



**UNIVERSIDAD DE  
MANIZALES**

Facultad de Ciencias e Ingeniería

**PROCESAMIENTO DE DATOS ACADÉMICO-ESPACIALES APLICANDO  
GEOBI A LAS FUENTES DE REGISTRO ACADÉMICO EN LA UNIVERSIDAD  
DE MANIZALES**

**JULIÁN ALBERTO ACEVEDO NOREÑA**



**UNIVERSIDAD DE MANIZALES  
FACULTAD DE INGENIERÍA  
MAESTRÍA EN TECNOLOGÍAS DE LA INFORMACIÓN GEOGRÁFICA  
MANIZALES  
2017**

PROTOCOLO INFORME FINAL DE INVESTIGACIÓN

**PROCESAMIENTO DE DATOS ACADÉMICO-ESPACIALES APLICANDO  
GEOBI A LAS FUENTES DE REGISTRO ACADÉMICO EN LA UNIVERSIDAD  
DE MANIZALES**

**JULIÁN ALBERTO ACEVEDO NOREÑA**

Trabajo de Grado presentado como opción para optar  
al título de Magíster en Tecnologías de la Información Geográfica

Presidente  
**CARLOS BETANCOURT CORREA**

**UNIVERSIDAD DE MANIZALES  
FACULTAD DE INGENIERÍA  
MAESTRÍA EN TECNOLOGÍAS DE LA INFORMACIÓN GEOGRÁFICA  
MANIZALES  
2017**

### TABLA DE CONTROL DE CAMBIOS

Nombre del Documento	Ubicación Específica	Criterio	Observación del Jurado	Cambio Propuesto
Informe Final Investigación	Título	¿El título es claro y acorde con el documento?	Jurado 1: Debe ser mas corto y al punto "Construcción de una base de datos geoespacial de negocios para la información académica de la Universidad de Manizales" Jurado 2: El título genera mucha expectativa pero al ver el documento se presentan muchas incertidumbres	PROCESAMIENTO DE DATOS ACADÉMICO-ESPACIALES APLICANDO GEOBI A LAS FUENTES DE REGISTRO ACADÉMICO EN LA UNIVERSIDAD DE MANIZALES
Informe Final Investigación	Título	¿El título tiene una longitud máximo de 15 palabras?	Jurado 1: Muchas mas Jurado 2: El título con los conectores tiene 22 palabras	17 palabras incluyendo artículos y preposiciones
Informe Final Investigación	Informe	¿El informe cumple con lo estipulado en la normatividad vigente en cuanto a la forma -normas (contec-, APA o IEEE)?	Jurado 1: Jurado 2: El documento no se ciñe a una norma de referencias, en algunos está en mayúsculas, en otros en minúsculas	Se reviso a cabalidad el documento y se realizaron las correcciones pertinentes
Informe Final Investigación	Informe	¿La redacción es clara, concisa y utiliza correctamente la terminología científica y técnica?	Jurado 1: Palabras en ingles deben evitarse y usar equivalentes en español. Solo se deben usar cuando son acrónimos Jurado 2:	Se repaso el informe y se efectuaron los cambios donde se consideraban pertinentes, ya que las citas textuales no es posible modificarse
Informe Final Investigación	Resumen	¿Los resúmenes son coherentes con el documento y respetan las normas del	Jurado 1: Al resumen le sobra el primer párrafo, y le falta ser concreto en decir	Se rehizo por completo el resumen del proyecto siguiendo en lo posible las

		idioma?	que se hizo y qué importancia tiene Jurado 2: El resumen no es claro, gran parte de la sección se habla de otras cosas y no del trabajo a realizar	observaciones de los jurados
Informe Final Investigación	Introducción	¿La introducción es adecuada y suficiente?	Jurado 1: Todos los acrónimos deben ser definidos. ¿Qué es SIG, qué es BI, qué es ETL, DW, Datamart, UTN? Así mismo usar equivalente en español de palabras en inglés: ¿White paper? Hay palabras escritas en portugués Jurado 2: Es adecuada, aunque factible de ser mejorada enfocándose en orientar al lector sobre el trabajo que se va a desarrollar	Se verifico el informe y se ejecutaron los cambios donde se consideraban pertinentes, ya que las citas textuales no es posible modificarse
Informe Final Investigación	Justificación	¿Se presenta claramente la novedad, interés y utilidad del proyecto?	Jurado 1: Jurado 2: No es claro la forma como se presenta los aportes novedosos del trabajo	Se reviso la sección y se ejecutaron los cambios donde se consideraban pertinentes
Informe Final Investigación	Procedimiento	¿Los procedimientos utilizados están debidamente descritos y son adecuados con los resultados obtenidos?	Jurado 1: Los procedimientos deben ser explicados de una forma tal que cualquiera pueda replicar el trabajo. No es el caso aquí pues solo se describen en forma teórica. Se deben incluir muestras del	Se incluyeron los pantallazos pertinentes

			código usado y pantallazos de los procesos ejecutados Jurado 2:	
Informe Final Investigación	Resultados	¿Los resultados obtenidos son plenamente coherentes con los objetivos planteados en la propuesta?	Jurado 1: Al parecer lo son, sin embargo solamente evaluando el desarrollo se podría afirmar esto. Así mismo el componente SIG de este trabajo no es muy claro Jurado 2: Es confuso al tratar de identificar los resultados del sistema y los resultados del caso de estudio	Se reviso la sección y se ejecutaron los cambios donde se consideraban pertinentes
Informe Final Investigación	Resultados	¿Los resultados están clara y ampliamente descritos y discutidos?	Jurado 1: Jurado 2: se podría hacer mayor énfasis en indicar la construcción de la propuesta y la solución; posteriormente indicar cómo se desarrolló el estudio de caso	Se reviso la sección y se ejecutaron los cambios donde se consideraban pertinentes
Informe Final Investigación	Resultados	¿Los resultados obtenidos dan respuesta pertinente y adecuada a la problemática indicada?	Jurado 1: Solo en forma teórica Jurado 2: Se plantearon seis objetivos, sin embargo en el documento se pudo constatar el cumplimiento de cuatro de ellos	Los resultados son consecuencia directa de los objetivos, algunos de ellos no están claros ya que se pueden evidenciar más detalladamente en el producto desarrollado, el cual será presentado en la sustentación
Informe Final Investigación	Conclusiones	¿Las conclusiones se derivan de la discusión de resultados y están claramente	Jurado 1: Debe haber una conclusión por cada objetivo y deben ser listadas en el mismo orden	Se reviso la sección y se ejecutaron los cambios donde se consideraban pertinentes

		presentadas?	de ejecución Jurado 2: Alinear los resultados a los objetivos y sobre ello se podría lograr una mejor discusión	
Informe Final Investigación	Recomendaciones	¿Las recomendaciones se derivan de la discusión de resultados y están claramente presentadas?	Jurado 1: No esta claro si la metodología Kimball sirvió o no sirvió, las dos primeras recomendaciones deben ser mejor explicadas pues no es claro que problema quieren resolver con ellas Jurado 2:	Se reviso la sección y se ejecutaron los cambios donde se consideraban pertinentes
Informe Final Investigación	Bibliografía	¿Las fuentes bibliográficas son suficientes, confiables, pertinentes y actualizadas?	Jurado 1: Jurado 2: No están actualizadas y no están estandarizadas para su presentación	Se reviso la sección y se ejecutaron los cambios donde se consideraban pertinentes

## CRÉDITOS

Las personas que participaron en este proyecto fueron las siguientes:

<b>NOMBRE COMPLETO</b>	<b>FUNCIÓN EN EL PROYECTO</b>	<b>DIRECCIÓN DE CONTACTO</b>	<b>CORREO ELECTRÓNICO</b>
Julián Alberto Acevedo Noreña	Autor	Universidad de Manizales	jacevedo@umanizales.edu.co
Carlos Betancourt Correa	Presidente	Universidad de Manizales	cbc@umanizales.edu.co

## PÁGINA DE ACEPTACIÓN

---

---

---

---

---

---

---

---

**<NOMBRE COMPLETO>  
JURADO**

---

**<NOMBRE COMPLETO>  
JURADO**

---

**<NOMBRE COMPLETO>  
JURADO**

Manizales, <día> de <mes> de <año>

## CONTENIDO

INTRODUCCIÓN .....	15
1. ÁREA PROBLEMÁTICA .....	17
2. OBJETIVOS .....	19
2.1 OBJETIVO GENERAL .....	19
2.2 OBJETIVOS ESPECÍFICOS .....	19
3. JUSTIFICACIÓN .....	20
4. FUNDAMENTOS TEÓRICOS .....	23
4.1 INTELIGENCIA DE NEGOCIOS .....	23
4.2 SISTEMAS DE INFORMACIÓN GEOGRÁFICA .....	25
4.3. ANTECEDENTES .....	27
5. METODOLOGÍA .....	44
5.1 TIPO DE TRABAJO .....	44
5.2 PROCEDIMIENTO .....	44
6. RESULTADOS .....	57
6.1 DESCRIPCIÓN DE RESULTADOS .....	57
6.2 DISCUSIÓN DE RESULTADOS .....	71
7. CONCLUSIONES .....	73
8. RECOMENDACIONES .....	75
BIBLIOGRAFÍA .....	76
ANEXO A .....	80

## LISTA DE FIGURAS

Figura 1. Automobile dashboard analogy (Quinn, 2009).....	24
Figura 2. Diagrama del ciclo de vida de Kimball (Kimball, 1998) .....	45
Figura 3. Modelo dimensional.....	47
Figura 4. Modelo físico.....	49
Figura 5. ETL dim_estudiante.....	50
Figura 6. ETL dim_periodo .....	50
Figura 7. ETL dim_estrato .....	50
Figura 8. ETL dim_programa .....	50
Figura 9. ETL dim_campo_formacion.....	51
Figura 10. Job .....	51
Figura 11. Job proyeccion.....	51
Figura 12. ETL dim_egresado .....	52
Figura 13. ETL fact_egresado .....	52
Figura 14. Job egresado .....	53
Figura 15. ETL matriculas.....	53
Figura 16. ETL aspirante .....	53
Figura 17. Job aspirante .....	54
Figura 18. Job estudiante .....	54
Figura 19. Job periodo.....	54
Figura 20. PostgreSQL Database.....	55
Figura 21. Pentaho Data Integration (Spoon).....	56
Figura 22. Qlik Sense Desktop .....	56
Figura 23. Beyond Traditional GIS/Imagery and Mass Markets (Imaging Notes, 2008).....	58
Figura 24. Modelo dimensional del Indicador 1 .....	62
Figura 25. Modelo dimensional del Indicador 2 .....	63
Figura 26. Modelo dimensional del Indicador 3.....	64
Figura 27. Modelo dimensional del Indicador 4 .....	65
Figura 28. Modelo dimensional del Indicador 5 .....	65
Figura 29. Modelo físico del datawarehouse .....	66
Figura 30. Transformación para cargar la tabla de dimensión dim_período del Indicador 1 .....	67
Figura 31. Transformación para cargar la tabla de dimensión dim_estudiante del Indicador 1 .....	67
Figura 32. Transformación para cargar la tabla de hechos fact_indicador_1 del Indicador 1 .....	67
Figura 33. Trabajo para cargar las tablas de dimensiones del Indicador 1 .....	68
Figura 34. Trabajo para ejecutar todo el ETL del Indicador 1 .....	68
Figura 35. Cuadro de Mando Operativo del Indicador 1.....	69
Figura 36. Cuadro de Mando Operativo del Indicador 2.....	69
Figura 37. Cuadro de Mando Operativo del Indicador 3.....	70
Figura 38. Cuadro de Mando Operativo del Indicador 4.....	70

Figura 39. Cuadro de Mando Operativo del Indicador 5..... 71

## **LISTA DE TABLAS**

Tabla 1. Cronograma de Actividades .....	45
Tabla 2. Indicador 1 .....	59
Tabla 3. Indicador 2 .....	59
Tabla 4. Indicador 3 .....	60
Tabla 5. Indicador 4 .....	61
Tabla 6. Indicador 5 .....	61

## RESUMEN

Para nadie es un secreto que hoy en día, la información se ha convertido en el baluarte y apoyo más indispensable para una compañía que tenga como estrategia de negocio el mantenerse altamente competitiva; cualquiera sea el entorno en que se desenvuelva, es esencial conocer las necesidades y exigencias del medio, por esta razón los datos históricos, sumados a los que producen constantemente las múltiples operaciones de sus sistemas de información se convierten en una fuente inmensa de información que debe ser analizada y procesada.

En este sentido, y puesto que los cambios que se producen actualmente en las tecnologías y sistemas de información son demasiado rápidos, dentro del presente proyecto se planteó una metodología clara y organizada para lograr desarrollar un almacén de datos (data warehouse) con un elemento geográfico aplicado a un entorno académico universitario; inicialmente fundamentado en un principio teórico de conocimientos, y familiarización con las principales tecnologías a emplear, obteniendo con ello unas sólidas bases sobre el que construir el desarrollo que constituye el propósito del proyecto; seguido a ello se describen con detalle cada una de las fases de la metodología empleada para su construcción; y para finalizar los resultados muestran la viabilidad de implementar un Almacén de Datos Geográfico (DWG) para la recopilación y análisis de datos, generados por su propia data histórica, y producidos por los sistemas operacionales tradicionales internos y foráneos de interés para la organización, pretendiendo con todo ello contribuir a suplir y brindar apoyo oportuno a los reportes de indicadores académicos para toma de decisiones de Registro Académico en la Universidad de Manizales.

**PALABRAS CLAVES:** Data Warehouse Geográfico, GeoBI, Universidad de Manizales, SIGUM

## **ABSTRACT**

It is no secret that today, information has become the stronghold and most essential support for a company that has as a business strategy to stay highly competitive; whatever the environment in which it is unwrap, it is essential to know the needs and demands of the environment, and for this reason, the historical data, together with those that constantly produce the multiple operations of their information systems turns into an immense source of information that must be analyzed and processed.

In this regard, and since the changes currently taking place in the information technology and information systems are too fast, within this project a clear and organized methodology was raised to develop a data warehouse (data warehouse) with a geographical element applied to a university academic environment; initially based on a theoretical principle of knowledge, and familiarization with the main technologies to be used, thereby gaining a solid basis on which to build the development that is the purpose of the project; continued to it each of the phases are described in detail of the methodology used for its construction; and finally, the results show the feasibility of implementing a Geographic Data Warehouse (GDW) for data collection and analysis, generated by its own historical data, and produced by the traditional internal and foreign operational systems of interest to the organization, trying with all this to help to fulfill and offer opportune support to the reports of academic indicators for decision making of Registro Académico in the Universidad de Manizales.

**KEYWORDS:** Geographic Data Warehouse, GeoBI, Universidad de Manizales, SIGUM

## INTRODUCCIÓN

De acuerdo con la Harvard Business Review uno debe “Convertirse en un competidor analítico: Utilizar tecnología sofisticada de recopilación de datos y análisis para exprimir hasta la última gota de valor de todos sus procesos de negocio” (Harvard Business Review, 2011). En la actualidad las organizaciones necesitan datos precisos, actualizados y completos que las lleven a tomar decisiones clave del negocio. Sin embargo, los datos residen en diferentes bases de datos y en múltiples formatos. No existe una manera realmente fácil de crear una visión global de la organización integrada e inteligente. Lo que, es más, hay un lapso considerable entre necesitar la información y realmente conseguirla, lo que conlleva a las acciones retardadas y posiblemente pérdida de oportunidades de negocio.

La necesidad informacional de una empresa conlleva a generar gestiones innovadoras de inteligencia empresarial en ETL<sup>1</sup> y desarrollo de aplicaciones; proporcionando acceso a todos los datos de la organización, presentándolos a los facultados para tomar las decisiones de una manera organizada, inteligente que fomente medidas de acción rápidas y precisas.

Pero ¿Cuál es la base para haber aplicado SIG<sup>2</sup> y BI<sup>3</sup> en la Universidad de Manizales?, hoy por hoy casi todos los mercados dependen de la geografía para tener éxito; la ubicación de las tiendas, sucursales, clientes y más, todo debe ser analizado e interpretado para optimizar el rendimiento del negocio. Así pues, se deben complementar y ampliar las herramientas de BI existentes, de manera que proporcionen capacidades de análisis y cartografía espacial, limpieza e integración de datos, bases de datos demográficas y de mercado, y más.

En este contexto, se puede utilizar el proceso de análisis en dos enfoques bien definidos: descriptivo o predictivo. En el primer enfoque, se intenta caracterizar las situaciones de manera tal de poder entender cuáles son las variables principales que describen el comportamiento de una situación particular. El segundo enfoque implica la utilización del modelo para establecer de antemano una situación problemática. (La Red Martínez, Karanik, Giovannini, & Pinto, 2015)

Todo esto lleva a observar la carencia que existe de poseer una solución de procesamiento de información que aplique BI y que se encuentre construida con

---

<sup>1</sup> Extract, Transform and Load (Extracción, Transformación y Carga)

<sup>2</sup> Sistemas de Información Geográfica

<sup>3</sup> Business Intelligence (Inteligencia de Negocios)

capacidades espaciales (Geo-BI<sup>4</sup>), que permita utilizar la información geográfica para obtener ventajas sobre la competencia, tales como:

- ¿Por qué algunos lugares funcionan mejor que otros?
- Los patrones de comportamiento y de compra de sus clientes
- ¿Dónde se encuentra la mejor nueva oportunidad de mercado?
- Objetivos de fusiones y adquisiciones lucrativas
- El rendimiento de las estrategias de publicidad y marketing

En el Informe Blanco: GIS and Business Intelligence: The Geographic Advantage publicado por ESRI, empresa líder en el sector a nivel mundial y que en la actualidad desarrolla y distribuye software SIG, registra que:

Históricamente, la inteligencia empresarial y los sistemas de información geográfica han seguido su desarrollo e implementación por caminos separados. Solicitudes de clientes para una imagen más completa de funcionamiento y la capacidad de ser más proactivo han dado lugar a la combinación de estas dos tecnologías. Los requerimientos regulatorios también han aumentado la visibilidad de ambas tecnologías dentro de muchas organizaciones. En respuesta a los usuarios de BI y SIG, los proveedores líderes de BI han estado integrando las dos tecnologías y proporcionado soluciones innovadoras a un número cada vez mayor de usuarios finales. Los usuarios están respondiendo con nuevas aplicaciones que aprovechan la sinergia de las tecnologías combinadas. (GIS and Business Intelligence: The Geographic Advantage, 2006)

---

<sup>4</sup> Según la OGC (Open Geospatial Consortium), GeoBI es Inteligencia de Negocios que hace uso de información geoespacial

## 1. ÁREA PROBLEMÁTICA

De acuerdo con IDC (International Data Corporation), “De 2005 a 2020, el universo digital crecerá en un factor de 300, de 130 exabytes de 40.000 exabytes, o 40 billones de gigabytes (más de 5.200 gigabytes por cada hombre, mujer y niño en el año 2020). A partir de ahora hasta el año 2020, el universo digital será aproximadamente el doble cada dos años.” (Reinsel & Gantz, 2012)

Con el crecimiento continuo y constante de grandes volúmenes de datos dentro de las organizaciones se vuelve cada vez más compleja las tareas de almacenamiento, seguridad, administración y transformación de los mismos, escenario que dificulta aún más la limpieza, integración, análisis e interpretación de los datos generados por sistemas de información transaccionales y operacionales de las compañías, con la única finalidad de generar información oportuna, pertinente y veraz que asista a los altos mandos de la organización en la toma de decisiones significativas para el bienestar financiero y estratégico de la misma.

Actualmente en la Universidad de Manizales (UM), concretamente en la dependencia de Admisiones, Registro y Control Académico (ARCA, también conocido como Registro Académico), no se encuentra implementada ninguna solución de reportes inteligentes de apoyo a la toma de decisiones y que ayude a la administración de los numerosos datos provenientes de los procesos académicos que allí se cumplen diariamente. Las razones para que esté sucediendo de este modo es que primero, en ninguno de los sistemas informáticos transaccionales que hoy por hoy emplea la UM en sus procesos misionales se encuentra embebida la tecnología SIG o algunos de sus componentes, segundo no se cuenta con una plataforma integral que mediante el uso del BI permita la integración de los datos distribuidos en diferentes bases de datos y tercero, es que la confiabilidad de los datos almacenados no es totalmente fiable ni relevante para la toma de decisiones en los diversos procesos académicos de la UM.

La información académica que administra la UM tiene más de 40 años, y se encuentra en constante aumento debido a que periódicamente se registran nuevos programas, nuevos planes de estudios, más asignaturas, registro de aspirantes, estudiantes y egresados, etc. Mucha de esta información académica al igual que los demás tipos de información de la Universidad (financiera, médica, administrativa, etc.) se emplea como la principal fuente para la elaboración de muchos de los reportes de indicadores académicos determinados por las autoridades de la UM. Esta información se encuentra en su mayoría en las bases de datos que emplean los diferentes sistemas informáticos operacionales que utilizan las distintas dependencias (Secretaría General, Planeación, División Administrativa y Financiera, División de Desarrollo Humano, Docencia, Proyección

Social, Investigaciones y Postgrados, Mercadeo Institucional y Comunicaciones, ARCA y Tecnologías de la Información, entre otras). Además, no existe cruce alguno de información dado que los datos, como ya se indicó con anterioridad, no están integrados en su totalidad y al tratar de analizarlos sobrecargan los sistemas operacionales actuales, al igual que las plataformas de gestión y administración de las bases de datos. (Acuña Salinas, 2014)

La Universidad de Manizales precisa implementar un procedimiento informático por medio del cual pueda de manera rápida y eficaz administrar y organizar todo el volumen de datos que día tras día es generado a través de la iteración de usuarios finales con sus sistemas de información propietarios y comerciales, y que siendo en la actualidad la localización geográfica (georreferenciada) un factor clave para tomar disposiciones, transforme dichos datos y acerque la información geoespacial a estos usuarios, aprovechándose como apoyo e insumo para decisiones idóneas y oportunas por parte de los altos mandos administrativos de la institución.

## **2. OBJETIVOS**

### **2.1 OBJETIVO GENERAL**

Analizar, diseñar, desarrollar e implementar una solución Geo-BI (Geo-Business Intelligence) aplicada a las fuentes de datos de los sistemas operacionales de Registro Académico de la Universidad de Manizales, con el fin de optimizar el procesamiento de datos geoespaciales y académicos, que sirva como herramienta de apoyo a la toma de decisiones argumentada en los datos.

### **2.2 OBJETIVOS ESPECÍFICOS**

- Describir las características, componentes, aplicación e integración de las tecnologías Geoespaciales y BI que se considerarán en la solución Geo-BI.
- Analizar y caracterizar las fuentes que contienen los datos que interesan, administran e interactúan en el área de estudio.
- Identificar, definir y priorizar 5 indicadores en cobertura, calidad, pertinencia y/o eficiencia académica de los requerimientos de información en Registro Académico.
- Diseñar e implementar un modelo lógico y físico multidimensional y desnormalizado basado en tablas de hechos y dimensiones con el fin de aplicar las correspondientes tareas ETL que permitan poblarlo.
- Esbozar, desarrollar y desplegar un Cuadro de Mando Operativo (CMO) fundamentado y focalizado en los indicadores académicos definidos.
- Implementar mediante el uso de una herramienta de software libre y de código abierto (open source) la solución Geo-BI, estructurando como base de dicha puesta en marcha el Data Warehouse Geográfico (DWG).

### 3. JUSTIFICACIÓN

El ambiente de convergencia de muchas de las necesidades de las grandes y medianas organizaciones pertenecientes al sector público y privado debe concordar con el proceso de toma de decisiones garantes y adecuadas dentro de la UM, permitiendo un oportuno e idóneo tratamiento de la información y de los indicadores de eficiencia involucrados en la transformación de datos de enseñanza y aprendizaje proveniente de diferentes fuentes de información para que se pueda aplicar su integración en un menor tiempo, permitiendo la posibilidad de reaccionar y adaptarse a los cambios o ingreso nuevos indicadores pertinentes.

Por dicha causa el considerar la presentación de información organizada y filtrada al procedimiento de “toma de decisiones responsables y responsablemente, al tiempo que se evalúa el estado actual de la inteligencia de negocios en [la Universidad de Manizales], contribuye a diagnosticar y plantear alternativas de gestión frente al conocimiento y a la estrategia empresarial” (Lozano Mejía, 2011), considerándola más apropiada a los menesteres académicos actuales de la UM, enfocándola a los conjuntos de procesos específicos y llevando finalmente a la implementación de un DW como el motor de datos del SIG orientado al BI (Solución Geo-BI).

Las universidades y en general las instituciones de educación superior se afanan cada vez más por progresar y satisfacer con su propósito misional de enseñar a su entorno; ahora bien, dar a conocer los datos acerca de dicha encomienda, como el conocer: ¿Cuántos estudiantes se han graduado? ¿Qué particularidad posee su población escolar? ¿La institución dirige sus propios dineros?) es un factor primordial de la esencia institucional diariamente. Hacer esto conlleva a la generación de un reto específico, más aún hoy en día, cuando se producen cada vez más datos y cuyas fuentes son aún más heterogéneas; sin embargo, a continuación, se muestran algunas formas de como las universidades y las instituciones de educación superior pueden emplear el análisis y la visualización de datos para contar mejor sus historias:

- Rastreo de la tendencia de inscripciones: Los datos de inscripción de nuevos estudiantes son algunos de los más importantes en una institución. Además, se convierten en información clave para la planificación y el enganche de aspirantes, y se vuelve en un medio predilecto para conocer cuándo se modifican significativamente los datos demográficos.
- Visualización de información de la universidad: Los posibles aspirantes podrían verificar la información pertinente a la matrícula, el cupo del grupo, promedios de estudiantes egresados y otras mediciones coherentes a la toma de decisión sobre si deben ingresar o no a una determinada universidad.

- Persuasión de posibles donadores: Para lograr cumplir sus objetivos, una institución debe ofrecer gran variedad de estadísticas y/o consultas; a través de ellas, los posibles donadores pueden tener acceso a las tasas de retención y graduación, así como poder identificar de donde provienen los estudiantes.
- Entrega de resultados de encuestas: Empleando la visualización interactiva, cada persona puede encontrar en las encuestas los datos que crea relevantes. Por ejemplo, los docentes pueden estar interesados en las actitudes de los estudiantes, mientras que los directivos pueden necesitar el nivel de satisfacción de estos.
- Análisis del empleo del espacio: Es importante para una universidad comprender de qué manera usa el espacio físico ya que es fundamental para un funcionamiento eficaz, desde la planificación de clases al establecimiento de planes de emergencia.
- Comparativa frente a instituciones pares: Ya que todas las universidades en Colombia deben presentar informes al Sistema Nacional de Información de la Educación Superior (SNIES) existe gran riqueza de información disponible para hacer comparativas.
- Muestra de los estados fiscales: En una época cuando el costo de las matrículas sube, los estudiantes, padres y becantes están interesados en saber qué tan bien administran el dinero las instituciones. La visualización de los datos financieros ayuda a que los interesados comprendan el panorama general.
- Mejorar las comunicaciones y el compañerismo: Si se comparten los datos, se puede facilitar el desarrollo de una cultura colaborativa y así mismo compartir los datos puede conducir a distribuir ideas y trabajar en conjunto.

En el presente desarrollo tecnológico enmarcado dentro de una investigación articulada en una de las tres áreas de la formación de la maestría: Gestión de la Información, y a partir de una de las tres líneas de investigación de la misma: Desarrollo de Software; el maestrante considera correcto y apropiado aportar desde el contexto del aprendizaje individual académico al desarrollo institucional con esta tesis de grado acerca de las tecnologías de los sistemas de información geográficos y la inteligencia de negocios para ser aplicada a las fuentes de datos de los sistemas operacionales de Registro Académico de la Universidad de Manizales.

El propósito [de este] trabajo de [grado es] implementar [una solución Geo-BI] como [herramienta de] apoyo a la toma de decisiones [argumentada en los datos], que permita monitorear el estado de los indicadores del proceso enseñanza y aprendizaje [planteados en la fase de identificación de requerimientos], cuyos reportes periódicos faciliten evaluar el desarrollo de los procesos académicos de la [Universidad de Manizales]. Estos reportes son evidencias necesarias para determinar las medidas de contingencia más adecuadas para el área académica. (Acuña Salinas, 2014)



## **4. FUNDAMENTOS TEÓRICOS**

### **4.1 INTELIGENCIA DE NEGOCIOS**

De acuerdo con ESRI “El término inteligencia de negocios fue acuñado a mediados de la década de 1990 para describir la práctica emergente de transformar datos en bruto a partir de los datos operativos heterogéneos de una organización en un almacén de datos común que podría utilizarse para descubrir y reportar información.” (GIS and Business Intelligence: The Geographic Advantage, 2006).

La Universidad de Manizales debe estar al tanto que los requerimientos del negocio en el cual está involucrada como organización privada no son estáticos, sino por el contrario deben estar en continua adecuación teniendo en cuenta las necesidades variables del entorno, por lo que eventualmente es inevitable llegar a la conclusión que no es posible insistir en un enfoque administrativo y de gestión que no sea ágil ni flexible.

#### **4.1.1 Concepto**

La inteligencia de negocios es una evolución de los sistemas de soporte a las decisiones, por lo tanto, BI es un conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de la información, permitiendo tomar mejores decisiones a las organizaciones mediante el análisis y explotación de los datos. (Acuña Salinas, 2014)

En 2009, Kevin Quinn acertó indicando que la inteligencia de negocios es usada de tres maneras diferentes principalmente: estratégica, analítica y operacional, estando conectadas de forma directa y trabajando conjuntamente. Llevando así a que el análisis estratégico impulse el BI analítico, mientras que el BI analítico dirige el enfoque de las iniciativas operacionales.

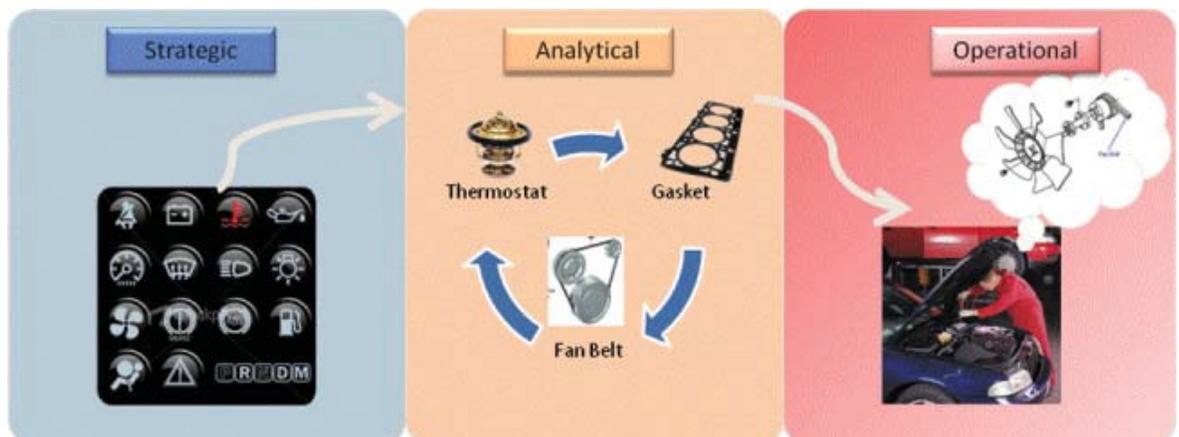


Figura 1. Automobile dashboard analogy (Quinn, 2009)

#### 4.1.2 Componentes

Almacén de datos: existen diversas definiciones de DW en la literatura tradicional y digital, de las cuales sobresalen como las más adoptadas y representativas las expuestas por William Inmon y Ralph Kimball en sus diferentes estudios: Un DW es una colección de datos orientada a temas, integrada, con variación de tiempo y no volátil, en apoyo al proceso de toma de decisiones de la administración. (Inmon, 1992), Un almacén de datos es una copia de datos transaccionales específicamente diseñada para consultas y análisis. (Kimball, 1998). Se evidencia claramente que las anteriores definiciones están concentradas alrededor de los datos, pero un data warehouse es mucho más que solo datos, un DW además de ser el dato es un administrador del procesamiento de este haciendo disponible la información y permitiendo a las personas tomar decisiones. (Anahory, 1997)

ETL: este proceso consiste en reunir las diversas fuentes de datos con las que interaccionan los diferentes sistemas transaccionales de la organización, y posteriormente realizar la depuración y homogenización de estos, para que a continuación sean cargados al data warehouse; todas estas fases suceden a menudo en una “staging area” (área de ensayo), un sitio de tránsito en donde los datos sufren las transformaciones ETL en su traslado desde de la fuente original al DW. Aunque generalmente en el proceso ETL solo se conocen las acciones de extracción, transformación y carga (Extraction, Transformation y Load, por sus siglas en inglés), también hacen parte de este componente del BI el descubrir (donde se analizan las fuentes de información) y transportar (como se llevan los datos ya transformados al DW).

OLAP: este tipo de operaciones permite realizar un análisis de forma multidimensional de los datos con el fin de llegar a obtener información en un alto grado de detalle, de manera rápida y fácil, procesando transacciones en

tiempo real y acomodando los datos en diferentes puntos de vista para el usuario; las OLAP pueden ser clasificadas en ROLAP (Procesamiento Analítico Relacional En Línea) y MOLAP (Procesamiento Analítico Multidimensional En Línea).

Consultas y Reportes: estas herramientas permiten la creación de informes, reportes, listados de grandes volúmenes de información ya transformada por el componente ETL, de forma granular como agregada. Los artefactos o software de consultas y reportes generan una invocación al DW, extrayendo los datos que son más pertinentes y efectúa sobre ellos cálculos adicionales, manipulándolos si es el caso y mostrando los resultados ágilmente en un formato claro y conciso.

#### **4.1.3 Medición**

La implementación de los componentes de BI puede llegar a ser valorado dependiendo del grado de criterio con que la organización muestre frente a la utilización de las tecnologías de la comunicación y la información. Para llevar a cabo esta valoración existen dos métodos claros y bien definidos: los reportes en sí y las herramientas analíticas; con los reportes la organización solamente recupera información sobre qué ocurrió y en algunos casos detallan dicha situación, con las herramientas analíticas la organización adquiere fundamentos y hechos que la pueden llevar a anticiparse al acontecer de nuevos imprevistos.

### **4.2 SISTEMAS DE INFORMACIÓN GEOGRÁFICA**

Los SIG son una tecnología que nació en los departamentos de informática de las universidades ya finalizando los años sesenta, y fue concebida principalmente con el fin de poder incorporar datos a mapas geográficamente referenciados; las propiedades de los SIG son representaciones gráficas de elementos reales como rutas, cuerpos hídricos y zonas boscosas, y componentes conceptuales como fronteras políticas terrestres y/o marítimas.

Abordando el panorama particular de la UM, el SIGUM<sup>5</sup> opera la información académica de la institución, estos datos alfanuméricos son brindados totalmente por sus sistemas de información transaccionales, y poder integrar a dichos datos la propiedad geoespacial, además de añadir un valor agregado incrementaría su uso. Esta interpretación geográfica desplegada en un mapa mostraría relaciones espaciales y predomios que no son identificados en los cuadros tabulares habituales de los datos.

---

<sup>5</sup> Sistema de Información Gerencial de la Universidad de Manizales

### **4.2.1 Concepto**

Un Sistema de Información geográfica es básicamente una tecnología multipropósito desarrollada para guardar, administrar y emplear datos geográficos digitalizados por medio de un grupo de subsistemas útiles para capturar, almacenar, analizar, visualizar y graficar diversa información espacial georreferenciada.

Un SIG es un conjunto de programas y aplicaciones informáticas que permiten la gestión de datos organizados en base de datos referenciados espacialmente, y que pueden ser visualizados mediante mapas (Moldes, 1995) y usan dos formas de representación de datos: ráster y vectorial. Ráster representa un conjunto de unidades regulares constituidas por celdillas o píxeles en forma de mosaico. Vectorial es la estructura de datos que utiliza puntos, líneas y polígonos para describir la información geográfica (Yañez Coto y Taboada, 2005).

### **4.2.2 Componentes**

**Hardware:** el componente de hardware dentro de un sistema de información geográfica está comprendido por el conjunto de aparatos constituido por una computadora y sus periféricos (Real Academia Española, 2017) en el cual el SIG se ejecuta; actualmente puede ser centralizado para el ámbito empresarial por medio de un servidor robusto con características de hardware excepcionales que se encuentre corriendo bajo sistemas operativos tipo UNIX o WINDOWS, o simplemente ser un PC de escritorio; estos tipos de hardware pueden estar operando de manera aislada o en una red.

**Software:** el software SIG es el compendio de artefactos y funciones que le permiten a los usuarios resguardar, analizar y desplegar toda clase de información geográfica georreferenciada; el software SIG se puede abarcar en dos grandes grupos dependiendo del formato que emplean para modelar los elementos geográficos: sistemas ráster y vectoriales, algunos pueden incluso tratar ambos escenarios.

**Datos:** uno de los más importantes componentes de un SIG son los datos en sí, siendo absolutamente esencial que estos sean precios y abundantes ya que constituyen una representación simplificada del mundo real. El éxito del SIG no se garantiza de una manera precisa si los datos no se mantienen actualizados periódicamente. Su representación es otro factor significativo que se debe tener en cuenta en el momento de manipular y/o ingresar la información en el sistema ya que los datos pueden ser de diversos tipos: datos tipo vector, tipo ráster, tipo imagen y tipo atributo o metadatos.

Recursos Humanos: el personal constituye una pieza clave en el funcionamiento del SIG, ya que dicha tecnología limitaría su valor y funcionalidad claramente sin las personas que puedan administrar el sistema y desarrollar planes para su aplicación. Básicamente existen dos tipos de usuarios bien definidos: los especializados y el público en general, los primeros son aquellos técnicos altamente calificados que interactúan el SIG en algunas de sus etapas, los segundos son quienes en algún momento de su vida tuvieron que solicitar información geográfica, bajo este supuesto no tendrían una formación específica.

Metodología: los métodos son planes y aplicaciones de reglas de negocio específicas y bien diseñadas, que describen como dicha tecnología es empleada; sin una metodología de trabajo específica, todo el SIG podría convertirse en una simple base de datos relacional en donde se conservarían los datos.

#### **4.2.3 Medición**

Al existir una referencia espacial para cada elemento dentro de un SIG se hace posible cuantificar otros parámetros igualmente espaciales. Siendo el más básico la distancia, podría hallarse la longitud entre dos puntos específicos o la distancia entre componentes individuales como polígonos o líneas. Aparte de la distancia es factible medir distintos factores como: área, perímetro, longitud de un recorrido no lineal, factores de forma. La medición de diferentes elementos contesta preguntas como:

- ¿Qué cantidad de superficie de zonas con árboles existen en determinado territorio municipal?
- ¿Cuántos kilómetros de extensión posee la red vial colombiana?
- ¿Poseen las zonas de usos de suelo formas espesas o son primordialmente prolongadas y de tipo puntiagudo?

### **4.3. ANTECEDENTES**

#### **4.3.1 ANÁLISIS Y DISEÑO DE UN DATA MART PARA EL SEGUIMIENTO ACADÉMICO DE ALUMNOS EN UN ENTORNO UNIVERSITARIO**

En este proyecto hacemos un análisis y diseño de una herramienta llamada Data Warehouse. Esta tecnología de la información representa el último avance dentro de las bases de datos, y se configura como el entorno idóneo para la consulta y el análisis de la información procedente tanto de los sistemas transaccionales internos, como de las fuentes de información externas de interés para la empresa ... cabe mencionar la existencia del Data Mart que por ahora podríamos decir que son una versión más reducida de un Data Warehouse. Estos Data Mart a menudo contienen información específica de algún departamento concreto de la organización ... a fin de mantener

consistencia de datos corporativos y mantener la seguridad e integridad de la información que se está usando. (Rodríguez Sanz, 2010)

La idoneidad al incluir este antecedente se hace evidente al comprender que los DW<sup>6</sup> y los DM (Data Mart) son elementos fundamentales de la tecnología BI, según Sinnexus:

Un Datawarehouse es una base de datos corporativa que se caracteriza por integrar y depurar información de una o más fuentes distintas, para luego procesarla permitiendo su análisis desde infinidad de perspectivas y con grandes velocidades de respuesta ... Un Datamart es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. (Sinnexus, 2016)

Teniendo en cuenta las conclusiones presentadas en este antecedente se tendrá que plantear una metodología lo suficientemente clara y organizada con la finalidad de lograr desplegar un DM aplicado a un ambiente universitario posibilitando así a los usuarios la obtención de información precisa.

#### **4.3.2 ESTUDIO DE PERFIL DE RENDIMIENTO ACADÉMICO UN ABORDAJE DESDE DATA WAREHOUSING**

En este trabajo se describe un modelo de Data Warehouse para determinar los perfiles de rendimiento académico en la asignatura Algoritmos y Estructuras de Datos de la carrera Ingeniería en Sistemas de Información de la UTN-Facultad Regional Resistencia (UTN<sup>7</sup>-FRRe) ... La característica principal del modelo es que combina aspectos relacionados a las calificaciones del estudiante junto a otras variables tales como factores socioeconómicos, demográficos, culturales, entre otros, en base a lo cual permite clasificar diferentes perfiles de alumnos. (La Red Martínez, Karanik, Giovannini, & Pinto, 2015)

Dadas las heterogéneas fuentes de datos que coexisten actualmente en la Universidad de Manizales se hace absolutamente necesario implementar una solución técnica basada en DW que conduzca a una integración y armonización de los datos mucho más acorde con la necesidad informacional de Registro Académico de la UM, posibilitando así el cruce de distintos tipos de información que sirva como insumo para “los principales productos de Business Intelligence que existen hoy en día [que] son:

- Cuadros de Mando Integrales (CMI)

---

<sup>6</sup> DW, por sus siglas en inglés de Data Warehouse

<sup>7</sup> Universidad Tecnológica Nacional (Argentina)

- Sistemas de Soporte a la Decisión (DSS)
- Sistemas de Información Ejecutiva (EIS)” (Sinnexus, 2016)

### **4.3.3 ANÁLISIS DE RENDIMIENTO ACADÉMICO ESTUDIANTIL USANDO DATA WAREHOUSE Y REDES NEURONALES**

Cada día las organizaciones tienen más información porque sus sistemas producen una gran cantidad de operaciones diarias que se almacenan en bases de datos transaccionales ... Una de las acciones más utilizadas en las instituciones educacionales para dar valor a la información y dar apoyo a la toma de decisiones, es la confección de reportes. La confección de los reportes es una acción exploratoria, es decir, se hacen ciertos cruces de datos y, dependiendo de los resultados, se van analizando otros criterios hasta que se llega a un punto en el cual los resultados son satisfactorios para tomar decisiones sobre la organización ... Las características distintivas de los DW respecto a los sistemas descritos anteriormente es que son flexibles, integran todos los aspectos organizacionales de interés, pueden manejar grandes volúmenes de datos eficientemente, permiten la creación y cálculo de indicadores de gestión. (Zambrano Matamala, Rojas Díaz, Carvajal Cuello, & Acuña Leiva, 2011)

La necesidad informacional de Registro Académico de la UM es el principal escenario en el que se pretende intervenir con este proyecto ya que en este momento los reportes institucionales son generados de dos formas principales: la primera de ellas se basa en la manipulación directa de las bases de datos relacionales que emplean los sistemas operacionales de la UM a través del uso del lenguaje estructurado de consultas (SQL, Structured Query Language), y la segunda forma es mediante el empleo de múltiples hojas de cálculo y datos tabulados dentro de ellas, automatizando procesos repetitivos y complejos.

### **4.3.4 DATA WAREHOUSE Y DATA MINING APLICADOS AL ESTUDIO DEL RENDIMIENTO ACADÉMICO Y DE PERFILES DE ALUMNOS**

El desigual aprovechamiento de las TICs observado en los alumnos ... han motivado la aplicación de técnicas de Almacenes de Datos (DataWarehouses: DW) y de Minería de Datos (Data Mining: DM) basadas en clustering, entre otras, para la búsqueda de perfiles de los alumnos ... según su rendimiento académico, situación demográfica y socio económica, con el propósito de determinar a priori situaciones potenciales de éxito o de fracaso académico. (La Red Martínez, Acosta, Cutro, Uribe, & Rambo, 2010)

Las técnicas de minería de datos y almacenes de datos pueden ser empleadas en muchos ámbitos actualmente utilizando mecanismos de exploración

apoyados en rastreo de información útil dentro de vastos volúmenes de datos, es el caso de la necesidad actual de la UM ya que las bases de datos que soportan todos los procesos académicos y administrativos no proveen de manera directa dichas funcionalidades y por el contrario se deben empezar a construir bajo las técnicas adecuadas e idóneas para tal fin.

#### **4.3.5 DESARROLLO DE UNA BASE DE DATOS INTEGRADA DE CENSO Y ENCUESTA MEDIANTE EL USO DE ELEMENTOS DE INTELIGENCIA DE NEGOCIOS Y SIG**

En este trabajo se plantea el desarrollo de una solución de Inteligencia de Negocios y Sistemas de Información Geográfica (SIG) para gestionar integradamente los datos ... La propuesta plantea el uso de una base de datos integrada [por varias] fuentes mediante tecnología ETL (Extract, Transform and Load), permitiendo obtener datos [homogenizados de la UM] en niveles de desagregación que esta no provee originalmente ... Utilizando los códigos de localización espacial incorporados a un SIG, se añaden herramientas de visualización cartográfica que facilitan la observación y análisis de las relaciones espaciales entre las unidades geográficas, así como la observación y análisis de las particularidades. (Cornejo, Navarrete, Valdivia, Aroca, & Aracena, 2014)

Al obtener datos homogenizados, integrados y desagregados espacialmente se podrá analizar la información geográficamente referenciada y almacenada en las bases de datos de la Universidad de Manizales pertenecientes a todos los actores que intervienen con la generación de datos en la misma, permitiendo así una la capacidad de construir modelos o representaciones de visualización a partir de la información acumulada en el mini mundo de la UM aplicando una sucesión de procedimientos determinados que llegarán a generar valiosa información para el análisis.

#### **4.3.6 SISTEMA DE INFORMACIÓN GEOGRÁFICA PARA LA AYUDA DE TOMA DE DECISIONES EN POLÍTICAS SOCIALES**

El objetivo de esta tesis es dar un panorama general de la construcción de un SIG usando software open source. Este SIG está orientado a la WEB por lo que es un sistema multiplataforma, que se conforma por varias tecnologías que lo hace un sistema único ... lo cual lo hace más eficiente y adaptable, ya que permite tener una imagen manipulable ya sea para realizar un acercamiento sin perder su resolución como el movimiento dentro del mapa (panning), además de proporcionarle interacción entre la aplicación y el usuario ... será de vital importancia al momento de realizar estudios ... gracias a estos ... se

podrá llegar a tomar ciertas decisiones en base a la visualización de mapas. (Leija Luna, 2010)

La importancia que han tenido en los últimos años los sistemas de información geográfica conlleva a que el uso de los mismos tenga cada vez mayor auge, ya que en la actualidad los servicios de servidores de mapas por internet posibilitan visualizar fácilmente distintas capas temáticas, consultar varios de sus atributos y en ciertos casos se puede llegar a realizar operaciones directamente a una base de datos geográfica, permitiendo la posibilidad de visualizar y seleccionar registros que cumplan con determinadas condiciones; esto último siendo de gran interés para el caso específico de la Universidad de Manizales ya que al ser el objetivo de este proyecto la creación de un DWG<sup>8</sup>, el servidor de mapas debe permitir conectarse al data warehouse y éste a su vez a las distintas bases de datos que componen los sistemas transaccionales con los que interactúan los diversos actores de la UM.

#### **4.3.7 IMPLEMENTACIÓN DEL SISTEMA DE INFORMACIÓN EJECUTIVA ACADÉMICO BASADO EN INTELIGENCIA DE NEGOCIOS: CASO UNIVERSIDAD PERUANA UNIÓN**

Hoy en día en el ámbito de los negocios se requieren aplicaciones capaces de analizar, explotar y brindar información eficiente para la toma de decisiones, mayor visibilidad de la gestión y dar soporte a las estrategias. La inteligencia de negocios ofrece escenarios, pronósticos, monitoreo del negocio mediante indicadores y reportes que sirven de apoyo a la toma de decisiones, lo cual se traduce en una ventaja competitiva dentro de la empresa.

El presente trabajo de investigación contiene información sobre las etapas que deben considerarse en la implementación de un Sistema de información ejecutiva de los indicadores académicos de la Universidad Peruana Unión, con el fin de brindar información precisa para la toma de decisiones en la gestión del área académica. (Acuña Salinas, 2014)

Actualmente la Universidad de Manizales posee diversos problemas informacionales sobre el entorno académico que han ido adquiriendo mayor complejidad con el pasar del tiempo, esto sumado al uso diario de sus sistemas transaccionales genera un crecimiento desbordado, en ocasiones desorganizado y continuo de datos que sobrepasa por mucho la capacidad de análisis de las directivas para la toma de decisiones. Es por lo anterior que se pretende en el marco de la inteligencia de negocios, los SIG, la calidad de información e indicadores estratégicos exponer claramente los aspectos metodológicos para la implementación del DWG sujeto a las etapas planteadas por Kimball y al análisis

---

<sup>8</sup> Data Warehouse Geográfico

de la calidad de información de la UM. Procurando aportar a mejorar la gestión y administración de ARCA, así como de las facultades, ofreciendo un apoyo pertinente con los reportes geográficos con indicadores académicos para la toma de decisiones.

#### **4.3.8 DESENVOLVIMENTO DE DATA WAREHOUSE E FERRAMENTA OLAP PARA A ANÁLISE DA PRODUÇÃO ACADÊMICA DE PESQUISADORES ESTUDO DE CASO NO PPGC**

El objetivo de este trabajo es crear una solución para ayudar a analizar la [información] académica ... La solución se basa en el concepto de modelado dimensional en el desarrollo de un almacén de datos para almacenar datos de [información] académica y utiliza una herramienta OLAP para el análisis de datos. Estos datos provienen de [sistemas transaccionales propios y foráneos], cuyo objetivo es recolectar información ... La solución desarrollada en este trabajo permite la visualización, de forma flexible y dinámica, del desempeño productivo [de la UM] a lo largo del tiempo, asistiendo en las decisiones de gestión relativas a la [información] académica realizada. (Rocha Kuplich, 2013)

En el transcurso del progreso tecnológico el hombre ha conseguido desarrollar diversos métodos de análisis de datos que posibilitan optimizar la obtención de información, este objetivo está directamente relacionado con la inteligencia de negocios, dado que en ésta se halla fundamentada la toma de decisiones. Actualmente existen distintas metodologías que logran implementar la acumulación de datos a partir de varias fuentes (bases de datos) con el fin de realizar su análisis histórico y cambiante en el tiempo (almacenes de datos), entre ella está la metodología Kimball que nos muestra el ciclo de vida de un almacén de datos y las fases que hacen falta para su óptima implementación; dentro de dichas fases aparece el modelado dimensional del cual habla Rocha Kuplich.

#### **4.3.9 UTILIZACIÓN DE LA TECNOLOGÍA DATA WAREHOUSE EN INSTITUCIONES EDUCATIVAS**

El contenido del presente trabajo se enfoca al uso de la tecnología Data Warehouse (DW), en instituciones educativas, considerando como caso de estudio al IHESyS (Instituto Hidalguense de Educación Media Superior y Superior). El proyecto surge por la necesidad de solucionar el problema de no contar con un sistema, que solucione las necesidades de tipo informativo dentro de la institución para ayudar en el soporte a la toma de decisiones. El sistema por implementar se enfoca básicamente, a solucionar necesidades de tipo OLAP (On Line Analitic Process) ... Las consultas respecto a múltiples dimensiones necesarias para la toma de decisiones, no se pueden realizar sobre las bases de datos de los sistemas operacionales con que cuenta la

institución. Esto, debido a que los datos no son completamente históricos y la información no está estructurada de forma adecuada para este tipo de consultas. (Cruz Guerrero, 2003)

Claramente se puede determinar que la tecnología DW es aplicable a cualquier ámbito computacional en el cual se haga recopilación y almacenamiento de datos en bases de datos relacionales alimentadas por diferentes fuentes que pueden ir desde simple archivos de texto plano hasta sofisticadas aplicaciones comerciales, sin dejar de lado o en segundo plano los sistemas transaccionales propios o propietarios de las mismas organizaciones. Ámbito en que los datos de la UM se mueven, ya que es poseedora de diversas herramientas y aplicaciones y todas sus dependencias pero que, por varios motivos, que no serán tocados en este proyecto, no se encuentran centralizados ni homogenizados, impidiendo su cruce y complementación unos a otros.

#### **4.3.10 DATA WAREHOUSE COMO INSTRUMENTO DE SUPORTE À AVALIAÇÃO ACADÊMICA**

Junto con la globalización, la información pasó a tener suma importancia para todas las personas e instituciones ... Sin embargo, el exceso de información es desinformación. En cuanto a la cantidad de datos, números, estadísticas, nombres, direcciones, fechas, valores, cantidades y cuentas disponibles ... sumadas a las informaciones disponibles en las organizaciones, instituciones y personas tienden a quedar sin información. Así, proveer información de manera organizada, seleccionada, suficiente y precisa es ciertamente una contribución esencial a las instituciones.

Este trabajo presenta la tecnología Data Warehouse como instrumento de evaluación académica, habiendo desarrollado un estudio de caso junto al CESPE<sup>9</sup>, a la Secretaría de Planificación (SPL) y al Departamento de Ciencias de la Computación (CIC) de la Universidad de Brasilia. (Kenji Fujiwara, 2006)

En muchas ocasiones el adquirir gran cantidad y volumen de datos no necesariamente debe interpretarse como poseer información, y mucho menos considerarla útil para contribuir en la toma de decisiones; en una organización donde sus sistemas transaccionales tradicionales son fuente continua de datos, éstos mismos sistemas deben cumplir con una serie de requisitos que conciben que la información sea confiable y útil para la toma de decisiones. Así pues, nos tropezamos con el requisito fundamental de utilidad, y para que la información sea útil debe ser pertinente, confiable, comprensible, productiva y oportuna.

---

<sup>9</sup> Centro de Seleção e de Promoção de Eventos

#### **4.3.11 DATA WAREHOUSE PARA EL ANÁLISIS ACADÉMICO DE LA ESCUELA POLITÉCNICA NACIONAL**

Actualmente la información se ha convertido en un elemento tan necesario para el surgimiento de un negocio ... por esta razón los datos históricos y la información operacionales de las bases de datos usadas en el negocio se convierte en una fuente inmensa de información valiosa que al ser procesada en un data warehouse puede convertirse en una herramienta que colabora al nivel directivo a tomar sus decisiones.

La Escuela Politécnica Nacional al igual que cualquier “negocio” puede hacer uso de las grandes cantidades de información que tiene a su disposición proveniente de las diferentes escuelas de formación y en cada una sus respectivas carreras, esta información puede llegar a ser de gran utilidad para descubrir posibles falencias en el funcionamiento de la Escuela.

El proceso de desarrollo de este proyecto posee una base teórica a la cual nos referiremos para implementar cada una de las etapas necesarias para ... la construcción de un Data warehouse que en definitiva es una colección de datos que nos ayudará a la tan mencionada toma de decisiones. Para la creación de esta colección de datos se procederá a la recolección de los datos de las diferentes carreras, su correspondiente depuración y transformación y a la culminante presentación de los datos en un formato claro, entendible para el usuario final. (Guevara Lenis, & Valencia Arcos, 2007)

La mayoría de las organizaciones necesitan poner mucha determinación en la toma de decisiones sobre sus negocios y para llegar apropiarse de dichas decisiones se deben tener datos concretos. Así que es imprescindible poseer información apropiada y analizada cuidadosamente, procedimiento que tomaría mucho tiempo y esfuerzo; por tal motivo se requieren artefactos y/o herramientas que ayuden a minimizar el tiempo de dicho análisis, aplicando velocidad y precisión. Asimismo, las instituciones tienen un factor dependiente denominado conocimiento que se halla basado en información comprensible, detallada y relevante, y al existir el conocimiento acertado significa beneficiarse de respuestas precisas para realizar decisiones estratégicas dentro de la empresa.

#### **4.3.12 PERFILES DE RENDIMIENTO ACADÉMICO: UN MODELO BASADO EN MINERÍA DE DATOS**

El rendimiento académico es un factor crítico teniendo en cuenta que, frecuentemente, el bajo rendimiento académico está asociado a una alta tasa de deserción ... En este trabajo se propone la utilización de técnicas de minería de datos sobre información del desempeño de los alumnos ... con el propósito de caracterizar los perfiles de alumnos exitosos (buen rendimiento

académico) y de aquellos que no lo son (bajo rendimiento académico). La determinación de estos perfiles permitiría a futuro definir acciones específicas tendientes a revertir el bajo rendimiento académico, una vez detectadas las variables asociadas al mismo. En este artículo se describen los modelos de datos y de minería de datos utilizados y se comentan los principales resultados obtenidos. (La Red Martínez, Karanik, Giovannini, & Pinto, 2015)

Dentro del entorno de la educación superior se desarrolla el reto de fomentar día tras día la alta calidad académica, sino lo es mejorarla, esto precisa, verificar los temas, estrategias y procedimientos de enseñanza buscando siempre estándares de calidad apropiados que tengan como producto final la constitución de profesionales calificados y útiles para la sociedad. Es por ello por lo que disponer de perfiles se convierte en una actividad muy extendida y es equivalente a la técnica de determinación y clasificación de modelos que se utilizan en el área de la inteligencia artificial y del aprendizaje de máquinas, ya que estos algoritmos retornan información de mucho interés al momento de tomar decisiones, esta información es organizada en un data warehouse que, después de una previa depuración, se analiza por estos mismos algoritmos que realizan minería de datos.

#### **4.3.13 EL IMPACTO DE LAS HERRAMIENTAS DE INTELIGENCIA DE NEGOCIOS EN LA TOMA DE DECISIONES DE LOS EJECUTIVOS**

De acuerdo a Kielstra (2007), los ejecutivos de las pequeñas y medianas empresas no disponen de la información pertinente necesaria para tomar las mejores decisiones de una manera puntual, además según Soto (2001), se debe de proporcionar a los directivos de las empresas unas herramientas adecuadas para la explotación y análisis de los datos que les permitan obtener el conocimiento necesario en el proceso de toma de decisiones estratégicas, por tanto, el problema es que los ejecutivos no toman buenas decisiones debido a que no cuentan con información de calidad y herramientas tecnológicas que les permitan la explotación de la misma para tomar decisiones bien fundamentadas e informadas. (Calzada Cantú & Abreu, 2009)

La trascendencia que tiene la información la convierte en un recurso vital para las organizaciones ya que hacen uso de ella para el avance en sus tareas habituales; esta información se transforma cada vez más en parte fundamental de la empresa, pero para que ésta pueda tener una mayor competitividad y poseer probabilidades de desarrollo, la información que adquiere debe ser pasar por una técnica de sistematización. Ahora bien, como la Asociación de Tecnología de la Información de América (ITAA) define la tecnología de la información como “el estudio, diseño, desarrollo, implementación, soporte o dirección de los sistemas de información computarizados, en particular de software de aplicación y hardware de computadoras.”; esta investigación se encamina en como el empleo de la tecnologías de la información en conjunto con las herramientas de BI posibilitan el

aprovechamiento de la información que tiene una organización para apoyar la toma de decisiones de sus directivos.

#### **4.3.14 ESTADO ACTUAL DE TEMÁTICAS PARA EL ANÁLISIS ESPACIAL EN LA TOMA DE DECISIONES**

Las necesidades de las organizaciones ... han planteado retos en cuanto al análisis de la información ... tener una base de datos de todos los movimientos transaccionales realizados en un histórico de tiempo se volvió algo muy importante. Sumado a esto, surgió un factor que permitiría tener control sobre el lugar de los sucesos y así situar los esfuerzos donde en realidad se necesitan, este factor se identifica con el espacio geográfico almacenado en las bases de datos como tipo de dato geométrico o espacial. A partir de este tipo se crean nuevas formas de análisis y manejo de los datos como la inteligencia de negocios, las bodegas de datos, las consultas, modelado y minería de datos enfocados en el descubrimiento de conocimiento espacial. (Niño Peña & Cáceres Castellanos, 2014)

Actualmente los datos geográficos se convirtieron, tanto o más importantes que los alfanuméricos, debido a ello cada vez un número mayor de organizaciones y empresas están interesadas ampliamente en que sus sistemas transaccionales no solo capturen y almacenen solo datos compuestos por números, letras y otros tipos de símbolos, sino también todo tipo de datos espaciales o geográficamente referenciados; estos distintos tipos de datos al ser acumulados y leídos en periodos de tiempo variantes deben ser almacenados en bases de datos espaciales que, primero puedan soportar la carga y volumen de datos que van guardar en un espacio de tiempo no determinado, segundo logren o permitan que los usuarios lleguen a dichos datos (consultarlos) de una manera fácil e intuitiva, y tercero permitan la expansión de la lógica del negocio ya que éste vive en un entorno en constante cambio evolutivo y normativo; es por lo anterior que las tecnologías y técnicas de análisis espacial aumentan las capacidades de las bases de datos tradicionales permitiendo producir análisis multidimensionales, y simplificando como los usuarios interactúan con los datos a través de consultas alfanuméricas y espaciales.

#### **4.3.15 INTELIGENCIA DE NEGOCIOS: UNA OPCIÓN ANTE LA TOMA RACIONAL DE DECISIONES RESPONSABLES**

Mediante la exploración de la toma de decisiones y la Inteligencia de negocios, se pretende poner sobre el papel algunos elementos que deberían ser centrales en las organizaciones actuales, y que, a causa de factores que afectan su desarrollo, están siendo remplazados por la mecánica toma de

decisiones que, finalmente, conserva poco de los elementos racionales que se supone debería tener. (Lozano Mejía, 2011)

Si se habla de inteligencia de negocios y toma de decisiones se puede intuir como la diferencia principal que se deslumbra entre estas es la utilización de las tecnologías de la información, de aquí que se puede tratar de definir la razón por la cual este par de elementos es también conocido y empleado como un conjunto de tecnologías y metodologías que posibilitan la extracción, transformación y carga de datos de los sistemas tradicionales transaccionales en información estructurada.

#### **4.3.16 TRATAMIENTO DE INFORMACIÓN GEORREFERENCIADA EN EL SISTEMA DE INFORMACIÓN DOCENTE DE LA EDUCACIÓN SUPERIOR EN CUBA**

El Ministerio de Educación Superior cuenta con el «Sistema de Información Docente de la Educación Superior en Cuba», que brinda los servicios necesarios para el apoyo a la toma de decisiones por parte de sus especialistas, pero no cuenta con un soporte para datos espaciales e información georreferenciada. Por otra parte, la integración de los Sistemas de Información Geográfica y los de Procesamiento Analítico en Línea ha conllevado al desarrollo de las herramientas de Procesamiento Analítico Espacial en Línea. Esto permite obtener información más precisa y efectiva, pues los datos representados en mapas son más fáciles de interpretar que los presentados en forma tabular. (Rubiera Hernández et al., 2014)

Al alcanzar una total integración entre los Sistemas de Información Geográfica y el Procesamiento Analítico en Línea se podrá aprovechar las grandes ventajas que estas tecnologías conllevan y, el tener este tipo elementos embebidos en base de datos especiales ofrece un abanico de oportunidades para lograr producir diferentes tipos de mapas espaciales dinámicos que ayuden en el análisis de información tanto académica, como de distribución territorial de la demanda de posibles aspirantes a la educación superior.

#### **4.3.17 DESARROLLO DE UNA SOLUCIÓN GEOBI PARA LA OBSERVACIÓN Y ANÁLISIS DE LA INFORMACIÓN DE CENSO EN CHILE**

Se describe una solución de Inteligencia de Negocios para los censos realizados en Chile en los años 1992 y 2002, con el objetivo de integrar la información de ambos periodos para facilitar la observación y análisis espaciotemporal de los datos. La solución involucró la implementación de un Data Mart que incluye información geoespacial (GeoBI), lo que permite el uso de tecnología SIG y herramientas espaciales de procesamiento analítico en

línea (SOLAP) para la visualización de la información. (Aracena Aguirre, Valdivia Pinto, Navarrete Álvarez, & Cornejo Yáñez, 2014)

Así mismo como en el antecedente acabado de evaluar, en este trabajo se pretende demostrar el desarrollo e implementación de una solución open source que integre las tecnologías SIG e Inteligencia de Negocios con el fin de manipular datos a través del acceso recurrente a un data warehouse geográfico tratado como repositorio integrado de información perteneciente a distintas fuentes y periodos de tiempo, con capacidades y posibilidades de mostrar desagregaciones espaciales. (Geobi, De Geospatial Bussines Intelligence; Wickramasuriya Et Al., 2013).

#### **4.3.18 DATA WAREHOUSING INVESTIGACIÓN Y SISTEMATIZACIÓN DE CONCEPTOS – HEFESTO METODOLOGÍA PROPIA PARA LA CONSTRUCCIÓN DE UN DATA WAREHOUSE**

Debido a que, para llevar a cabo BI, es necesario gestionar datos guardados en diversos formatos, fuentes y tipos, para luego depurarlos e integrarlos, además de almacenarlos en un solo destino, depósito o base de datos que permita su posterior análisis y exploración, es imperativo y de vital importancia contar con una herramienta que satisfaga todas estas necesidades.

Esta herramienta es el Data Warehouse (DW), que básicamente se encarga de consolidar, integrar y centralizar los datos que la empresa genera en todos los ámbitos de una actividad de negocios (Compras, Ventas, Producción, etc.), para luego ser almacenados mediante una estructura que permite el acceso y exploración de la información requerida con buena performance, facilitando posteriormente, una amplia gama de posibilidad de análisis multivariados, que permitirá la toma de decisiones estratégicas y tácticas. (Bernabeu, 2007)

En la mayoría de casos es aplicable y certero decir que la construcción en sí y la implementación de un data warehouse puede ser adaptada de manera fácil e idónea a cualquier parte del ciclo de vida del desarrollo de software, teniendo muy en cuenta que en ciertas etapas, las tareas que se deben ejecutar pueden llegar a ser muy diferentes; igualmente lo que se debe observar detenidamente es no dar comienzo al empleo de metodologías que adviertan extensas fases, ya sean de toma de requerimientos, de desarrollo monótono o de despliegue muy extensas; a o que se quiere llegar es a ofrecer una implementación inicial que complazca una fracción de los requisitos originales para mostrar algunas de las ventajas del DW y estimular en los principales beneficiarios su uso.

#### **4.3.19 DISEÑO DE UN DATAMART PARA EL APOYO EN LA TOMA DE DECISIONES DEL DEPARTAMENTO DE ADMISIÓN DE LA UNIVERSIDAD PERUANA UNIÓN, FILIAL TARAPOTO**

La presente investigación tiene como objetivo diseñar un Datamart (DM) para el apoyo en la toma de decisiones del Departamento de Admisión de la Universidad Peruana Unión, Filial Tarapoto, que sigue el proceso normal de una solución de Inteligencia de Negocios (IN); reunir información de las distintas fuentes de datos, diseñar un prototipo de extracción, transformación y carga de data (ETL), para después mostrar la información a través de gráficos estadísticos. (Guadaña Quiroz, 2016)

En 2011, Zambrano, Rojas, Carvajal, & Acuña, ya mostraban como realizar el análisis de rendimiento académico de estudiantes a través del uso de DW para detallar datos de sistemas transaccionales que mediante distintas operaciones informáticas generaban muy poca o ineficiente información, la cual era guardada en bases de datos relacionales, con el propósito de llegar pronosticar en base al volumen de información almacenada en el tiempo, el rendimiento académico de los estudiantes y las asignaturas aprobadas; en el presente trabajo se pretende ahondar aún más en los diferentes estados por los cuales el estudiantado universitario pasa en su transcurso por la educación superior.

#### **4.3.20 DATAWAREHOUSE CON GEOLOCALIZACIÓN Y CLUSTERING**

Cada día las organizaciones tienen más información porque sus sistemas producen una gran cantidad de operaciones diarias que se almacenan generalmente en bases de datos transaccionales. Con el fin de analizar esta información histórica, una alternativa interesante es implementar un Data Warehouse. Por otro lado, Data Warehouse por sí mismos no soportan análisis geográfico, ni técnicas de clustering. Sin embargo, en la etapa ETL se pueden aplicar técnicas de geolocalización y técnicas de machine learning para clasificar y agrupar información con el fin de mejorar la calidad del análisis histórico. (Zambrano, Rojas & Varas, 2017)

En la actualidad, una de las operaciones más empleadas por parte de las instituciones de educación superior para poder llegar a dar al menos un poco de valor a los datos almacenados y generar información que sirva como apoyo a la toma de decisiones, es la preparación y elaboración de reportes. La creación de los reportes en sí es solo una tarea de reconocimiento, es decir, se van realizando intersecciones y procedimientos de un gran número de datos y según sean los resultados que se vayan observando, se van comparando o añadiendo pautas o criterios hasta que finalmente es posible llegar al estado en cual los resultados obtenidos pueden ser leídos como propicios para tomar decisiones en la organización.

#### **4.3.21 SISTEMA DE INFORMACIÓN GEOGRÁFICA PARA LA INTELIGENCIA DE NEGOCIOS, CASO DE ESTUDIO EN LA SUPERINTENDENCIA DE ECONOMÍA POPULAR Y SOLIDARIA**

La implementación de un sistema de soporte a la toma de decisiones que involucre la Inteligencia de Negocios más los Sistemas de Información Geográfica nos da una visión integral de los indicadores claves tanto de información interna como de información externa de la Organización, lo que nos lleva a reducir el empirismo en la toma de decisiones. Para ello en este trabajo se han utilizado herramientas de Inteligencia de Negocios con el componente Geográfico, con el objetivo de mostrar la estrecha relación entre ambas tecnologías y lo útil que resulta para el usuario final el generar información a través de reportes estáticos, reportes interactivos, tableros de mando utilizando datos transaccionales más relaciones espaciales. Las relaciones espaciales, nos ayudan a identificar características que a través de vistas de datos tabulares no son muy explícitas, es por ello por lo que hablamos de una visión integral de la información. (González Ramírez, 2014)

Ya es bien conocido que el BI da apoyo a la toma de decisiones y que está fundamentada tanto en datos estructurados y como en no estructurados, comúnmente se piensa que basta con hacer inteligencia de negocios con datos planos que son adquiridos de arquetipos de datos que no tienen en cuenta el factor geográfico, ahora bien, lo que se pretende exponer con éste trabajo de investigación es que si se logra acoplar adecuadamente los SIG y BI la toma de decisiones logrará ser objetiva ya que se podrán apropiar elementos que solo pueden ser visualizados si se emplea de manera conveniente el componente geográfico.

#### **4.3.22 ESTADO ACTUAL DE LAS TECNOLOGÍAS DE BODEGA DE DATOS Y OLAP APLICADAS A BASES DE DATOS ESPACIALES**

Las organizaciones requieren de una información oportuna, dinámica, amigable, centralizada y de fácil acceso para analizar y tomar decisiones acertadas y correctas en el momento preciso. La centralización se logra con la tecnología de bodega de datos. El análisis lo proporcionan los sistemas de procesamiento analítico en línea, OLAP (On Line Analytical Processing). Y en la presentación de los datos se pueden aprovechar tecnologías que usen gráficos y mapas para tener una visión global de la compañía y así tomar mejores decisiones. Aquí son útiles los sistemas de información geográfica, SIG, que están diseñados para ubicar espacialmente la información y representarla por medio de mapas. Las bodegas de datos generalmente se implementan con el modelo multidimensional para facilitar los análisis con

OLAP. Uno de los puntos fundamentales de este modelo es la definición de medidas y de dimensiones, entre las cuales está la geografía. (Abril Frade y Pérez Castillo, 2007)

Se debe dar vital importancia a la manera en que los datos, los análisis ejecutados y la información va a ser presentada a los directivos de las organizaciones, ya que para que esta información sea estudiada eficazmente los usuarios finales deben poder verla, examinarla y comprenderla de un modo amigable y accesible, para tal fin los instrumentos o mecanismos a emplear para cumplir tal finalidad, además de que permitan simplemente visualizar la información, no deben requerir por parte de los ejecutivos que los usaran competencias muy técnicas al momento de interactuar con ellos. Para ello se hace indispensable el empleo de recursos como mapas, tablas y gráficos para la presentación de los datos, igualmente es totalmente inevitable que los datos sean mostrados de un modo dinámico, renovándose de forma automática o con cierta periodicidad programada en las herramientas implementadas.

#### **4.3.23 CRM CON TECNOLOGÍAS SPATIAL DATA WAREHOUSE Y DATA MINING BAJO UN ENTORNO DISTRIBUIDO**

Una empresa dispone de infinidad de datos pertenecientes a un cliente: datos transaccionales, datos corporativos y datos departamentales. Por medio de estos datos es posible identificar los atributos y características. En ocasiones la organización carece de soluciones prácticas para la toma de decisiones, afectando a la veracidad de los resultados y a la velocidad de procesamiento de la información. Surge la idea de realizar una integración orientada a la toma de decisiones a través de la aplicación de las tecnologías de Spatial Data Warehouse y Data Mining en un entorno modularizado y distribuido. (Nigro, Miramont and Velázquez, 2003)

Comúnmente las aplicaciones que se encuentran desarrolladas e implementadas en un ambiente transaccional habitualmente lo único que permiten es mecanizar y automatizar los procedimientos o técnicas ya existentes en la empresa, debido a ello es como se logra concebir por qué un DWG (también conocido como Spatial Data Warehouse) converge y se posiciona mucho más adecuadamente a la estrategia de negocios de la organización de lo que podrían llegar a estarlo las anteriores herramientas. Ahora bien, para conseguir lo descrito atrás, el diseño del DWG o SDW tiene que estar dirigido a siempre tratar de optimizar las consultas que guarden de alguna forma correlación con las peculiaridades del negocio de la empresa que se pretender examinar, sin olvidar que las generalidades de los datos tienden a tomar algún tipo representación y posición en el espacio (georreferenciación).

#### **4.3.24 PROPUESTA DE IMPLEMENTACIÓN DE UN WAREHOUSE CON EXTRACCIÓN DE DATOS DE UNA BASE DE DATOS ESPACIAL Y RELACIONAL PARA LA GESTIÓN TERRITORIAL**

El Data Warehouse Espacial (DWE) entrega datos geodescriptivos, un aporte interesante en su implementación es contar con una metodología de construcción que describa paso a paso sus requerimientos. Todo ello, porque en su creación se requiere un acabado análisis de cada uno de los componentes que intervienen directamente en cada proceso, nos referimos a un estudio profundo de las bases de datos espaciales y los conceptos de Data Warehouse Espacial, material que sirven de plataforma para estructurar los procedimientos, los procesos e indicadores que deben ser tomados en cuenta cuando existe interés en este tipo de iniciativa. (Urrutia Sepúlveda y Medina, 2009)

En el momento de diseñar un data warehouse geográfico cobra una relevante trascendencia el considerar una combinación lógica de pasos que posibiliten a los diseñadores y desarrolladores del mismo la apropiada implementación de una base de datos para la administración de información espacial; por ello, cobra gran importancia el anterior estudio para el presente proyecto, ya que plantea un método (conjunto de pasos) que propicia implementar un DW con un componente geoespacial con el fin de realizar consultas de datos geográficos y alfanuméricos de una dependencia específica de la UM.

#### **4.3.25 INTELIGENCIA DE NEGOCIOS ESPACIAL APLICADA A LA PROGRAMACIÓN Y EJECUCIÓN DE PROYECTOS EN EL INSTITUTO COLOMBIANO DE BIENESTAR FAMILIAR**

La aplicación de inteligencia de negocios espacial dentro de la programación y ejecución de proyectos es clave para analizar el estado y la inversión realizada sobre estos en una ubicación geográfica. La inteligencia de negocios espacial es base para la generación de estrategias que permitan estimar los recursos que deben asignarse para el cumplimiento de sus objetivos ... Se busca con ello tener mayor control y organización, identificar problemas presupuestales y de planeación, y contribuir en la toma de decisiones para mejorar la gestión de los recursos y contribuir en la ampliación o creación de nuevos proyectos de acuerdo con las necesidades de los beneficiarios. (Rodríguez Torres y Cáceres Castellanos, 2015)

En muchos casos el empleo de la informática y la tecnología se halla restringida prácticamente al resguardo, consulta, y en algunos casos, formateo de datos, y cada vez más se omite la oportunidad para que los sistemas de información transaccionales se transformen en robustos sistemas estratégicos que verdaderamente asistan la toma de decisiones de la organización. El caso

específico del documento estudiado se encuentra cimentado en las técnicas de preparación y cumplimiento de los objetivos misionales del Instituto Colombiano de Bienestar Familiar (ICBF); con la realización de estos procedimientos se reúne gran volumen de datos acerca de los planes proyectados y emprendidos, y que se localizan geográficamente en diversos municipios. Debido a lo anterior es sumamente primordial que los directivos de la empresa puedan acceder a la información organizada y compilada, de una manera fácil y rápida, que cuenten con algún tipo software que suministre inteligencia de negocios espacial, y apoyen la toma de decisiones de la organización.

## 5. METODOLOGÍA

### 5.1 TIPO DE TRABAJO

Este proyecto se encuentra tipificado dentro de la categoría de Desarrollo Tecnológico, según Salinas:

El desarrollo tecnológico es la modificación que se hace a los resultados de una investigación con el objeto de mejorar su aplicación y su extensión. Se le considera también como el traslado a la práctica, de ser posible en forma masiva, de los resultados de una investigación. (Salinas, 2007)

La Universidad de Manizales cuenta con el Sistema de Información Gerencial de la Universidad de Manizales (SIGUM), que por el momento cumple con las necesidades básicas como apoyo a la toma de decisiones por parte de sus directivos, pero este no cuenta con una plataforma que permita visualizar datos espaciales e información georreferenciada, al igual que no tiene una implementación de Inteligencia de Negocios que considere el uso de sus principales componentes: DM y DW. Además, la integración de las tecnologías de los Sistemas de Información Geográfica y los de Procesamiento Analítico en Línea está directamente implicada en el impulso de los nuevos métodos de Procesamiento Analítico Espacial en Línea. En conjunto estos elementos permiten adquirir información más precisa, oportuna y efectiva, pues los datos que son representados a través de mapas cartográficos y cuadros de mando son más cómodos de interpretar que los mostrados en forma tabular. (Aracena Aguirre, Valdivia Pinto, Navarrete Álvarez, & Cornejo Yáñez, 2014)

### 5.2 PROCEDIMIENTO

BI enmarca un conjunto de actividades y procesos que deben llevarse a cabo para ayudar a las entidades a conocer y aprender de sus negocios, dar valor a sus datos y que sean convertidos en información, y posteriormente en conocimiento, en tanto contribuyan y apoyen la toma de decisiones del negocio; resaltando entre estos procesos acciones de análisis, transformación y generación de reportes, tableros BSC<sup>10</sup>, entre otros.

Ralph Kimball es uno de los considerados expertos originales del DW y BI, y junto a W. H. Inmon son llamados los padres del data warehousing. Kimball y su metodología se transformaron en el estándar para la creación de DW que proporcionen apoyo a la toma de decisiones organizacionales. En el libro “The Data Warehouse Lifecycle Toolkit” del año 1998 se concibe como ciclo de vida a

---

<sup>10</sup> Balanced Scorecard (BSC / Cuadro de Mando Integral)

seguir en el desarrollo de un DW. Siguiendo la metodología Kimball, el proyecto se realizará en 6 fases principales, proponiendo una solución BI + SIG enmarcada en el plano institucional de la UM.

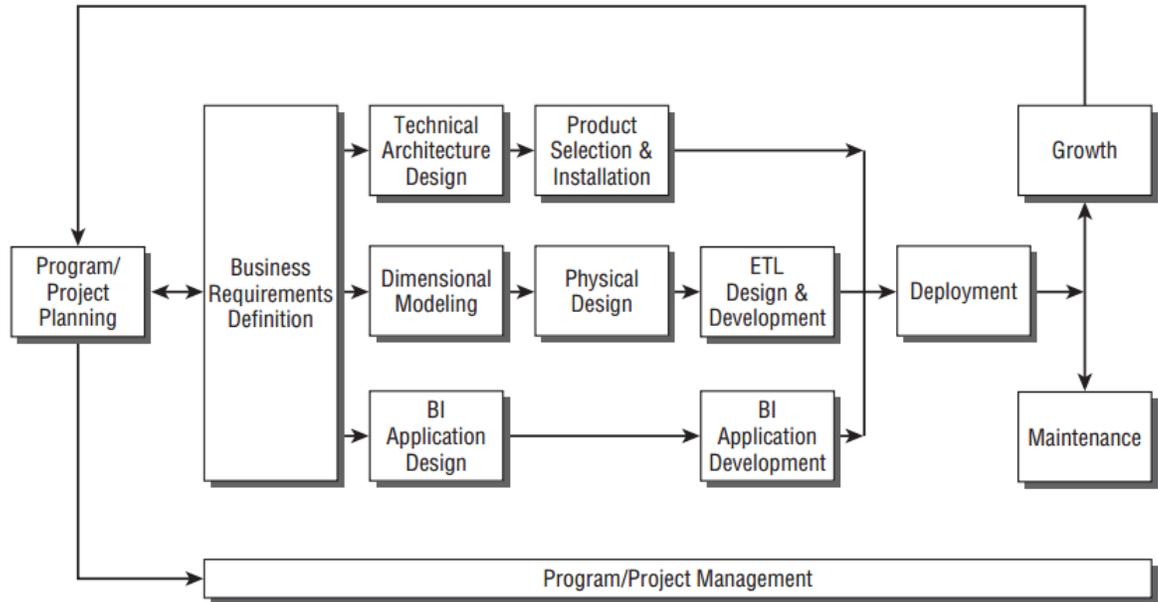


Figura 2. Diagrama del ciclo de vida de Kimball (Kimball, 1998)

### 5.2.1 Fase 1. Planificación del Proyecto

En esta fase se definirá la intención primordial del proyecto, las metas y alcance de este, los peligros potenciales y la necesidad de información que se desea solventar.

A este nivel, en la planificación se establecerá el personal a involucrar, igualmente se identificarán y programarán las tareas, y el cómo será asignada dicha carga laboral, también se determinará la forma de hacer gestión con actividades como: seguimiento y monitorización de procesos, búsqueda de inconvenientes y el mismo progreso del plan del proyecto.

Tabla 1. Cronograma de Actividades

ACTIVIDAD	MESES							
	1	2	3	4	5	6	7	8
<b>Fase 1.</b>								
Planificación del Proyecto	X	X						
<b>Fase 2.</b> Definición de los Requerimientos del Negocio		X						

<b>Fase 3.</b> Modelado Dimensional	X	X			
<b>Fase 4.</b> Diseño Físico			X		
<b>Fase 5.</b> Diseño del Sistema de Extracción, Transformación y Carga				X	X
<b>Fase 6.</b> Especificación de la Arquitectura Técnica		X	X	X	X

### 5.2.2 Fase 2. Definición de los Requerimientos del Negocio

La definición de los requerimientos se dará en gran parte luego de indagar y definir a que personal de la organización se debería entrevistar, para ello se tendrá que dominar la terminología del negocio y el organigrama de la del mismo. Existe un grupo de individuos en particular con el que se deberá hablar desde el inicio: el personal de TI, ya que son las personas que saben realmente qué problemas de datos e información es posible hallar.

Indicadores por contemplar:

- Participación de estudiantes matriculados por procedencia geográfica
- Participación según estratos socioeconómicos de los estudiantes matriculados
- Cobertura de la matrícula financiera por su procedencia académica geográfica
- Tasa de retención de estudiantes
- Proyección de matrículas

### 5.2.3 Fase 3. Modelado Dimensional

La producción en sí de un modelo dimensional se convertirá en un procedimiento dinámico y repetitivo en donde la especificación de los requerimientos del negocio descritos en la fase anterior determinará los datos indispensables que satisfarán los requerimientos analíticos de los usuarios finales.

Básicamente, el proceso de creación iniciará con un modelo dimensional de alto nivel que se obtendrá, fundamentalmente a partir de los indicadores generados en el punto previo, desde aquí se convertirá en una técnica iterativa en donde se identifican esencialmente cuatro pasos: seleccionar el proceso del negocio, donde se detallará la dimensión de cada indicador, y fijará el nivel de granularidad del mismo, y por último se establecerán las dimensiones y las

tablas de hechos; así se dará forma al modelo dimensional o mapa dimensional.

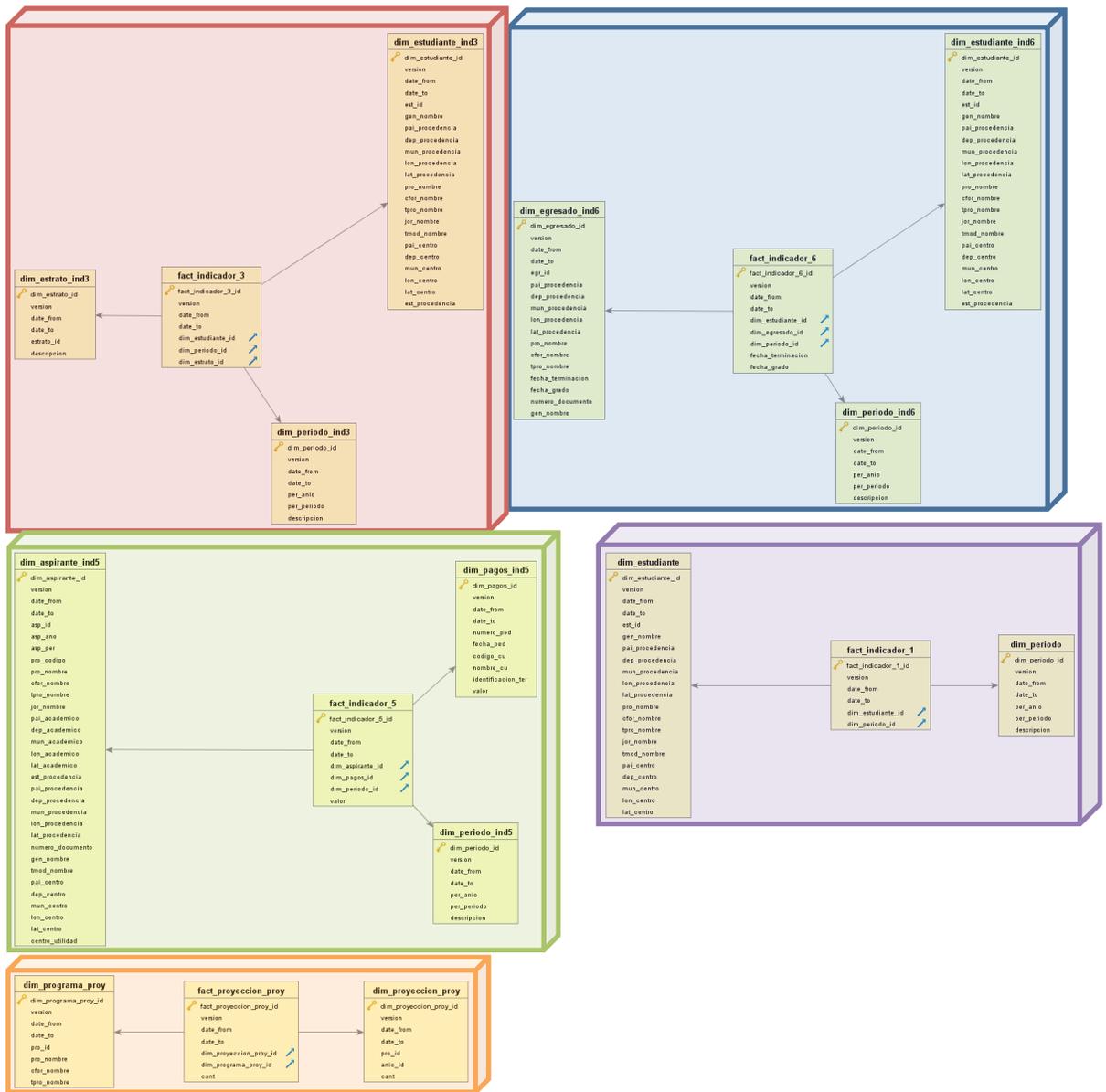


Figura 3. Modelo dimensional

### 5.2.4 Fase 4. Diseño Físico

La fase del diseño físico se centralizará en la elección de la configuración necesaria que sostendrá el diseño lógico ya planteado. Una pieza clave en este proceso es la especificación de estándares dentro el contexto de la base de datos, principalmente la indexación y el particionamiento a utilizar (en el

particionamiento o agregación se debe llevar la información a un nivel atómico).

Ahora bien, se debe procurar contestar los siguientes interrogantes:

- ¿De qué forma se podrá transformar el modelo de datos lógico en un modelo de datos físico en la base de datos?
- ¿Qué se necesitará para obtener un estándar de indexación?
- ¿Será correcto emplear la agregación o partición en tablas relacionales de la base de datos?

Con el fin de resolver estas incógnitas se debe tener en cuenta que la solución a las mismas conduce casi siempre al desarrollo de tablas agregadas; éstas tablas deberán ser interpretaciones reducidas del modelo dimensional, asociando la granularidad de las dimensiones con la tabla de hechos, añadiendo indicadores de detalle a un nivel superior.

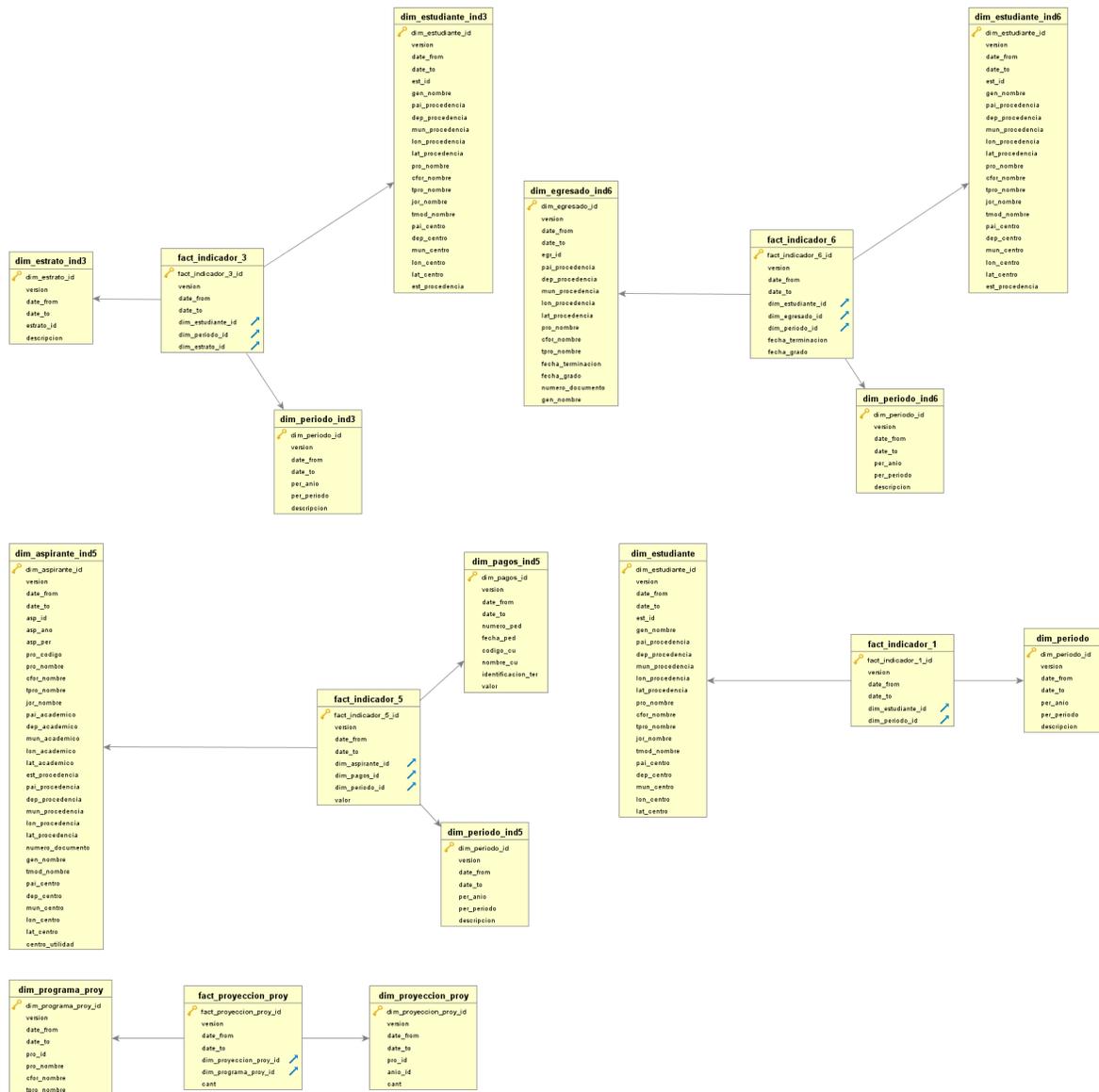


Figura 4. Modelo físico

### 5.2.5 Fase 5. Diseño del Sistema de Extracción, Transformación y Carga

Esta etapa del proyecto es tradicionalmente una de las más menospreciadas, sin embargo, es la raíz de la cual se nutre el DW. El sistema ETL se deberá diseñar de manera adecuada, es decir que permita, implementar en los procesos de extracción las funcionalidades necesarias para producir datos que posibiliten realizar su modificación; igualmente, definir los procedimientos de transformación para poder alterar o ajustar los datos provenientes de diversas fuentes a fin de lograr su carga de forma efectiva; por último, ejecutar diferentes reglas y técnicas de carga de datos al modelo físico

diseñado en la fase anterior; todas estas acciones son eminentemente críticas ya que son las tareas requeridas para poblar el data warehouse.

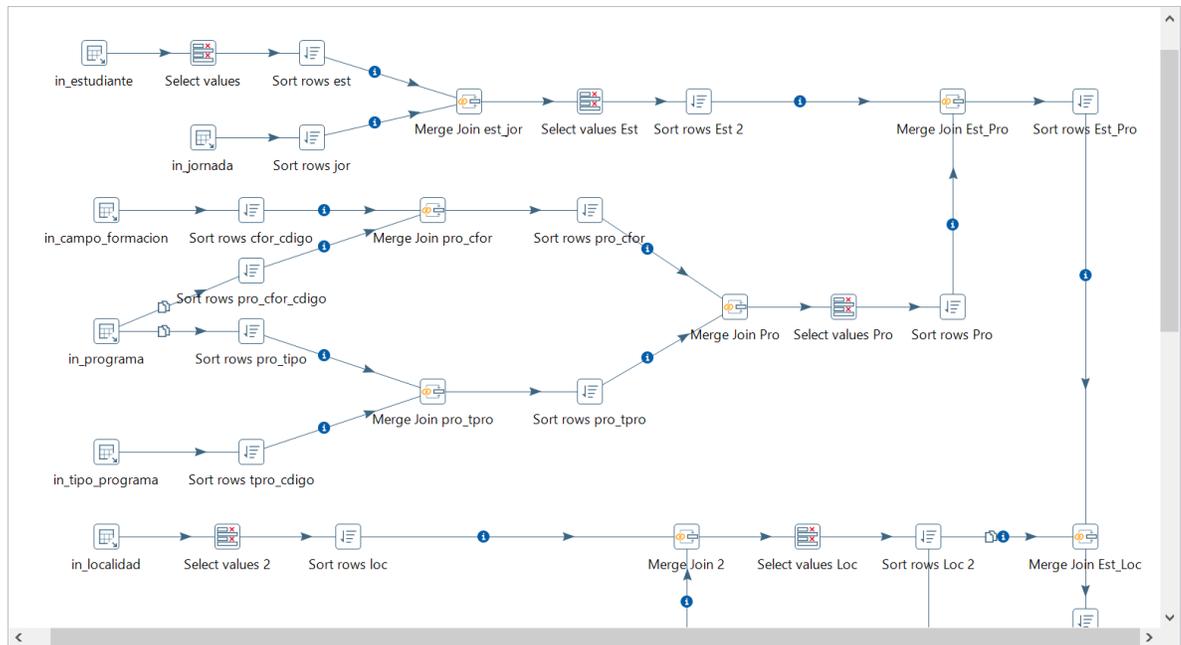


Figura 5. ETL dim\_estudiante

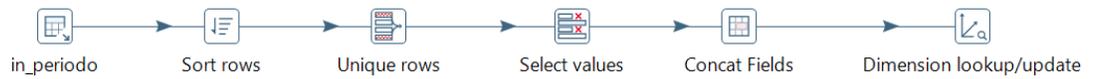


Figura 6. ETL dim\_perodo



Figura 7. ETL dim\_estrato



Figura 8. ETL dim\_programa

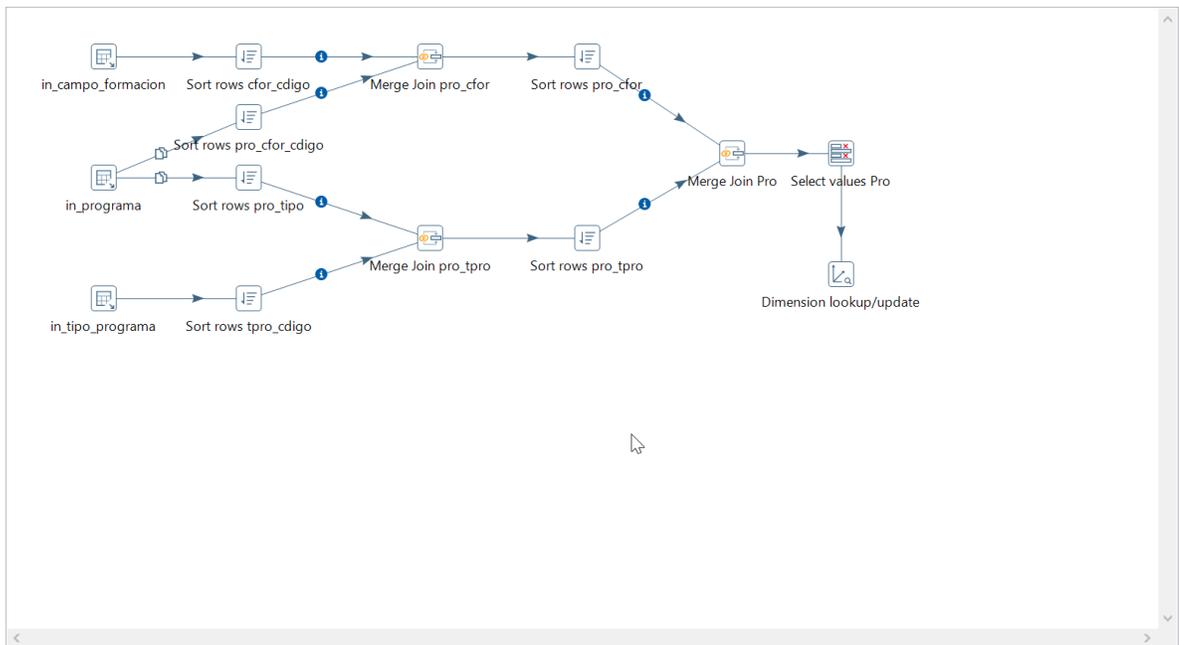


Figura 9. ETL dim\_campo\_formacion

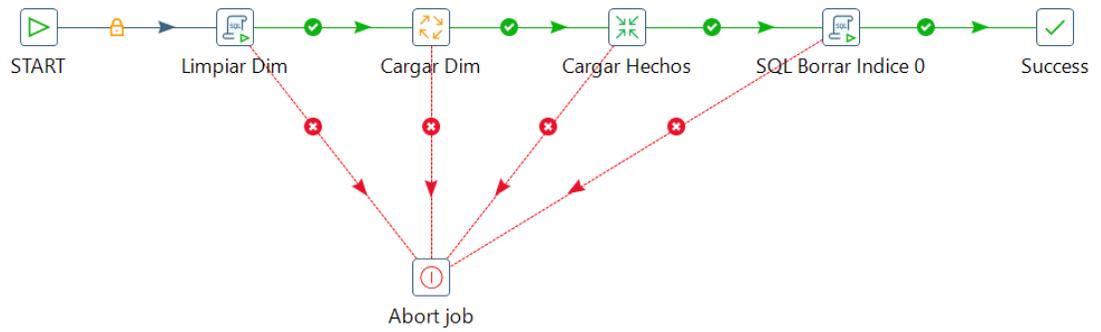


Figura 10. Job

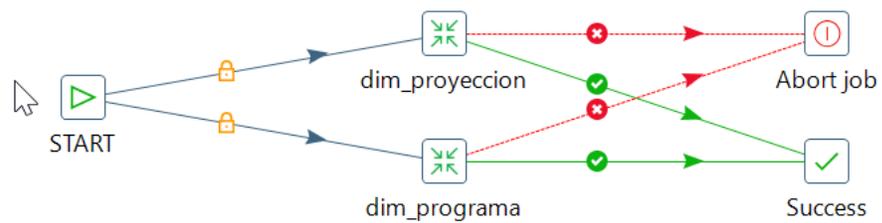


Figura 11. Job proyeccion



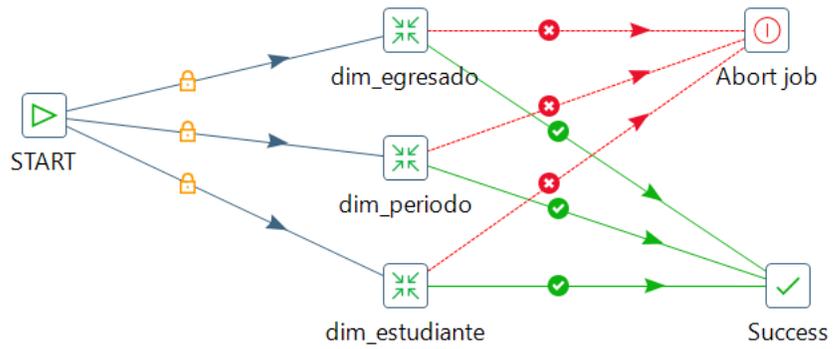


Figura 14. Job egresado

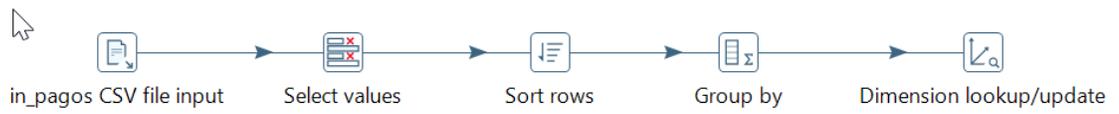


Figura 15. ETL matriculas

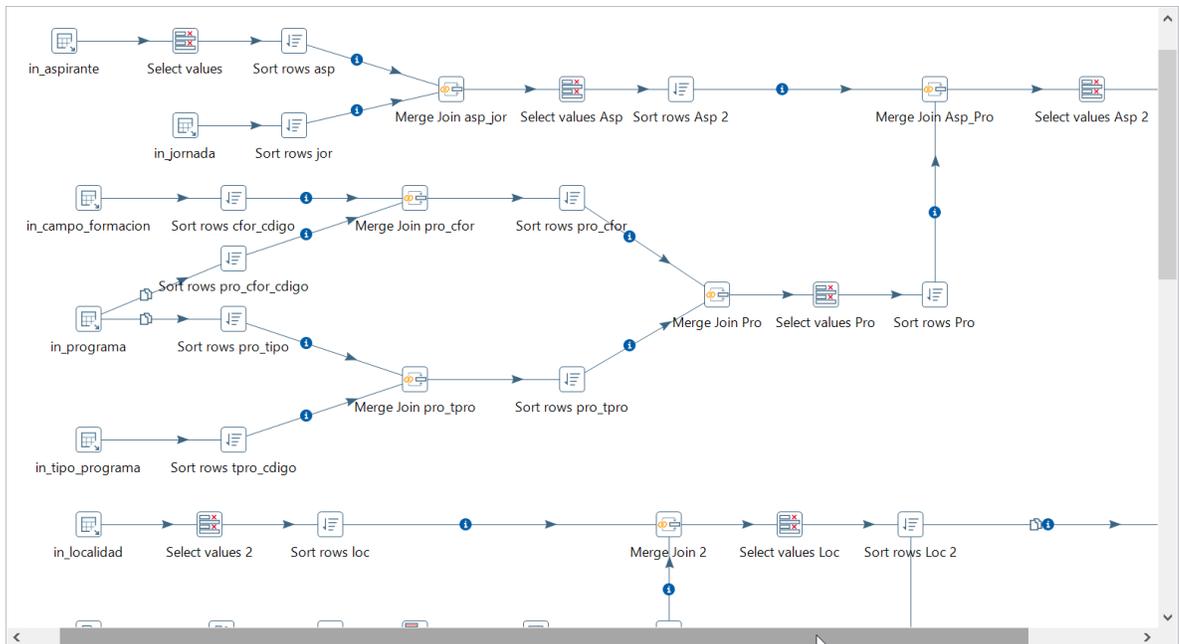


Figura 16. ETL aspirante

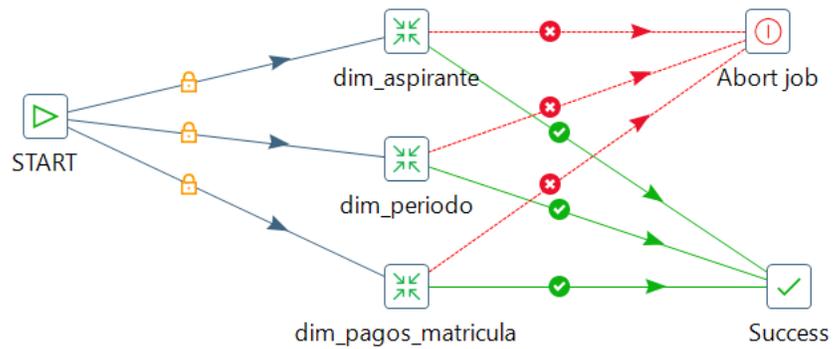


Figura 17. Job aspirante

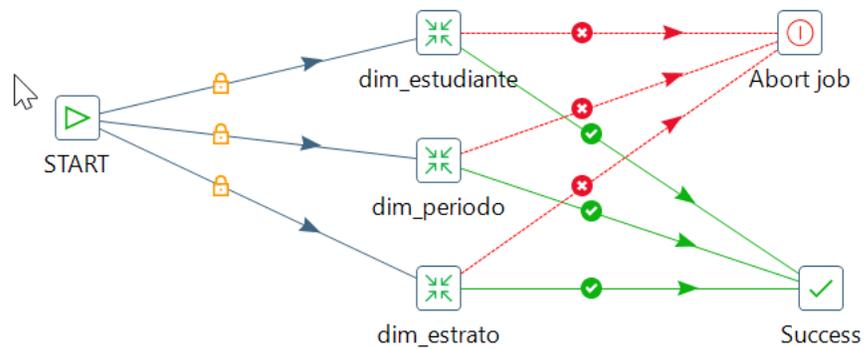


Figura 18. Job estudiante

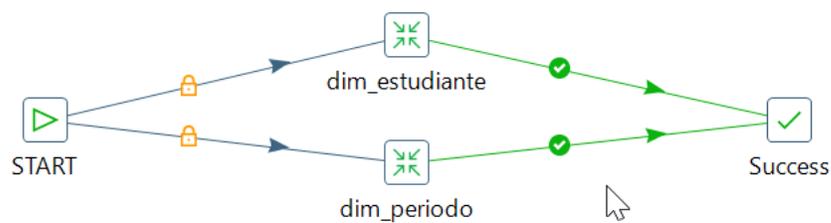


Figura 19. Job periodo

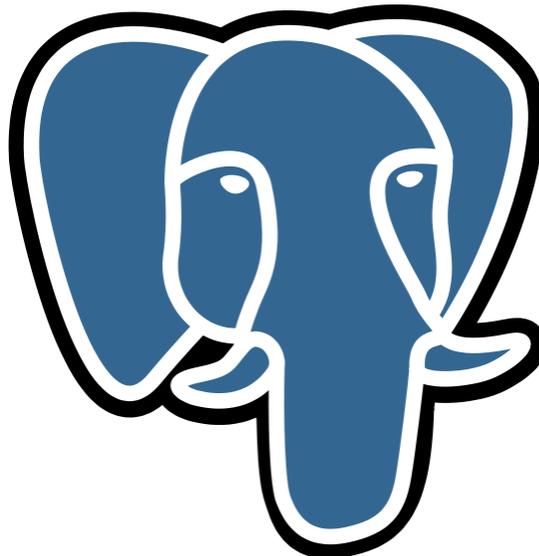
5.2.6 Fase 6. Especificación de la Arquitectura Técnica

El ámbito de implementación de un DW requiere la cohesión de diferentes tecnologías, convirtiéndose en parte fundamental del proyecto y proporcionando a los distintos actores o usuarios del negocio una forma más

organizada y, por ende, más simple de acceder e interactuar con el almacén de datos.

Así pues, se deben tener en cuenta tres elementos esenciales para lograr implantar la especificación de la arquitectura técnica del DW: los requerimientos del negocio definidos en la segunda fase del proyecto, conocer la forma en que los requisitos del negocio impactan en la arquitectura, y finalmente, determinar los componentes técnicos específicos y convergir dicha tecnología, datos y aplicaciones haciéndolos accesibles para los usuarios de la organización.

Bajo la premisa de emplear herramientas de software libre se opta por definir como Sistema Gestor de Base de Datos el motor PostgreSQL; Spoon para los procesos ETL perteneciente a la plataforma Pentaho Data Integration; y emplear Qlik Sense Desktop para visualización.



*Figura 20. PostgreSQL Database*



Pentaho Data  
Integration



Pentaho  
server

*Figura 21. Pentaho Data Integration (Spoon)*



Qlik<sup>®</sup> Sense

*Figura 22. Qlik Sense Desktop*

## 6. RESULTADOS

### 6.1 DESCRIPCIÓN DE RESULTADOS

Alrededor del 80 por ciento de los datos tienen un componente geográfico, con inteligencia geoespacial<sup>11</sup> las preguntas clave para el negocio: dónde, qué y por qué pueden ser respondidas; los datos espaciales ahora pueden ser capturados rápida y fácilmente, cada localización geográfica tiene gran cantidad de información; por eso debe hacerse visible y accesible tan fácil como sea posible, al crear mapas geográficos que presenten información espacial se pueden usar para apoyar procesos de toma de decisiones dentro de la organización.

GeoBI es BI que utiliza información geoespacial. La Inteligencia de Negocios soporta una mejor toma de decisiones que conduce a nuevas oportunidades de ingresos, una mejor visibilidad de los costos y una mejor gestión de riesgos. La inteligencia de negocios (BI) se refiere a las técnicas computarizadas que se emplean para identificar, extraer y analizar datos. Las funciones comunes de las tecnologías BI son los informes, cuadros de mando, cuadros de indicadores, OLAP, minería de datos, minería de procesos, análisis de gestión del rendimiento empresarial y análisis predictivo, entre otras.

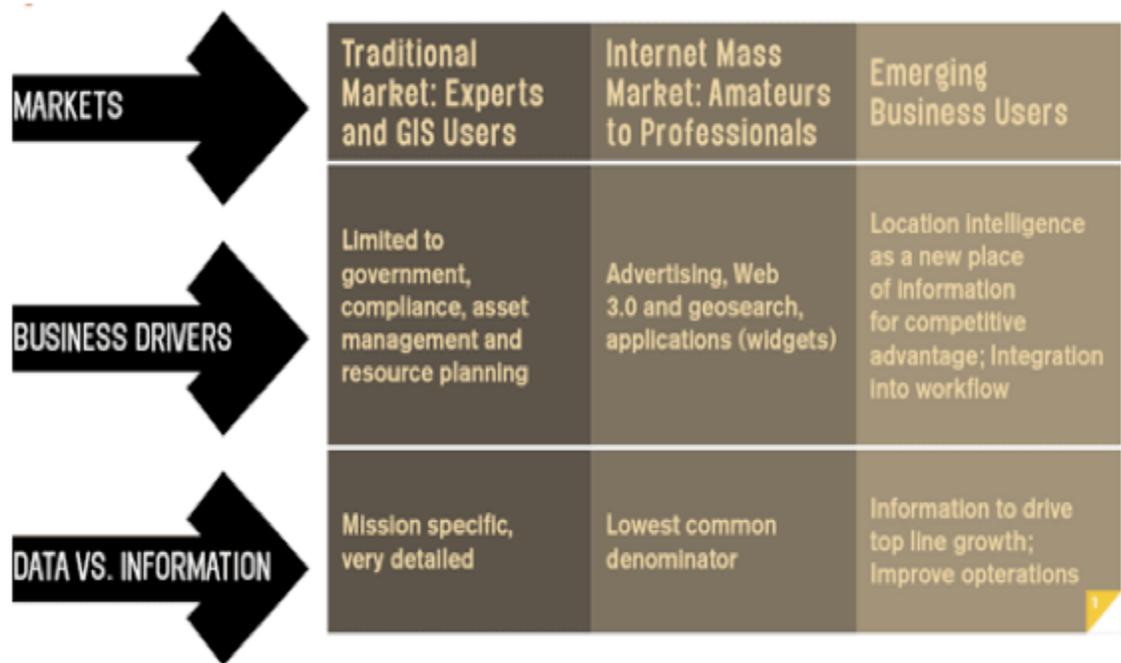
Los sistemas tradicionales de información geográfica (GIS) y el software de procesamiento de imágenes proporcionan potentes herramientas para tratar los datos geoespaciales; con el advenimiento de la Web, y en parte debido a los estándares de OGC<sup>12</sup>, los datos geoespaciales se han convertido en un factor determinante para muchas aplicaciones de mercado masivo, como el mapeo web y los servicios basados en ubicación (LBS<sup>13</sup>).

---

<sup>11</sup> Capacidad de obtener información a partir del componente geográfico de los datos, con base en un parámetro de ubicación, éstos se pueden visualizar, vincular y enriquecer con otras fuentes de datos.

<sup>12</sup> Open Geospatial Consortium

<sup>13</sup> Location Based Services



Source: ITF Advisors

Figura 23. Beyond Traditional GIS/Imagery and Mass Markets (Imaging Notes, 2008)

Mediante el empleo de la metodología Kimball para la construcción de un DW se logró obtener un claro y detallado análisis y posterior caracterización de los datos pertenecientes a las fuentes de información con las cuales interactúan los sistemas transaccionales y operacionales de Registro Académico de la Universidad de Manizales; por ejemplo, la área financiera cuenta con el sistema APOTEOSYS, sistema de información para la gestión financiera y administrativa adaptado a la Universidad, este software está enfocado a empresas del sector público, salud, educación y servicio financieros, el soporte se realiza por una empresa externa a la Universidad, en cuanto se refiere a modificaciones de la funcionalidad; el sistema de Registro Académico, fue desarrollo, complementado y modificado por la dependencia de Tecnologías de la Información de la Universidad, actualmente el mantenimiento y mejoras se realiza por parte de la misma dependencia.

Fruto de ésta metodología se generó el siguiente procedimiento como marco ideal para la producción y total ejecución de los objetivos planteados en la propuesta investigativa, ésta serie de pasos a seguir se encuentran delimitados por fases de trabajo, seis para ser exactos, dentro de las cuales se cumple con parte de la metodología Kimball, ya que ahondar en ella no permitiría llegar a resultados de manera ágil y rápida; las fases anteriormente mencionadas se definieron teniendo en cuenta una secuencia lógica de eventos a cumplir para producir el objetivo general de la actual propuesta, éstas fases son:

- Fase 1. Planificación del Proyecto
- Fase 2. Definición de los Requerimientos del Negocio
- Fase 3. Modelado Dimensional
- Fase 4. Diseño Físico
- Fase 5. Diseño del Sistema de Extracción, Transformación y Carga
- Fase 6. Especificación de la Arquitectura Técnica

Una vez realizada la socialización del proyecto con los actores de Registro Académico que intervienen en la producción de información a partir de los datos, se seleccionaron, identificaron, definieron y priorizaron los siguientes 5 indicadores dentro de los cuales se logran evidenciar intervenciones en cobertura, calidad, pertinencia y/o eficiencia tanto académica como poblacional:

Tabla 2. Indicador 1

Nombre del Indicador	Participación de estudiantes matriculados por procedencia geográfica		
<b>Definición</b>	Este indicador muestra la cantidad de estudiantes que se encuentran matriculados en los diferentes programas de pregrado y posgrado de la Universidad distribuido geográficamente por su lugar o municipio de procedencia		
<b>Forma de Cálculo</b>	Cociente resultante entre el número de estudiantes matriculados en un programa y el número total de estudiantes matriculados, en un periodo determinado, multiplicado por 100% expresado como porcentaje (%)	<b>Componentes de la Fórmula de Cálculo</b>	
		<b>Numerador</b>	Población estudiantil efectivamente matriculada en un programa de pregrado o posgrado
		<b>Fuente del Numerador</b>	SIGUM
		<b>Denominador</b>	Población total estudiantil efectivamente matriculada
		<b>Fuente del Denominador</b>	SIGUM
<b>Unidad de Medida</b>	Por 100 (%)		
<b>Niveles de Desagregación Geográfica</b>	Nacional	<b>Periodicidad</b>	<b>Serie Disponible</b>
	Departamental Municipal	Semestral	1974 - 2017
<b>Observaciones</b>	En algunos casos se deberá suplir el dato de procedencia geográfica con los datos del lugar de nacimiento ya que en series de datos muy antiguos dicha información no se solicitaba o requería, adicionalmente se podrá tener una desagregación académica por tipo de programa al que se encuentra vinculado el estudiante		
<b>Versión: 1.3</b>	<b>Fecha:</b> 16/09/2017	<b>Elaborado por:</b> Julián A. Acevedo Noreña	

Tabla 3. Indicador 2

Nombre del Indicador	Participación según estratos socioeconómicos de los estudiantes matriculados
<b>Definición</b>	Este indicador muestra la participación que existe de cada uno de los estratos socioeconómicos de los estudiantes matriculados en los programas de pregrado y posgrado de la Universidad, según su lugar de procedencia

<b>Forma de Cálculo</b>	Fracción resultante entre el número de estudiantes diferenciados socioeconómicamente y el número total de estudiantes matriculados, en un periodo determinado, multiplicado por 100% expresado como porcentaje (%)	<b>Componentes de la Fórmula de Cálculo</b>	
		<b>Numerador</b>	Número de estudiantes en un estrato socioeconómico determinado
		<b>Fuente del Numerador</b>	SIGUM
		<b>Denominador</b>	Población total estudiantil efectivamente matriculada
<b>Fuente del Denominador</b>	SIGUM		
<b>Unidad de Medida</b>	Por 100 (%)		
<b>Niveles de Desagregación Geográfica</b>	Nacional Departamental Municipal	<b>Periodicidad</b>	<b>Serie Disponible</b>
		Semestral	2017
<b>Observaciones</b>	En algunos casos se deberá suplir el dato del estrato socioeconómico de procedencia con los datos del estrato socioeconómico de residencia ya que en series de datos muy antiguos dicha información no se solicitaba o requería, adicionalmente se podrá tener una desagregación académica por tipo de programa al que se encuentra vinculado el estudiante		
<b>Versión: 1.0</b>	<b>Fecha: 16/09/2017</b>	<b>Elaborado por: Julián A. Acevedo Noreña</b>	

Tabla 4. Indicador 3

<b>Nombre del Indicador</b>	<b>Cobertura de la matrícula financiera por su procedencia académica geográfica</b>		
<b>Definición</b>	Este indicador muestra la cobertura de la matrícula financiera que ingresa a la Universidad en los programas de pregrado desagregada por su procedencia académica geográfica, con el fin de obtener espacialmente representado lo lugares donde entra más dinero por concepto de matrículas		
<b>Forma de Cálculo</b>	Producto final de sumar el valor de la matrícula de los estudiantes en un periodo determinado	<b>Componentes de la Fórmula de Cálculo</b>	
		<b>Sumando</b>	Valor de matrícula de los estudiantes
		<b>Fuente del Sumando</b>	APOTEOSYS
		<b>Agregador</b>	Procedencia académica
<b>Fuente del Agregador</b>	SIGUM		
<b>Unidad de Medida</b>	Valor Absoluto		
<b>Niveles de Desagregación Geográfica</b>	Nacional Departamental Municipal	<b>Periodicidad</b>	<b>Serie Disponible</b>
		Semestral	1997 - 2017
<b>Observaciones</b>	En algunos casos los estudiantes financian el valor de la matrícula para el período en curso y solo cancelan una cuota inicial, sin embargo, se tomará como valor total de la matrícula del estudiante, el valor lo que deberá pagar a la Universidad antes de culminar sus estudios en un periodo determinado		
<b>Versión: 1.4</b>	<b>Fecha: 16/09/2017</b>	<b>Elaborado por: Julián A. Acevedo Noreña</b>	

Tabla 5. Indicador 4

<b>Nombre del Indicador</b>	<b>Tasa de retención de estudiantes</b>		
<b>Definición</b>	Este indicador muestra la tasa de retención de estudiantes de pregrado, es decir, estudiantes que ingresan por primera vez a la Universidad y continúan sus estudios hasta graduarse, dispuesto geográficamente por municipio de procedencia académica		
<b>Forma de Cálculo</b>	Cociente entre el número de estudiantes nuevos matriculados y el número total de estos estudiantes graduados, en un periodo determinado, multiplicado por 100% expresado como porcentaje (%)	<b>Componentes de la Fórmula de Cálculo</b>	
		<b>Numerador</b>	Número de estudiantes nuevos matriculados
		<b>Fuente del Numerador</b>	SIGUM
		<b>Denominador</b>	Población total estudiantil que obtuvieron grado
		<b>Fuente del Denominador</b>	SIGUM
<b>Unidad de Medida</b>	Por 100 (%)		
<b>Niveles de Desagregación Geográfica</b>	Nacional	<b>Periodicidad</b>	<b>Serie Disponible</b>
	Departamental Municipal	Semestral	1972 - 2017
<b>Observaciones</b>	En algunos casos se deberá suplir el dato de procedencia académica con los datos del lugar de procedencia ya que en series de datos muy antiguos dicha información no se solicitaba o requería, adicionalmente se podrá tener una desagregación académica por tipo de programa al que se encuentra vinculado el estudiante		
<b>Versión: 1.2</b>	<b>Fecha: 18/09/2017</b>	<b>Elaborado por: Julián A. Acevedo Noreña</b>	

Tabla 6. Indicador 5

<b>Nombre del Indicador</b>	<b>Proyección de matrículas</b>		
<b>Definición</b>	Este indicador de proyección muestra la cantidad de futuros estudiantes que se matricularían en los diferentes programas de pregrado de la Universidad desde el año 2018 hasta el año 2060		
<b>Forma de Cálculo</b>	Se seguirá el Método de Pronóstico Holt-Winters	<b>Componentes de la Fórmula de Cálculo</b>	
		<b>Valores Conocidos</b>	Cantidad de estudiantes matriculados en el tiempo
		<b>Fuente de los Valores Conocidos</b>	SIGUM
<b>Unidad de Medida</b>	Valor Absoluto		
<b>Niveles de Desagregación Geográfica</b>	Ninguna	<b>Periodicidad</b>	<b>Serie Disponible</b>
		Anual	2018 - 2060
<b>Observaciones</b>			
<b>Versión: 1.3</b>	<b>Fecha: 18/09/2017</b>	<b>Elaborado por: Julián A. Acevedo Noreña</b>	

Un vez realizado el proceso de toma de requerimientos se diseñó un modelo lógico del negocio, en donde las tablas de hechos y dimensiones son las grandes protagonistas, ya que gracias a ellas, se pueden obtener descriptores de los datos

solicitados y especificados en los indicadores de los requerimientos y así, almacenar y procesar grandes cantidades de datos de una manera adecuada e idónea para su consumo tecnológico; luego de la identificación, definición y descripción de las tablas a generar se lograron implementar los siguientes modelos multidimensionales en estrella:

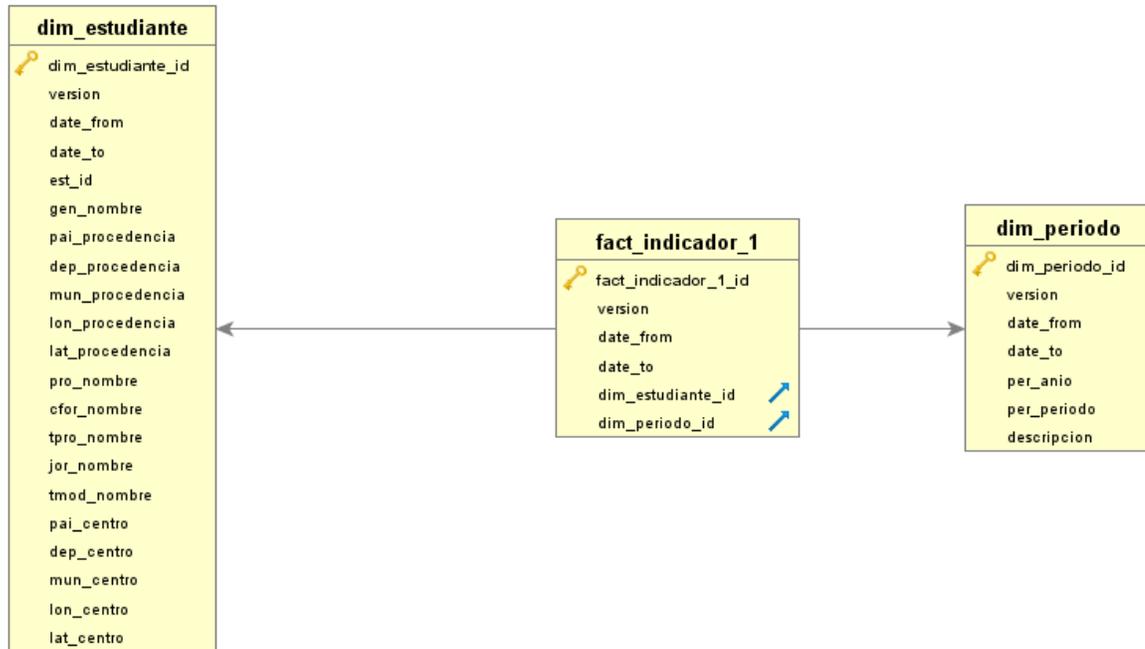
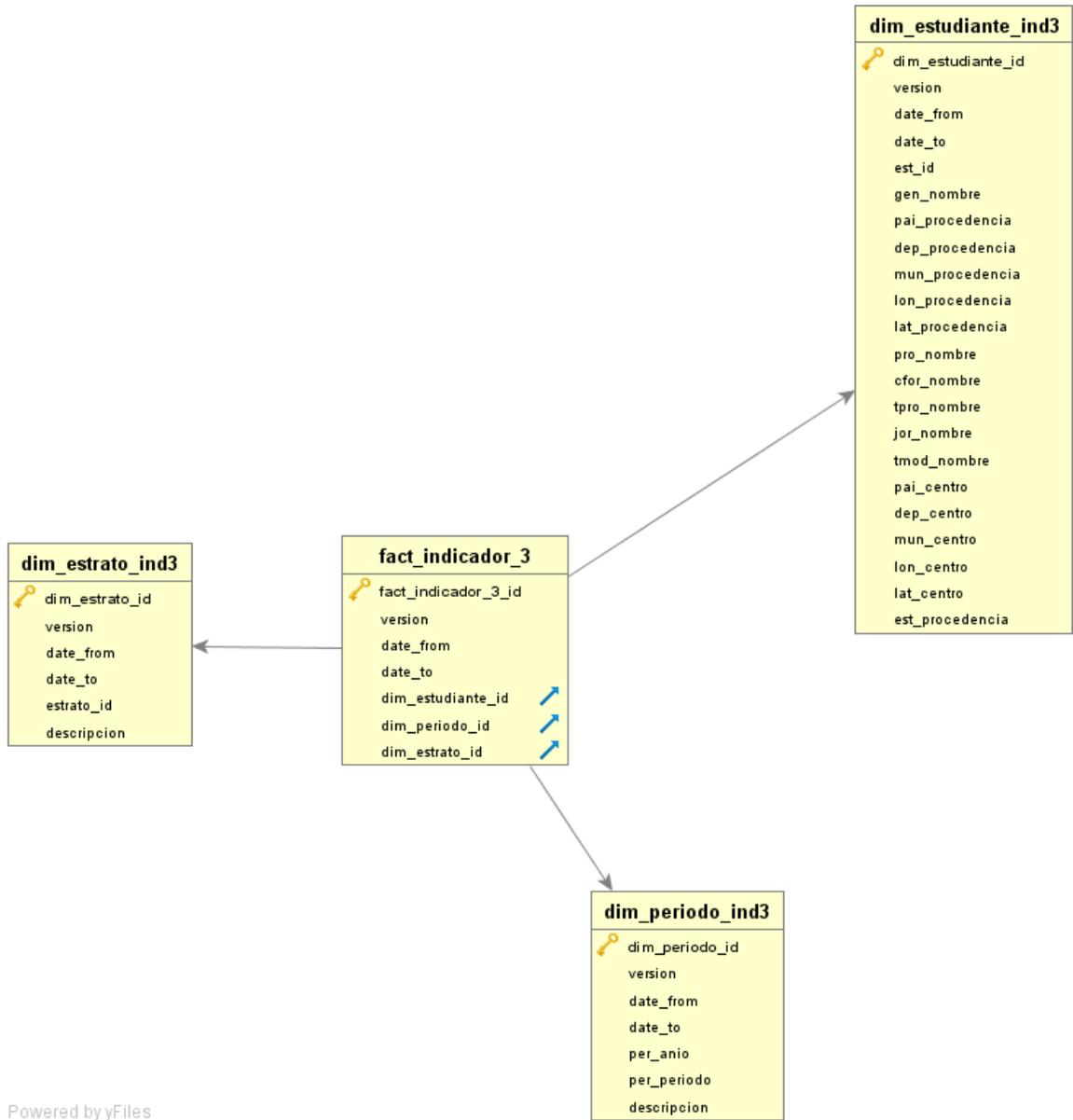


Figura 24. Modelo dimensional del Indicador 1



Powered by yFiles

Figura 25. Modelo dimensional del Indicador 2

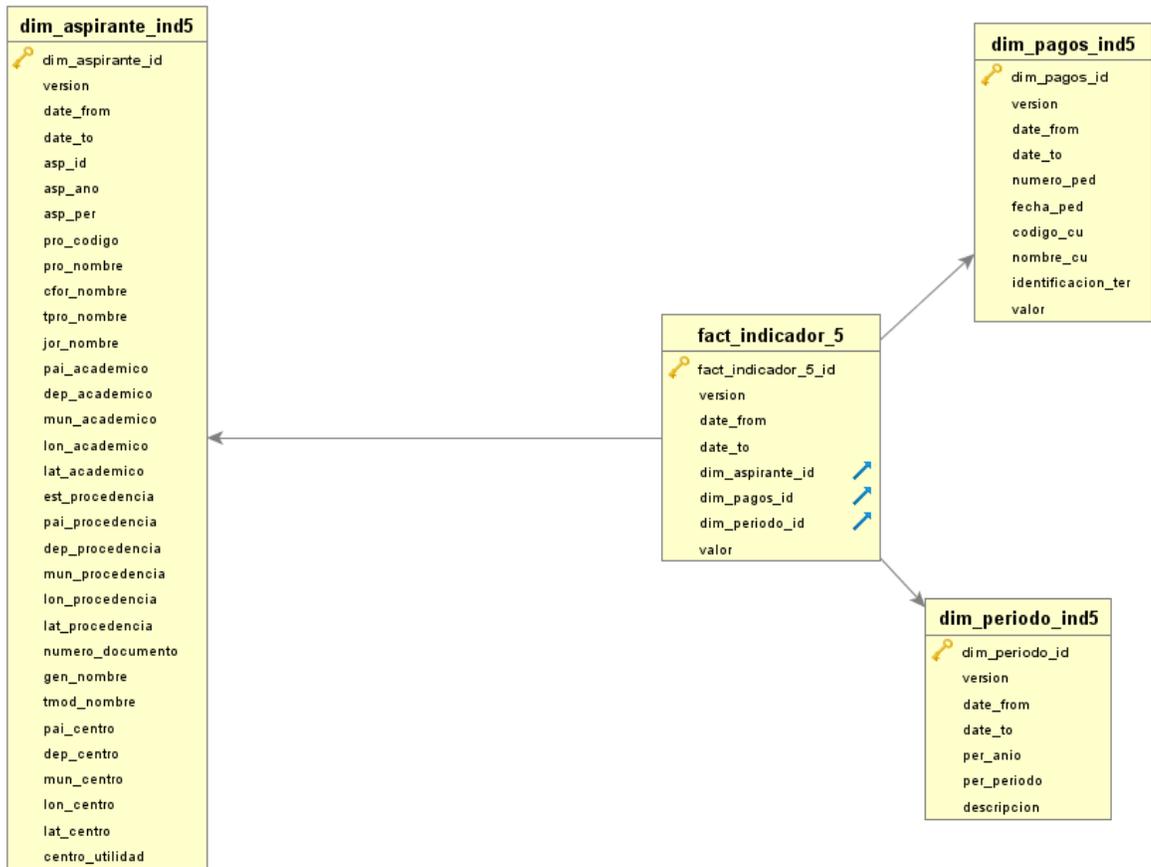


Figura 26. Modelo dimensional del Indicador 3

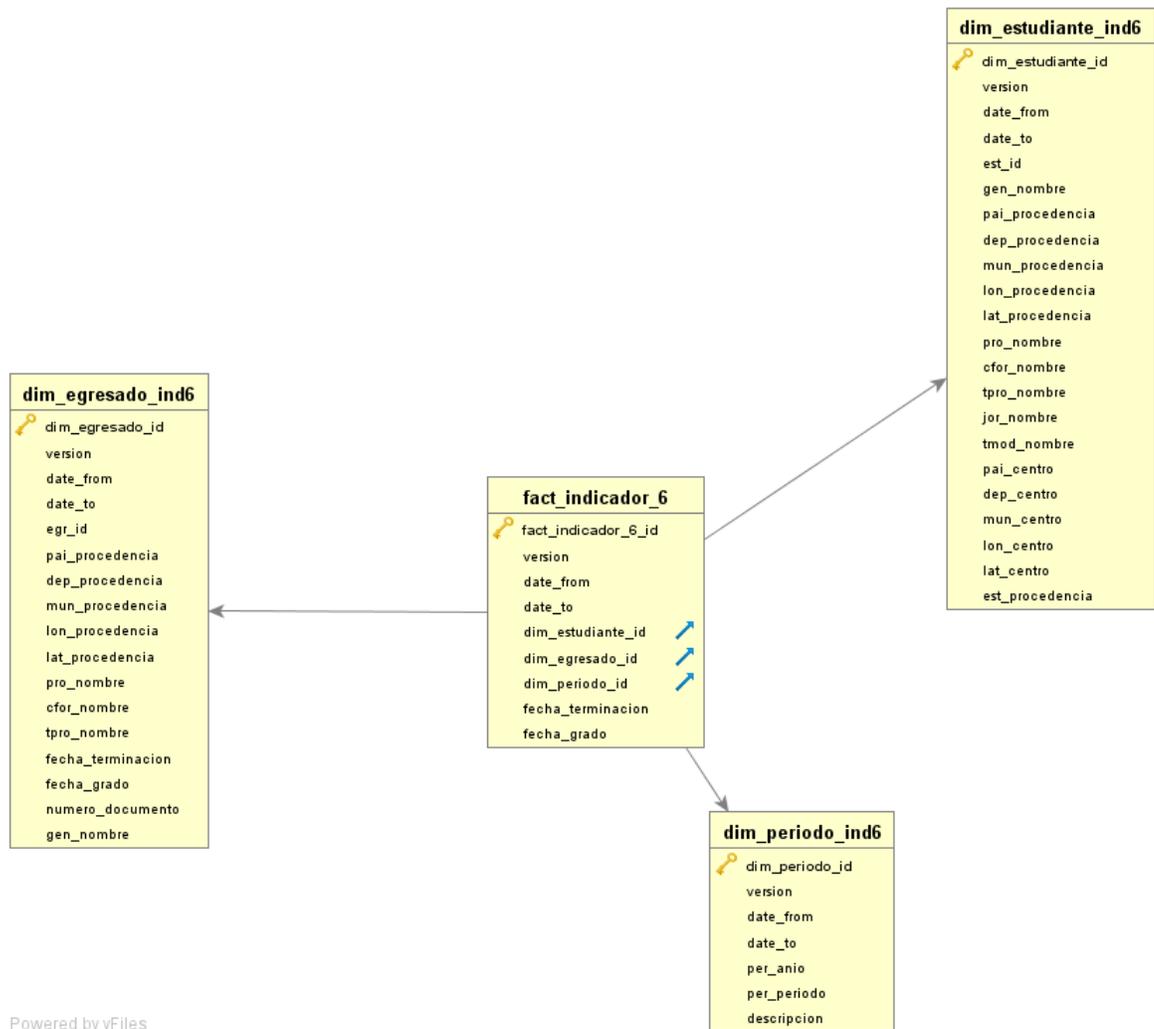


Figura 27. Modelo dimensional del Indicador 4



Figura 28. Modelo dimensional del Indicador 5

Ya que se han producido los diseños lógicos correspondientes a los indicadores de los requerimientos en los cuales fueron definidas y estructuradas las tablas de hechos y dimensiones, se procede entonces con la

última fase para terminar el diseño del DW, implementando el diseño físico en el Sistema Gestor de Base de Datos elegido, y verificando constantemente que se trate de la forma más eficaz para recuperar y almacenar los datos; obteniendo así un enfoque de tipo de arriba abajo (top-down), en donde cada DM será optimizado para un indicador individual:

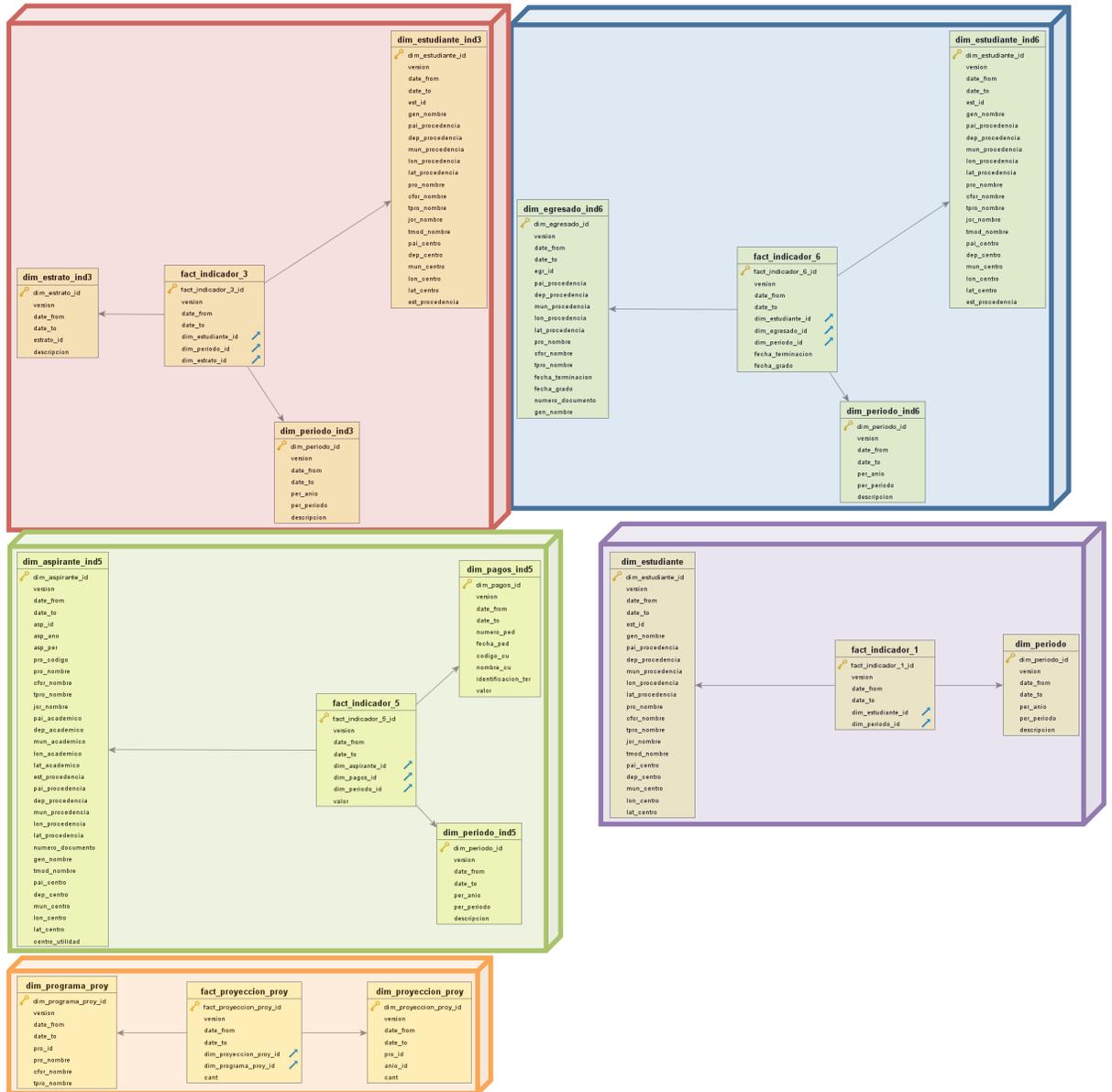


Figura 29. Modelo físico del datawarehouse

Con el fin de poblar las tablas de hechos y dimensiones que fueron diseñadas e implementadas en las fases anteriores se debe recurrir a herramientas tecnológicas específicas que permitan extraer los datos de las bases de datos

operacionales en las cuales los sistemas transaccionales realizan todas sus operaciones de entrada y salida de datos, y luego posibiliten realizar procesos de limpieza a los datos con la finalidad de obtener solo y únicamente la información que es necesitada con precisión, y para terminar ejecuten cargas masivas de datos al DW construido en la fase del diseño físico; por cuestiones meramente de espacio en el presente documento solo se incluirán las figuras del proceso ETL del indicador 1:

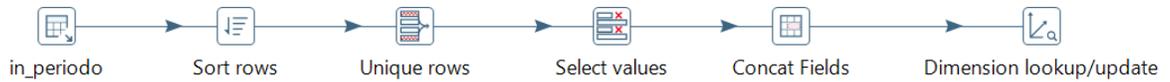


Figura 30. Transformación para cargar la tabla de dimensión dim\_período del Indicador 1

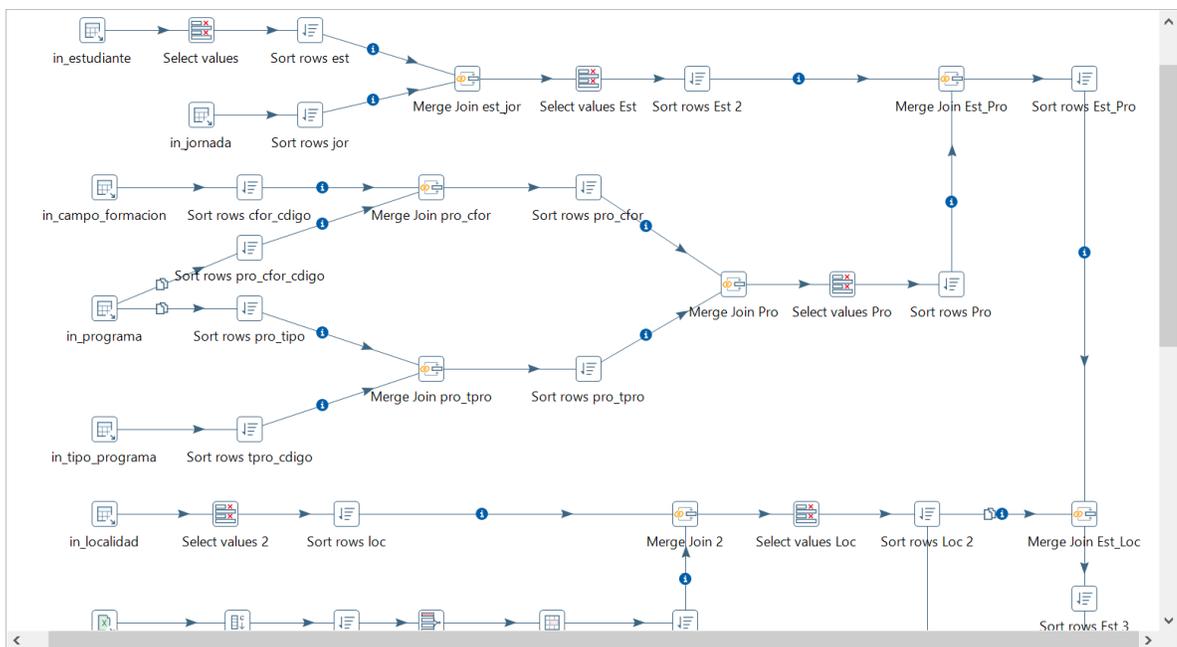


Figura 31. Transformación para cargar la tabla de dimensión dim\_estudiante del Indicador 1

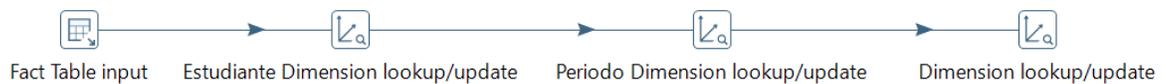


Figura 32. Transformación para cargar la tabla de hechos fact\_indicador\_1 del Indicador 1

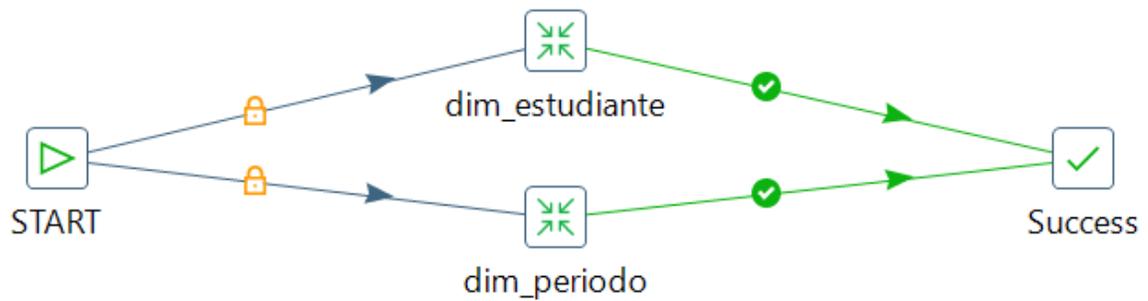


Figura 33. Trabajo para cargar las tablas de dimensiones del Indicador 1

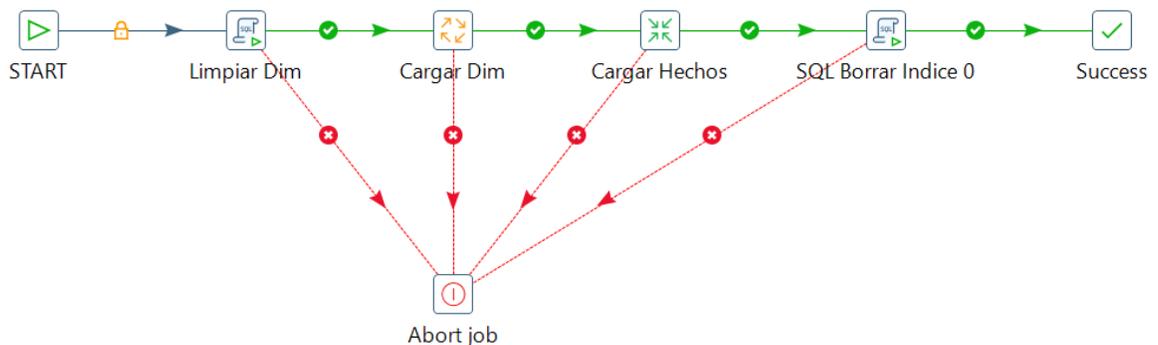


Figura 34. Trabajo para ejecutar todo el ETL del Indicador 1

Dándole la importancia que se merecen los tres elementos descritos anteriormente y bajo la premisa de emplear herramientas de software libre se opta por definir como Sistema Gestor de Base de Datos que albergará el corazón del DW el motor PostgreSQL; Spoon como el diseñador gráfico de transformaciones y trabajos para los procesos ETL, perteneciente a la plataforma Pentaho Data Integration; y emplear Qlik Sense Desktop para crear dashboards e informes interactivos con tablas y gráficos, con el fin de simplificar el análisis de datos y ayudar a la toma de decisiones.

Mediante el empleo de la metodología Kimball para la construcción de un DW se logró obtener un claro y detallado análisis de los datos pertenecientes a las fuentes de información con las cuales interactúan los sistemas transaccionales y operacionales de Registro Académico de la Universidad de Manizales, posteriormente se consiguió generar el diseño de los diferentes DM a ser utilizados y que se encontraban basados en la toma de requerimientos inicial y en el panorama global de la lógica del negocio a la que se estaba aplicando dicha metodología, ahora, con este análisis y diseño se pretendía implementar una solución tipo Geo-BI como insumo primordial para el desarrollo tecnológico a realizar, utilizando herramientas adecuadas para ello, en algunos casos alguna de éstas ofrece los recursos suficientes para todas las fases del procedimiento definido, sin embargo, en el caso del presente proyecto se emplearon varios de estos artefactos para llegar a un fin útil y factible, siendo este la implementación de

un producto Geo-BI y que tenga como fin optimizar el procesamiento de datos geoespaciales y académicos, y sirva como herramienta de apoyo a la toma de decisiones argumentada en los datos.

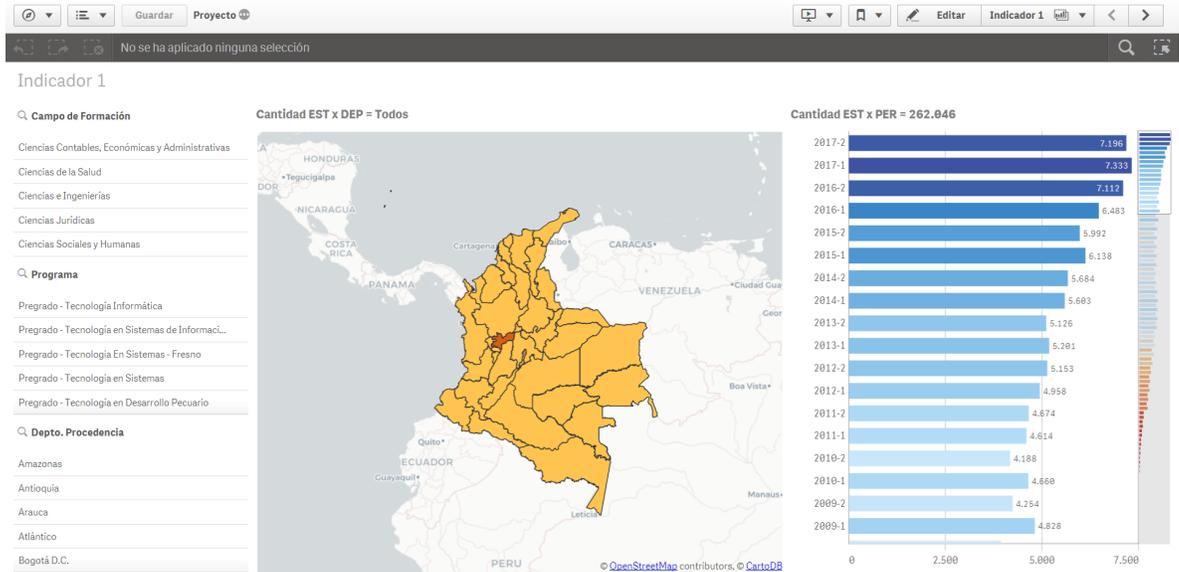


Figura 35. Cuadro de Mando Operativo del Indicador 1

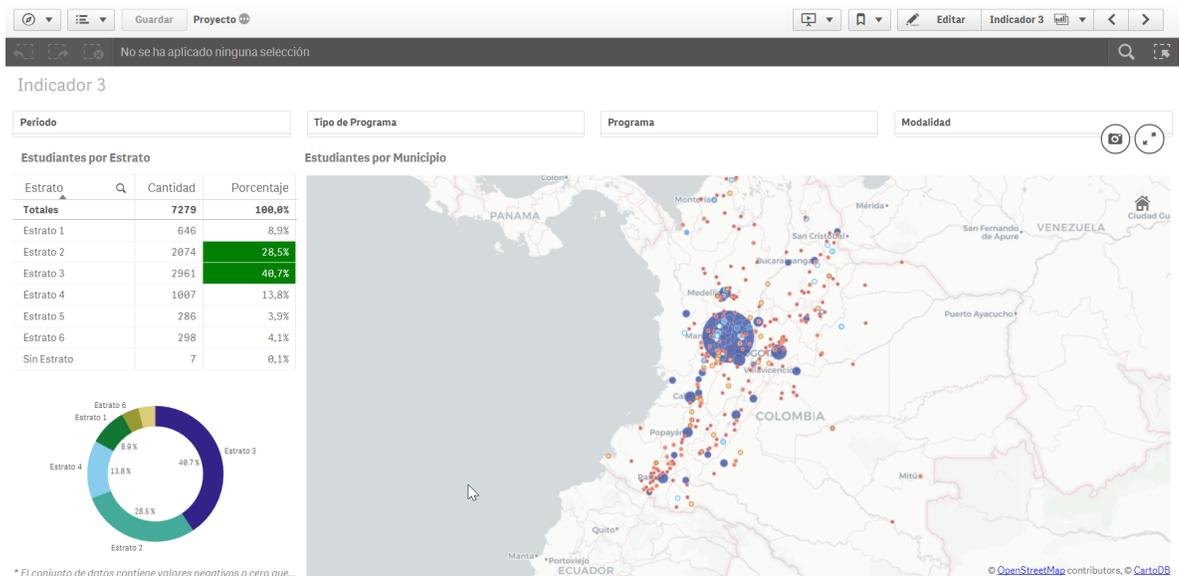


Figura 36. Cuadro de Mando Operativo del Indicador 2

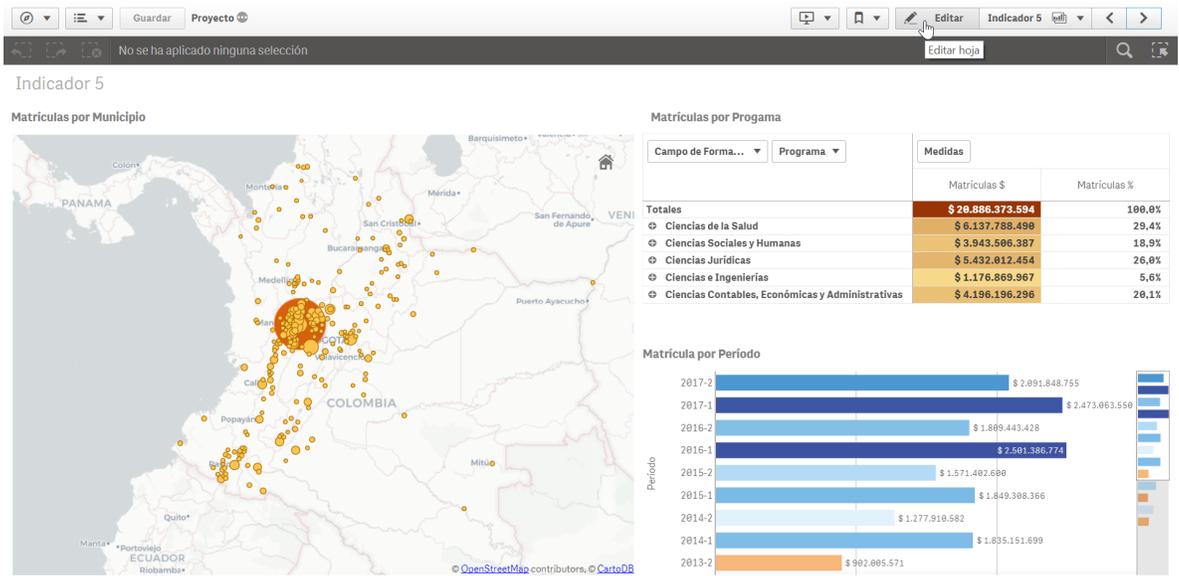


Figura 37. Cuadro de Mando Operativo del Indicador 3

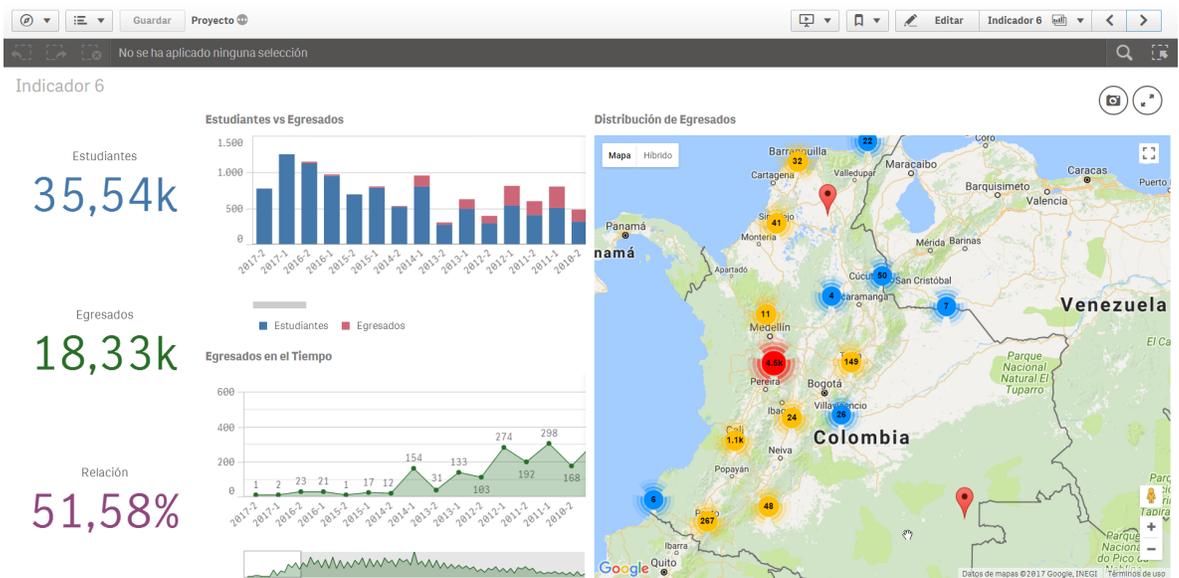


Figura 38. Cuadro de Mando Operativo del Indicador 4

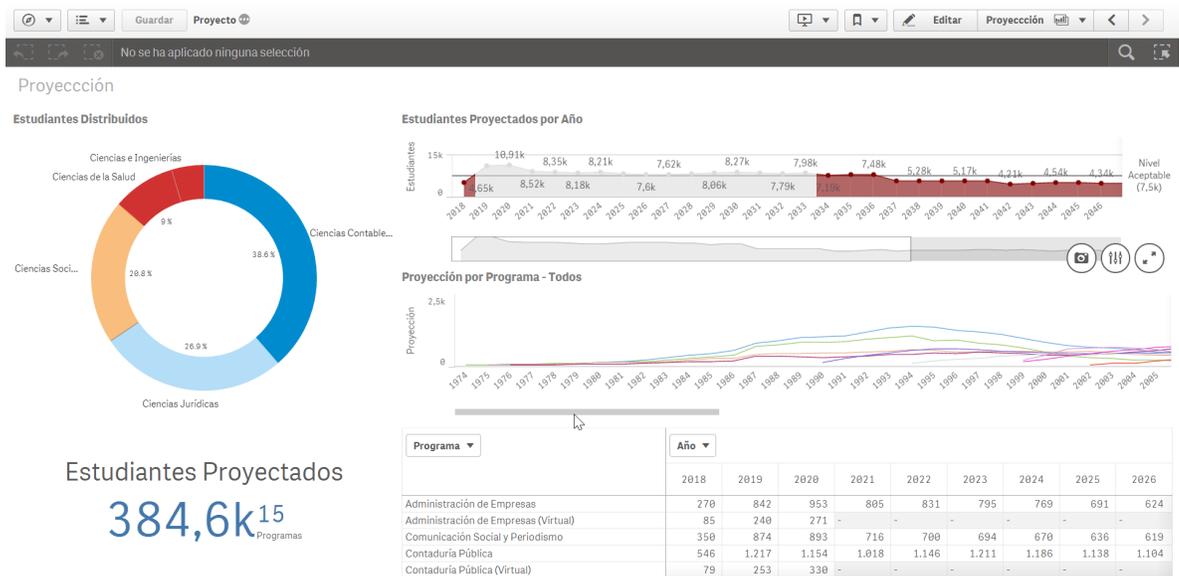


Figura 39. Cuadro de Mando Operativo del Indicador 5

Cabe de anotar que Qlik Sense, con su motor asociativo como elemento central, permite generar el conocimiento que las herramientas de BI tradicionales basadas en consultas simplemente no lo admiten, logra combinar fácilmente todas las fuentes de datos, independientemente del tamaño, en una única visualización; también se debe tener en cuenta que Qlik Sense posee un componente espacial llamado Qlik GeoAnalytics con el cual se puede revelar información geoespacial y exponer las relaciones geográficas que están ocultas en los datos tabulares para tomar mejores decisiones en cuanto a la ubicación geográfica, siendo este caso una mejora o recomendación a realizar para revisiones futuras; permitiendo con ello tomar muchas decisiones organizativas importantes, como la elección de una sede, la asignación de recursos o los plazos de entrega dependen de información relativa a la ubicación geográfica concreta.

Los mapas proporcionan un contexto visual que ayuda a tomar decisiones correctas y bien informadas acerca de la ubicación:

- Analizando las relaciones geoespaciales (y observar lo no relacionado también)
- Escoger una selección de operaciones geoanalíticas
- Combinar fuentes internas y externas en los cálculos

## 6.2 DISCUSIÓN DE RESULTADOS

Aunque inicialmente se pretendía cumplir con la metodología Kimball en su totalidad y al pie de la letra, resultó no ser la adecuada al cien por ciento para el tipo de proyecto que se estaba desarrollando, ya que la misma está implementada y concebida principalmente para un entorno en el cual se produjeran y manipularan datos de manera parcial o específicamente industriales, debido a ello

fueron de gran utilidad los antecedentes que se detallaron y tuvieron en cuenta al momento de visualizar el estado del arte en el cual el proyecto se desenvolvía, en especial los que incluían datos, desarrollos, temas o productos académicos y/o universitarios, dado que llevo al investigador por nuevos rumbos y parajes en la búsqueda de la resolución del problema de investigación que se tenía planteado, de todos modos y aunque la metodología en sí no está para nada errada en su concepción, si se debe tener la claridad y precaución de saber interpretar cuando, como y en qué casos emplearla, y sobre todo conocer en qué tipos de datos es posible aprovecharla a cabalidad.

## 7. CONCLUSIONES

- Las tecnologías BI y SIG forman parte fundamental e integral del proyecto y ofrecen las técnicas, tanto las que se emplean para identificar, extraer y analizar datos, como para tratar la información geoespacial.
- La caracterización de las fuentes de información con las cuales interactúan los sistemas transaccionales; tanto del área financiera como del sistema de registro académico.
- Se consiguió efectuar el análisis de requerimientos en Registro Académico de la Universidad de Manizales, empleando para ello datos académicos, financieros y geográficos, tanto actuales como antiguos, generando así los indicadores que se procedieron a evaluar.
- Se logró producir el diseño del modelo multidimensional para dar solución a los indicadores analizados, por consiguiente, se modelaron los DM basados en diagramas de tipo estrella.
- Elegir la herramienta o artefacto adecuado para ser empleado en el desarrollo de un DW no es sencillo, ya que necesariamente se tienen que respetar y estudiar distintos elementos. Uno de ellos, por ejemplo, se refiere a la comprobación de los procesos que posibiliten la ejecución de tareas como extracción, transformación, carga y exploración de datos. Determinadas herramientas, proporcionan la elaboración de uno solo de estos procedimientos, otras por el contrario aceptan el montaje de varios de ellos y algunas pocas en cambio admiten todos los procesos.
- Se consiguió desarrollar la extracción y transformación de datos de cada uno de los indicadores propuestos, lo cual se realizó con la herramienta Pentaho Data Integration, sin pasar por alto la carga de datos en los DM ya creados en el motor PostgreSQL.
- Se concibieron maneras de exploración de datos por parte de los usuarios finales a partir del DW creado, realizando despliegues de cuadros de mando operativos, esto se realizó empleando la plataforma Qlik Sense Desktop.
- Asimismo, se consiguió revelar que el más trabajoso de todos los procesos fue la transformación de los datos crudos. El inconveniente con éste cambia de una lógica de negocio a otra, obedeciendo este comportamiento a la proporción de información utilizada por la misma y que tan diversas sean las distintas bases de datos de la organización.

- Se evidenció que es conveniente y factible desarrollar e implementar una solución tipo Geo-BI en Registro Académico de la Universidad de Manizales. Demostrando con ello, que se podrá mejorar el procesamiento de datos geoespaciales y académicos para que sirva como apoyo a la toma de decisiones argumentada en dichos datos.

## 8. RECOMENDACIONES

Con el fin de que el DWG sea implantado y puesto en marcha de forma definitiva en la Universidad de Manizales, se hace imperativo efectuar diversas labores, a continuación, se especifican alguna de las más fundamentales:

- Realizar una carga de información real, veraz y actualizada con respecto a las bases de datos relacionales con la que operan en Registro Académico y en el Área Financiera, con el fin de realizar pruebas de veracidad y eficiencia de la solución implementada, ya que se pretende efectuar una implementación en producción del DWG.
- Implementar una solución con una arquitectura de tipo servidor de aplicaciones con la cual sea más sencillo y fácil acceder a los cuadros de mando operativos que sean desarrollados a futuro, y que sea utilizada por el usuario final, el software utilizado, Qlik Sense posee una opción empresarial denominada Qlik Sense Enterprise Server con la finalidad de crear una aplicación estandarizada y visualizaciones usando la arquitectura de Qlik Sense.
- Inicialmente se buscaba emplear la metodología Kimball en su totalidad, sin embargo, se infirió que no es la adecuada al cien por ciento para este proyecto en específico, ya que está concebida para un entorno en el cual se manipularan datos de manera parcial y no contantemente.
- En lo que respecta al análisis de la información necesaria para el DWG, se deben involucrar más indicadores y efectuar los procedimientos análisis y diseño multidimensional con el fin de generar las dimensiones necesarias para los DM.

## BIBLIOGRAFÍA

Abril Frade, D. y Pérez Castillo, J. (2007). Estado Actual De Las Tecnologías De Bodega De Datos Y Olap Aplicadas A Bases De Datos Espaciales. *Ingeniería e Investigación*, 27(1), p.58.

Acuña Salinas, C. (2014). Implementación Del Sistema De Información Ejecutiva Académico Basado En Inteligencia De Negocios: Caso Universidad Peruana Unión (Pregrado). Universidad Peruana Unión.

Anahory, Sam, Murray, Denis. (1997). *Data Warehousing In The Real World: A Practical Guide For Building Decision Support Systems* (1st ed.), Addison Wesley, Boston, MA.

Aracena Aguirre, S., Valdivia Pinto, R., Navarrete Álvarez, M., & Cornejo Yáñez, R. (2014). Desarrollo De Una Solución Geobi Para La Observación Y Análisis De La Información De Censo En Chile. *Interciencia*, 39(10).

Bernabeu, R. (2007). *Data Warehousing Investigación Y Sistematización De Conceptos – HEFESTO Metodología Propia Para La Construcción De Un Data Warehouse* (1st ed., p. 3). Córdoba: Ing. Bernabeu, Ricardo Dario. Recuperado de [http://www.linuxbi.com/files/datawarehouse\\_hefesto.pdf](http://www.linuxbi.com/files/datawarehouse_hefesto.pdf)

Calzada Cantú, L., & Abreu, J. (2009). El Impacto De Las Herramientas De Inteligencia De Negocios En La Toma De Decisiones De Los Ejecutivos. *Daena: International Journal Of Good Conscience*, 4(2), 16-52.

Cornejo, R., Navarrete, M., Valdivia, R., Aroca, P., & Aracena, S. (2014). Desarrollo De Una Base De Datos Integrada De Censo Y Encuesta Mediante El Uso De Elementos De Inteligencia De Negocios Y Sig. *Ingeniare. Revista Chilena De Ingeniería*, 22(2), 205-217. <http://dx.doi.org/10.4067/s0718-33052014000200007>

Cruz Guerrero, R. (2003). *Utilización De La Tecnología Data Warehouse En Instituciones Educativas* (Maestría). Universidad Autónoma Del Estado De Hidalgo.

*GIS and Business Intelligence: The Geographic Advantage*. (2006) (1st ed., p. 1). Redlands. Recuperado de <http://www.esri.com/library/whitepapers/pdfs/gis-and-business-intelligence.pdf>

González Ramírez, C. (2014). *Sistema De Información Geográfica Para La Inteligencia De Negocios, Caso De Estudio En La Superintendencia De Economía Popular Y Solidaria* (Maestría). Universidad San Francisco de Quito.

Guadaña Quiroz, I. (2016). Diseño De Un Datamart Para El Apoyo En La Toma De Decisiones Del Departamento De Admisión De La Universidad Peruana Unión, Filial Tarapoto. Revista De Investigación Ciencia, Tecnología Y Desarrollo, 2(2), 45.

Guevara Lenis, J., & Valencia Arcos, J. (2007). Data Warehouse Para El Análisis Académico De La Escuela Politécnica Nacional (Ingeniería). Escuela Politécnica Nacional.

Harvard Business Review. Hbr's 10 Must Reads. (2011) (p. 25). Boston, Mass.

Inmon, B. (1992). Building The Data Warehouse (1st ed.), Wiley and Sons, New Jersey.

Kenji Fujiwara, D. (2006). Data Warehouse Como Instrumento De Suporte À Avaliação Acadêmica (Licenciatura). Universidade De Brasília.

Kimball, R. (1998). The Data Warehouse Toolkit, Wiley and Sons, New Jersey.

“Junto com a globalização, a informação passou a ter suma importância para todas as pessoas e instituições. A desinformação é nevrálgica para qualquer indivíduo. As relações comerciais, de modo geral, pregam a competitividade de tal modo, que empresas e clientes tornaram-se extremamente ágeis e exigentes. Entretanto o excesso de informação é desinformação. Diante da quantidade de dados, números, estatísticas, nomes, endereços, datas, valores, quantidades e contas disponibilizadas por todos os meios de comunicação de massa, somadas as informações disponíveis nas organizações, instituições e pessoas tendem a ficarem sem informação. Assim, prover informações de maneira organizada, selecionada, suficiente e precisa é certamente uma contribuição essencial às instituições.

Este trabalho apresenta a tecnologia de Data Warehouse como Instrumento de Avaliação Acadêmica, tendo desenvolvido um estudo de caso junto ao CESPE, à Secretaria de Planejamento (SPL) e ao Departamento de Ciências da Computação (CIC) da Universidade de Brasília.” Traducción mía.

La Red Martínez, D., Acosta, J., Cutro, L., Uribe, V., & Rambo, A. (2010). Data Warehouse Y Data Mining Aplicados Al Estudio Del Rendimiento Académico Y De Perfiles De Alumnos. Presentation, Xii Workshop De Investigadores En Ciencias De La Computación.

La Red Martínez, D., Karanik, M., Giovannini, M., & Pinto, N. (2015). Perfiles De Rendimiento Académico: Un Modelo Basado En Minería De Datos. Campus Virtuales, IV (1), 12-30.

Leija Luna, P. (2010). Sistema De Información Geográfica Para La Ayuda De Toma De Decisiones En Políticas Sociales (Maestría). Centro De Investigación Y De Estudios Avanzados Del Instituto Politécnico Nacional Departamento De Computación.

Lozano Mejía, M. (2011). Inteligencia De Negocios: Una Opción Ante La Toma Racional De Decisiones Responsables (Magíster). Universidad Autónoma De Manizales.

Moldes, F. J. (1995). Tecnología De Los Sistemas De Información Geográfica. RA-MA. Madrid.

Nigro, O., Miramont, M. and Velázquez, F. (2003). Crm Con Tecnologías Spatial Data Warehouse Y Data Mining Bajo Un Entorno Distribuido.

Niño Peña, Christiam Alejandro & Cáceres Castellanos, Gustavo (2014). Estado Actual De Temáticas Para El Análisis Espacial En La Toma De Decisiones. En: Ventana Informática No. 31 (jul-dic). Manizales (Colombia): Facultad de Ciencias e Ingeniería, Universidad de Manizales. p. 59-74.

Quinn, K. (2009). How Business Intelligence Should Work The Connection Between Strategic, Analytical, and Operational Initiatives (1st ed., p. 3). Information Builders. Recuperado de [http://www.ebsolute.com/web/doc/InformationManagementWP\\_BI.pdf](http://www.ebsolute.com/web/doc/InformationManagementWP_BI.pdf)

Real Academia Española. (2014). Diccionario De La Lengua Española (23.a ed.). Consultado en <http://dle.rae.es/index.html>

Reinsel, D. & Gantz, J. (2012). The Digital Universe In 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East (p. 1). Framingham: International Data Corporation. Recuperado de <https://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>

Rocha Kuplich, C. (2013). Desenvolvimento de Data Warehouse e Ferramenta OLAP para a Análise da Produção Acadêmica de Pesquisadores Estudo de Caso no PPGC (Licenciatura). Universidad Federal de Río Grande del Sur. "O objetivo deste trabalho é criar uma solução para auxiliar na análise da produção acadêmica do Programa de Pós-Graduação em Computação (PPGC) do Instituto de Informática da UFRGS. A solução toma como base o conceito de modelagem dimensional no desenvolvimento de uma data warehouse para armazenar os dados de produção acadêmica e utiliza uma ferramenta OLAP para a análise dos dados. Esses dados têm origem no Aplicativo Coleta de Dados CAPES, cujo objetivo é coletar informações dos programas de pós-graduação no Brasil. A solução desenvolvida neste trabalho permite visualizar, de forma flexível e dinâmica, o desempenho da produção do PPGC ao longo do tempo, auxiliando nas

decisões gerenciais relativas às pesquisas acadêmicas realizadas no programa.”  
Traducción mía.

Rodríguez Sanz, M. (2010). Análisis Y Diseño De Un Data Mart Para El Seguimiento Académico De Alumnos En Un Entorno Universitario (Pregrado). Universidad Carlos III de Madrid.

Rodríguez Torres, S. y Cáceres Castellanos, G. (2015). Inteligencia De Negocios Espacial Aplicada A La Programación Y Ejecución De Proyectos En El Instituto Colombiano De Bienestar Familiar. Ingenio Magno, 6(1), p.61.

Rubiera Hernández, O., Álvarez Fernández, A., García Barrios, H., Muñoz Castillo, V., López Paz, C., & Wilford Rivera, I. (2014). Tratamiento De Información Georreferenciada En El Sistema De Información Docente De La Educación Superior En Cuba. Ciencias De La Información, 45(3), 3-10.

Salinas, P. (2007). Metodología De La Investigación Científica. Mérida, Venezuela: Universidad de Los Andes.

Sinnexus. ¿Qué es Business Intelligence?. (2016). Sinnexus.com. Recuperado 8 Julio 2016, de [http://www.sinnexus.com/business\\_intelligence/index.aspx](http://www.sinnexus.com/business_intelligence/index.aspx)

Urrutia Sepúlveda, A. Y Medina, r. (2009). Propuesta De Implementación De Un Warehouse Con Extracción De Datos De Una Base De Datos Espacial Y Relacional Para La Gestión Territorial. Gerencia Tecnológica Informática, 8(21), p.44.

Yañez Coto, J., Taboada, J. (2005). Sistemas de Información Medioambiental. Netbioblo. España.

Zambrano Matamala, C; Carvajal Cuello, K; Acuña Leiva, G; Rojas Díaz, D; (2011). Análisis De Rendimiento Académico Estudiantil Usando Data Warehouse Y Redes Neuronales. Ingeniare. Revista Chilena de Ingeniería, 19() 369-381. Recuperado de <http://www.redalyc.org/articulo.oa?id=77221486007>

Zambrano, C., Rojas, D., & Varas, M. (2017). Datawarehouse Con Geolocalización Y Clustering. Nuevas Ideas En Informática Educativa, 7(2011), 102.

## ANEXO A RESUMEN ANALÍTICO

Título del proyecto	PROCESAMIENTO DE DATOS ACADÉMICO-ESPACIALES APLICANDO GEOBI A LAS FUENTES DE REGISTRO ACADÉMICO EN LA UNIVERSIDAD DE MANIZALES
Presidente	BETANCOURT CORREA, Carlos cbc@umanizales.edu.co Magíster, Docente de Planta, Universidad de Manizales
Tipo de documento	Desarrollo Tecnológico
Referencia documento	ACEVEDO NOREÑA, Julián Alberto. Procesamiento de información académico-espacial aplicando GeoBI sobre las fuentes de datos de registro académico, caso de estudio en la Universidad de Manizales. Manizales, 2017, 61 paginas. Universidad de Manizales, Facultad de Ingeniería.
Institución	Maestría en Tecnologías de la Información Geográfica, Facultad de Ingeniería, Universidad de Manizales
Palabras claves	Data Warehouse Geográfico, GeoBI, Universidad de Manizales, SIGUM
Descripción	<p>Para nadie es un secreto que hoy en día, la información se ha convertido en el baluarte y apoyo más indispensable para una compañía que tenga como estrategia de negocio el mantenerse altamente competitiva; cualquiera sea el entorno en que se desenvuelva, es esencial conocer las necesidades y exigencias del medio, por esta razón los datos históricos, sumados a los que producen constantemente las múltiples operaciones de sus sistemas de información se convierten en una fuente inmensa de información que debe ser analizada y procesada.</p> <p>En este sentido, y puesto que los cambios que se producen actualmente en las tecnologías y sistemas de información son demasiado rápidos, dentro del presente proyecto se planteó una metodología clara y organizada para lograr desarrollar un almacén de datos (data warehouse) con un elemento geográfico aplicado a un entorno académico universitario; inicialmente fundamentado en un principio teórico de conocimientos, y familiarización con las principales tecnologías a emplear, obteniendo con ello unas sólidas bases sobre el que construir el desarrollo que constituye el propósito</p>

del proyecto; seguido a ello se describen con detalle cada una de las fases de la metodología empleada para su construcción; y para finalizar los resultados muestran la viabilidad de implementar un Almacén de Datos Geográfico (DWG) para la recopilación y análisis de datos, generados por su propia data histórica, y producidos por los sistemas operacionales tradicionales internos y foráneos de interés para la organización, pretendiendo con todo ello contribuir a suplir y brindar apoyo oportuno a los reportes de indicadores académicos para toma de decisiones de Registro Académico en la Universidad de Manizales.

#### Fuentes

Abril Frade, D. y Pérez Castillo, J. (2007). Estado Actual De Las Tecnologías De Bodega De Datos Y Olap Aplicadas A Bases De Datos Espaciales. *Ingeniería e Investigación*, 27(1), p.58.

Acuña Salinas, C. (2014). Implementación Del Sistema De Información Ejecutiva Académico Basado En Inteligencia De Negocios: Caso Universidad Peruana Unión (Pregrado). Universidad Peruana Unión.

Anahory, Sam, Murray, Denis. (1997). *Data Warehousing In The Real World: A Practical Guide For Building Decision Support Systems* (1st ed.), Addison Wesley, Boston, MA.

Aracena Aguirre, S., Valdivia Pinto, R., Navarrete Álvarez, M., & Cornejo Yáñez, R. (2014). Desarrollo De Una Solución Geobi Para La Observación Y Análisis De La Información De Censo En Chile. *Interciencia*, 39(10).

Bernabeu, R. (2007). *Data Warehousing Investigación Y Sistematización De Conceptos – HEFESTO Metodología Propia Para La Construcción De Un Data Warehouse* (1st ed., p. 3). Córdoba: Ing. Bernabeu, Ricardo Dario. Recuperado de [http://www.linuxbi.com/files/datawarehouse\\_hefesto.pdf](http://www.linuxbi.com/files/datawarehouse_hefesto.pdf)

Calzada Cantú, L., & Abreu, J. (2009). El Impacto De Las Herramientas De Inteligencia De Negocios En La Toma De Decisiones De Los Ejecutivos. *Daena: International Journal Of Good Conscience*, 4(2), 16-52.

Cornejo, R., Navarrete, M., Valdivia, R., Aroca, P., & Aracena, S. (2014). Desarrollo De Una Base De Datos Integrada De Censo Y Encuesta Mediante El Uso De Elementos De Inteligencia De Negocios Y Sig. *Ingeniare. Revista Chilena De Ingeniería*, 22(2), 205-217. <http://dx.doi.org/10.4067/s0718-33052014000200007>

Cruz Guerrero, R. (2003). Utilización De La Tecnología Data Warehouse En Instituciones Educativas (Maestría). Universidad Autónoma Del Estado De Hidalgo.

GIS and Business Intelligence: The Geographic Advantage. (2006) (1st ed., p. 1). Redlands. Recuperado de <http://www.esri.com/library/whitepapers/pdfs/gis-and-business-intelligence.pdf>

González Ramírez, C. (2014). Sistema De Información Geográfica Para La Inteligencia De Negocios, Caso De Estudio En La Superintendencia De Economía Popular Y Solidaria (Maestría). Universidad San Francisco de Quito.

Guadaña Quiroz, I. (2016). Diseño De Un Datamart Para El Apoyo En La Toma De Decisiones Del Departamento De Admisión De La Universidad Peruana Unión, Filial Tarapoto. Revista De Investigación Ciencia, Tecnología Y Desarrollo, 2(2), 45.

Guevara Lenis, J., & Valencia Arcos, J. (2007). Data Warehouse Para El Análisis Académico De La Escuela Politécnica Nacional (Ingeniería). Escuela Politécnica Nacional.

Harvard Business Review. Hbr's 10 Must Reads. (2011) (p. 25). Boston, Mass.

Inmon, B. (1992). Building The Data Warehouse (1st ed.), Wiley and Sons, New Jersey.

Kenji Fujiwara, D. (2006). Data Warehouse Como Instrumento De Suporte À Avaliação Acadêmica (Licenciatura). Universidade De Brasília.

Kimball, R. (1998). The Data Warehouse Toolkit, Wiley and Sons, New Jersey.

La Red Martínez, D., Acosta, J., Cutro, L., Uribe, V., & Rambo, A. (2010). Data Warehouse Y Data Mining Aplicados Al Estudio Del Rendimiento Académico Y De Perfiles De Alumnos. Presentation, Xii Workshop De Investigadores En Ciencias De La Computación.

La Red Martínez, D., Karanik, M., Giovannini, M., & Pinto, N. (2015). Perfiles De Rendimiento Académico: Un Modelo Basado En Minería De Datos. Campus Virtuales, IV (1), 12-30.

Leija Luna, P. (2010). Sistema De Información Geográfica Para La Ayuda De Toma De Decisiones En Políticas Sociales (Maestría). Centro De Investigación Y De Estudios Avanzados Del Instituto Politécnico Nacional Departamento De Computación.

Lozano Mejía, M. (2011). Inteligencia De Negocios: Una Opción Ante La Toma Racional De Decisiones Responsables (Magíster). Universidad Autónoma De Manizales.

Moldes, F. J. (1995). Tecnología De Los Sistemas De Información Geográfica. RA-MA. Madrid.

Nigro, O., Miramont, M. and Velázquez, F. (2003). Crm Con Tecnologías Spatial Data Warehouse Y Data Mining Bajo Un Entorno Distribuido.

Niño Peña, Christiam Alejandro & Cáceres Castellanos, Gustavo (2014). Estado Actual De Temáticas Para El Análisis Espacial En La Toma De Decisiones. En: Ventana Informática No. 31 (jul-dic). Manizales (Colombia): Facultad de Ciencias e Ingeniería, Universidad de Manizales. p. 59-74.

Quinn, K. (2009). How Business Intelligence Should Work The Connection Between Strategic, Analytical, and Operational Initiatives (1st ed., p. 3). Information Builders. Recuperado de [http://www.ebsolute.com/web/doc/InformationManagementWP\\_BI.pdf](http://www.ebsolute.com/web/doc/InformationManagementWP_BI.pdf)

Real Academia Española. (2014). Diccionario De La Lengua Española (23.a ed.). Consultado en <http://dle.rae.es/index.html>

Reinsel, D. & Gantz, J. (2012). The Digital Universe In 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East (p. 1). Framingham: International Data Corporation. Recuperado de <https://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>

Rocha Kuplich, C. (2013). Desenvolvimento de Data Warehouse e Ferramenta OLAP para a Análise da Produção Acadêmica de Pesquisadores Estudo de Caso no PPGC (Licenciatura).

Rodríguez Sanz, M. (2010). Análisis Y Diseño De Un Data Mart Para El Seguimiento Académico De Alumnos En Un Entorno Universitario (Pregrado). Universidad Carlos III de Madrid.

Rodríguez Torres, S. y Cáceres Castellanos, G. (2015). Inteligencia De Negocios Espacial Aplicada A La Programación Y Ejecución De Proyectos En El Instituto Colombiano De Bienestar Familiar. *Ingenio Magno*, 6(1), p.61.

Rubiera Hernández, O., Álvarez Fernández, A., García Barrios, H., Muñoz Castillo, V., López Paz, C., & Wilford Rivera, I. (2014). Tratamiento De Información Georreferenciada En El Sistema De Información Docente De La Educación Superior En Cuba. *Ciencias De La Información*, 45(3), 3-10.

Salinas, P. (2007). *Metodología De La Investigación Científica*. Mérida, Venezuela: Universidad de Los Andes.

Sinnexus. ¿Qué es Business Intelligence?. (2016). Sinnexus.com. Recuperado 8 Julio 2016, de [http://www.sinnexus.com/business\\_intelligence/index.aspx](http://www.sinnexus.com/business_intelligence/index.aspx)

Urrutia Sepúlveda, A. Y Medina, r. (2009). Propuesta De Implementación De Un Warehouse Con Extracción De Datos De Una Base De Datos Espacial Y Relacional Para La Gestión Territorial. *Gerencia Tecnológica Informática*, 8(21), p.44.

Yañez Coto, J., Taboada, J. (2005). *Sistemas de Información Medioambiental*. Netbioblo. España.

Zambrano Matamala, C; Carvajal Cuello, K; Acuña Leiva, G; Rojas Díaz, D; (2011). Análisis De Rendimiento Académico Estudiantil Usando Data Warehouse Y Redes Neuronales. *Ingeniare. Revista Chilena de Ingeniería*, 19() 369-381. Recuperado de <http://www.redalyc.org/articulo.oa?id=77221486007>

Zambrano, C., Rojas, D., & Varas, M. (2017). Datawarehouse Con Geolocalización Y Clustering. *Nuevas Ideas En Informática Educativa*, 7(2011), 102.

## Contenido

En la primera sección del presente trabajo se brinda una breve presentación de la tecnología GeoBI, donde básicamente se describe como el BI que utiliza información geoespacial, luego se describe y plantea el problema, en el cual se logró obtener un claro y detallado análisis y posterior caracterización de los datos pertenecientes a las fuentes de información con las cuales interactúan los sistemas transaccionales y operacionales de Registro Académico, se identifican los principales objetivos del proyecto, además mediante el empleo de la metodología Kimball para la

construcción de un datawarehouse se generó el procedimiento como marco ideal para la producción y total ejecución de los objetivos planteados en la propuesta investigativa. En el marco teórico se hace referencia a todas las fases anteriormente mencionadas se definieron teniendo en cuenta una secuencia lógica de eventos a cumplir para producir el objetivo general de la actual propuesta, estas fases son: Fase 1. Planificación del Proyecto, Fase 2. Definición de los Requerimientos del Negocio, Fase 3. Modelado Dimensional, Fase 4. Diseño Físico, Fase 5. Diseño del Sistema de Extracción, Transformación y Carga y la Fase 6. Especificación de la Arquitectura Técnica; se exponen los resultados del desarrollo del proyecto con sus respectivas conclusiones y recomendaciones.

**Metodología** BI enmarca un conjunto de actividades y procesos que deben llevarse a cabo para ayudar a las entidades a conocer y aprender de sus negocios, dar valor a sus datos y que sean convertidos en información, y posteriormente en conocimiento, en tanto contribuyan y apoyen la toma de decisiones del negocio; Kimball y su metodología se transformaron en el estándar para la creación de DW que proporcionen apoyo a la toma de decisiones organizacionales. En el libro “The Data Warehouse Lifecycle Toolkit” del año 1998 se concibe como ciclo de vida a seguir en el desarrollo de un DW. Siguiendo la metodología Kimball, el proyecto se realizará en 6 fases principales:

- Fase 1. Planificación del Proyecto: En esta fase se definirá la intención primordial del proyecto, las metas y alcance del mismo, los peligros potenciales y la necesidad de información que se desea solventar.
- Fase 2. Definición de los Requerimientos del Negocio: La definición de los requerimientos se dará en gran parte luego de indagar y definir a que personal de la organización se debería entrevistar, para ello se tendrá que dominar la terminología del negocio y el organigrama de la del mismo.
- Fase 3. Modelado Dimensional: La producción en sí de un modelo dimensional se convertirá en un procedimiento dinámico y repetitivo en donde la especificación de los requerimientos del negocio descritos en la fase anterior determinará los datos indispensables que satisfarán los requerimientos analíticos de los usuarios finales.
- Fase 4. Diseño Físico: La fase del diseño físico se centralizará en la elección de la configuración necesaria que sostendrá el diseño lógico ya planteado. Una pieza clave en este proceso es la especificación de estándares dentro el contexto de la base de datos, principalmente la indexación y el

particionamiento a utilizar (en el particionamiento o agregación se debe llevar la información a un nivel atómico).

- Fase 5. Diseño del Sistema de Extracción, Transformación y Carga: Esta etapa del proyecto es tradicionalmente una de las más menospreciadas, sin embargo, es la raíz de la cual se nutre el DW. El sistema ETL se deberá diseñar de manera adecuada, es decir que permita, implementar en los procesos de extracción las funcionalidades necesarias para producir datos que posibiliten realizar su modificación.
- Fase 6. Especificación de la Arquitectura Técnica: El ámbito de implementación de un DW requiere la cohesión de diferentes tecnologías, convirtiéndose en parte fundamental del proyecto y proporcionando a los distintos actores o usuarios del negocio una forma más organizada y, por ende, más simple de acceder e interactuar con el almacén de datos.

**Conclusiones** Las tecnologías BI y SIG forman parte fundamental e integral del proyecto y ofrecen las técnicas, tanto las que se emplean para identificar, extraer y analizar datos, como para tratar la información geoespacial.

La caracterización de las fuentes de información con las cuales interactúan los sistemas transaccionales; tanto del área financiera como del sistema de registro académico.

Se consiguió efectuar el análisis de requerimientos en Registro Académico de la Universidad de Manizales, empleando para ello datos académicos, financieros y geográficos, tanto actuales como antiguos, generando así los indicadores que se procedieron a evaluar.

Se logró producir el diseño del modelo multidimensional para dar solución a los indicadores analizados, por consiguiente, se modelaron los DM basados en diagramas de tipo estrella.

Elegir la herramienta o artefacto adecuado para ser empleado en el desarrollo de un DW no es sencillo, ya que necesariamente se tienen que respetar y estudiar distintos elementos. Uno de ellos, por ejemplo, se refiere a la comprobación de los procesos que posibiliten la ejecución de tareas como extracción, transformación, carga y exploración de datos. Determinadas herramientas, proporcionan la elaboración de uno solo de estos procedimientos, otras por el contrario aceptan el montaje de varios de ellos y algunas pocas en cambio admiten todos los procesos.

Se consiguió desarrollar la extracción y transformación de datos de cada uno de los indicadores propuestos, lo cual se realizó con la herramienta Pentaho Data Integration, sin pasar por alto la carga de datos en los DM ya creados en el motor PostgreSQL.

Se concibieron maneras de exploración de datos por parte de los usuarios finales a partir del DW creado, realizando despliegues de cuadros de mando operativos, esto se realizó empleando la plataforma Qlik Sense Desktop.

Asimismo, se consiguió revelar que el más trabajoso de todos los procesos fue la transformación de los datos crudos. El inconveniente con éste cambia de una lógica de negocio a otra, obedeciendo este comportamiento a la proporción de información utilizada por la misma y que tan diversas sean las distintas bases de datos de la organización.