año en relación a labor no instruccional-proyectos- labor instruccional)

Porcentajes de publicaciones por cuartiles

Docentes más productivos por cuartiles

No. de citas por áreas (cual área recibe más cita) cual autor es más citado)

Autores y áreas con mayor liderazgo

Grupos de investigación con más producción

Grupos de investigación con producción de liderazgo

Grupos de investigación con producción de calidad (Q1)

Coautoría (cuantos trabajan con otros investigadores Local-nacional internacional)

3. ¿Para qué utiliza esta información? ¿Cuál son los usos?

Toma de decisiones, generación de estrategias para aumentar la producción, simular los resultados en los rankings internacionales, entregar información a procesos de autoevaluación institucional, indicadores institucionales.

4. ¿Hay cuellos de botella para obtener la información?

Para obtener la información no, es más para manipularla, pues se requiere un trabajo muy manual.

5. ¿Cuánta información histórica es requerida?

Toda la que haya en la base de datos. (desde 1960)

6. ¿Qué oportunidades de mejora existen en lo referente al análisis de información para toma de decisiones?

Tener la información al día, que se actualice automáticamente, generar reportes en tiempo real

7. ¿Cuáles son los aspectos que encuentran más difíciles en el proceso actual y que considera que puede ser cambiada para mejor?

La manipulación de los datos no debería ser tan manual. Sería bueno crear los reportes y generarlos automáticamente cada vez que se requieran.

Encuesta realizada a Zulma Vanessa Rueda Vallejo

De: Zulma Vanessa Rueda Vallejo
Enviado el: viernes, 26 de febrero de 2016 1:37 p. m.
Para: Natalia Henao Saldarriaga <natalia.henao@upb.edu.co>
Asunto: Encuesta

1. ¿Cuáles son sus indicadores claves para la producción científica?
 - a. Artículo en revistas indexadas
 - b. Presentación en eventos académicos nacionales o internacionales
 - c. Diferentes formas de eventos o publicaciones para apropiación social del conocimiento (se hace muy poco para que realmente se haga apropiación social, no solo para mostrar un papel o un certificado, sino algo real).
 - d. Número de "Policy Brief" o resúmenes con los principales resultados del proyecto con el fin de que los tomadores de decisiones puedan decidir con base en los hallazgos de las investigaciones, para los proyectos cuyas temáticas apliquen. Este punto es algo que prácticamente no se trabaja casi en Colombia pero que creo esencial para que las investigaciones trasciendan y tengan un real impacto en la sociedad.
2. ¿Qué reportes genera a partir de la información de la base de datos scopus? Realmente no me gusta mucho la base de datos de Scopus pues se demora varios meses en tener actualizadas las publicaciones de los investigadores. Pero las veces que la he utilizado, la uso para saber el número de docentes que tienen publicaciones indexadas en Scopus y evaluar la producción por ECS.
3. ¿Qué datos son los más importantes del reporte de la producción científica de alto impacto indexada en scopus?
Yo uso número de publicaciones de los autores, número de citaciones en las revistas de Scopus, y el gráfico del índice H de los autores.
4. ¿Para qué utiliza esta información? ¿Cuál son los usos?
Básicamente la utilizo para hacer seguimiento a la producción de los profesores de la ECS. Esto sirve para los informes de investigación del área, para la renovación de los registros calificados y acreditación de los programas de pre y posgrado.
5. ¿Hay cuellos de botella para obtener la información?
Sí porque no se tiene acceso por fuera de la universidad. Y como mencioné antes, se demora hasta 6 meses en actualizar la información de publicaciones de los profesores.
6. ¿Cuánta información histórica es requerida?
No entiendo esta pregunta. Si es que entro a buscar por un límite de tiempo, sí. Si es para informe de investigación de la ECS lo hago anualmente, si es para la renovación de los registros calificados lo hago para 3 o 5 años dependiendo del programa.
7. ¿Qué oportunidades de mejora existen en lo referente al análisis de información para toma de decisiones?
Básicamente la utilizo para estimular en los investigadores la publicación en revistas indexadas en ISI o Scopus.
8. ¿Cuáles son los aspectos que encuentran más difíciles en el proceso actual y que considera que puede ser cambiada para mejor?
No sé si esta pregunta se refiere a Scopus, pero con relación a Scopus es que se actualice de forma más rápida, casi que en tiempo real.
Con relación a los procesos de publicación en revistas internacionales, veo tres cuellos de botella muy grandes: 1) el temor de los investigadores a publicar en revistas indexadas en ISI o Scopus, 2) la universidad debe invertir un fondo grande para traducción de artículos en revistas internacionales pues los investigadores escriben muy poco en otro idioma diferente al español, y no pagan las traducciones, 3) un fondo grande para pago en revistas Open Access.

ANEXO 2

CÓDIGO DE EXTRAXIÓN TRANSFORMACIÓN Y CARGA

CÓDIGO CAPA DE EXTRACCIÓN

PaísesRevistas

```
PaísesRevista:
LOAD
    Revista,
    País
FROM [lib://Fuentes de datos/PaisesRevistas.xlsx]
(ooxml, embedded labels, table is Hoja3);

store PaísesRevista into $(Extraccion)PaísesRevista.qvd(qvd);
drop table PaísesRevista;
```

Idioma

```
Idioma:
LOAD
    IdCategoría,
    Descripción
FROM [lib://Fuentes de datos/Idioma.xlsx]
(ooxml, embedded labels, table is Hoja1);

store Idioma into $(Extraccion)Idioma.qvd(qvd);
drop table Idioma;
```

Cuartiles

```
Cuartiles:
CrossTable (Año, Cuartil)
LOAD
    Revista,
    "1999",
    "2000",
    "2001",
    "2002",
    "2003",
    "2004",
    "2005",
    "2006",
    "2007",
    "2008",
    "2009",
    "2010",
    "2011",
    "2012",
    "2013",
    "2014",
    "2015"
FROM [lib://Fuentes de datos/Cuartiles.xlsx]
(ooxml, embedded labels, table is Cuartiles);

store Cuartiles into $(Extraccion)Cuartiles.qvd(qvd);
drop table Cuartiles;
```

Personas

```
Personas:
LOAD
    ID,
    NIT,
    PRIMER_APELLIDO,
    NOMBRE,
    TIPO_EMPLEADO,
    "Nivel Formacion" as NivelFormacion,
    ESTUDIOS
FROM [lib://Fuentes de datos/Personas.xlsx]
(ooxml, embedded labels, table is Personas);

store Personas into $(Extraccion)Personas.qvd(qvd);
drop table Personas;
```

Grupos

```
Grupos:
LOAD
    Escuela,
    "Grupo de Investigacion",
    Cédula
FROM [lib://Fuentes de datos/Int Grupos.xlsx]
(ooxml, embedded labels, table is Integrantes);

store Grupos into $(Extraccion)Grupos.qvd(qvd);
drop table Grupos;
```

NivelesFormación

```
NivelFormacion:
LOAD
   Codigo,
    Descripcion
FROM [lib://Fuentes de datos/Nivel Formacion.xlsx]
(ooxml, embedded labels, table is Hoja1);

store NivelFormacion into $(Extraccion)NivelFormacion.qvd(qvd);
drop table NivelFormacion;
```

Categorías

```
Categoria:
LOAD
    Revista,
    Categoria
FROM [lib://Fuentes de datos/Categorías.xlsx]
(ooxml, embedded labels, table is Categorías);
store Categoria into $(Extraccion)Categoria.qvd(qvd);
drop table Categoria;
```

Articulos

```
Articulos:
LOAD
    Seccional,
    Escuela,
    Titulo,
    "Autor UPB" as FirmaAutor,
    ID,
    Cedula,
    Apellidos,
    Nombre,
    F9,
    "Autores no",
    Authors,
    F12,
    F13,
    "Year",
    "Source title",
    "Cited by",
    Affiliations,
    "Authors with affiliations",
    ISSN,
    ISBN,
    CODEN,
    DOI,
    "Language of Original Document",
    IdCategoria,
    "Document Type",
    EID
FROM [lib://Fuentes de datos/Articulos.xlsx]
(ooxml, embedded labels, table is Hoja1);
store Articulos into $(Extraccion)Articulos.qvd(qvd);
drop table Articulos;
```

CÓDIGO CAPA DE TRANSFORMACIÓN

DimCategoria

```
DimCategoriaFuente:
LOAD
    Revista as $keyCategoria,
    Categoria

FROM [lib://Extracción/Categoria.qvd]
(qvd);

store DimCategoriaFuente into $(Transformacion)DimCategoriaFuente.qvd(qvd);
drop table DimCategoriaFuente;
```

DimCuartiles

```
DimCuartiles:
LOAD
    Año & ' - ' & Revista as $KeyCuartil,
    //     Revista,
    //     Año,
    'Q' & Cuartil as Cuartil
FROM [lib://Extracción/Cuartiles.qvd]
(qvd);
store DimCuartiles into $(Transformacion)DimCuartiles.qvd(qvd);
drop table DimCuartiles;
```

DimFuentes

```
DimFuentes:
LOAD
    Revista as $KeyRevista,
    País as Pais
FROM [lib://Extracción/PaisesRevista.qvd]
(qvd);

store DimFuentes into $(Transformacion)DimFuentes.qvd(qvd);
drop table DimFuentes;
```

DimIdioma

```
DimIdioma:
LOAD
    IdCategoria as $KeyIdioma,
    Descripción as Lenguaje
FROM [lib://Extracción/Idioma.qvd]
(qvd);

store DimIdioma into $(Transformacion)DimIdioma.qvd(qvd);
drop table DimIdioma;
```

DimTipoProducto

```
DimTipoProducto:
LOAD

    DISTINCT "Document Type" as %KeyDimTipoProducto
FROM [lib://Extracción/Articulos.qvd]
// group by %KeyDimTipoProducto
(qvd);

store DimTipoProducto into $(Transformacion)DimTipoProducto.qvd(qvd);
drop table DimTipoProducto;
```

Producto

```
Producto:
LOAD

    DISTINCT EID as EID,
    Titulo as TituloProducto,
    "Cited by" as Citacion

FROM [lib://Extracción/Articulos.qvd]
(qvd);

store Producto into $(Transformacion)Producto.qvd(qvd);
drop table Producto;
```

DimTiempo

```
tmpFecha:
LOAD
    max(Floor("Year")) as tmpMaxAño,
    min(Floor("Year")) as tmpMinAño
FROM [lib://Extracción/Articulos.qvd]
(qvd);

let vMaxFecha= Peek('tmpMaxAño',0,tmpFecha);
let vMinFecha= Peek('tmpMinAño',0,tmpFecha);

TempTiempo:
LOAD $(vMinFecha) + rowNo()-1 as tmpFecha
AutoGenerate ($(vMaxFecha)-$(vMinFecha)+1);

DimTiempo:
// el tmpFecha es el atributo de l tabla tempTiempo
Load tmpFecha as #KeyTiempo,
    tmpFecha as Año

Resident TempTiempo;
drop table TempTiempo;

store DimTiempo into $(Transformacion)DimTiempo.qvd(qvd);
drop table DimTiempo;
```

Investigadores Parte 1/2

```
TablaNiveles:
Mapping
LOAD
    Codigo,
    Descripcion
FROM [lib://Extracción/NivelFormacion.qvd]
(qvd);
|
Investigadores:
LOAD

    ID,
    NIT as Cedula,
    PRIMER_APELLIDO,
    NOMBRE,
    NOMBRE & ' ' & PRIMER_APELLIDO as NombreCompleto,
    TIPO_EMPLEADO as TipoPersona,
    // NivelFormacion,
    ApplyMap('TablaNiveles', NivelFormacion, null()) as NivelFormacion
FROM [lib://Extracción/Personas.qvd]
(qvd);

join(Investigadores)
LOAD
    Titulo as TituloProducto,
    Seccional,
    Escuela as EscuelaInvestigador,
    Cedula

FROM [lib://Extracción/Articulos.qvd]
(qvd);
```

Investigadores Parte 2/2

```
join(Investigadores)
LOAD
    Escuela as EscuelaGrupo,
    "Grupo de Investigacion",
    Cédula as Cedula
FROM [lib://Extracción/Grupos.qvd]
(qvd);

store Investigadores into $(Transformacion)Investigadores.qvd(qvd);
drop table Investigadores;
```

FactProducto

```
FactProducto:
LOAD
    Titulo as TituloProducto,
    "Year" as %KeyTiempo,
    "Source title" as %KeyRevista,|
    "Document Type" as %KeyDimTipoProducto,
    "Year" & ' - ' & "Source title" as %KeyCuartil,
    "Source title" as %keyCategoria,
    IdCategoria as %KeyIdioma

FROM [lib://Extracción/Articulos.qvd]
(qvd);
store FactProducto into $(Transformacion)FactProducto.qvd(qvd);
drop table FactProducto;
```

CÓDIGO CAPA DE CARGA O NEGOCIO

DimCategoria

```
DimCategoriaFuente:
LOAD
    %keyCategoria,
    Categoria
FROM [lib://Tranformacion/DimCategoriaFuente.qvd]
(qvd);
store DimCategoriaFuente into $(Negocio)DimCategoriaFuente.qvd(qvd);
```

DimCuartiles

```
DimCuartiles:
LOAD
    %KeyCuartil,
    Cuartil
FROM [lib://Tranformacion/DimCuartiles.qvd]
(qvd);
store DimCuartiles into $(Negocio)DimCuartiles.qvd(qvd);
```

DimFuentes

```
DimFuentes:
LOAD
    Pais,
    %KeyRevista
FROM [lib://Tranformacion/DimFuentes.qvd]
(qvd);
store DimFuentes into $(Negocio)DimFuentes.qvd(qvd);
```

DimIdioma

```
DimIdioma:
LOAD
    %KeyIdoma,
    Lenguaje
FROM [lib://Negocio/DimIdioma.qvd]
(qvd);

store DimIdioma into $(Negocio)DimIdioma.qvd(qvd);
```

DimTipoProducto

```
DimTipoProducto:
LOAD
    %KeyDimTipoProducto
FROM [lib://Tranformacion/DimTipoProducto.qvd]
(qvd);
store DimTipoProducto into $(Negocio)DimTipoProducto.qvd(qvd);
```

DimTiempo

```
DimTiempo:
LOAD
    %KeyTiempo,
    Año
FROM [lib://Tranformacion/DimTiempo.qvd]
(qvd);
store DimTiempo into $(Negocio)DimTiempo.qvd(qvd);
```

Producto

```
Producto:
LOAD
    TituloProducto,
    Citacion,
    EID
FROM [lib://Tranformacion/Producto.qvd]
(qvd);
store Producto into $(Negocio) Producto.qvd(qvd);
```

Investigadores

```
Investigadores:
LOAD
    Cedula,
    ID,
    PRIMER_APELLIDO,
    NOMBRE,
    NombreCompleto,
    TipoPersona,
    NivelFormacion,
    TituloProducto,
    Seccional,
    EscuelaInvestigador,
    EscuelaGrupo,
    "Grupo de Investigacion"
FROM [lib://Tranformacion/Investigadores.qvd]
(qvd);
store Investigadores into $(Negocio) Investigadores.qvd(qvd);
```

FactProducto

```
FactProducto:
LOAD
    TituloProducto,
    %KeyTiempo,
    %KeyRevista,
    %KeyDimTipoProducto,
    %KeyCuartil,
    %keyCategoria,
    %KeyIdioma
FROM [lib://Tranformacion/FactProducto.qvd]
(qvd);

store FactProducto into $(Negocio) FactProducto.qvd(qvd);
```