

UNIVERSIDAD PERUANA UNIÓN

FACULTAD DE INGENIERÍA Y ARQUITECTURA

Escuela profesional de Ingeniería de sistemas



Una Institución Adventista

Diseño de un Datamart para el apoyo en la toma de decisiones del
Departamento de Admisión de la Universidad Peruana Unión, Filial
Tarapoto

Por:

Ivan Guadaña Quiroz

Asesor:

Ing.Jenson Daniel Chambi Aguilar

Morales, octubre del 2016

DECLARACIÓN JURADA DE AUTORIA DEL INFORME DE TESIS


Mtro. Jenson Daniel Chambi Aguilar, de la Facultad de Ingeniería y Arquitectura, Escuela Profesional de Ingeniería de Sistemas, de la Universidad Peruana Unión.

DECLARO:

Que el presente informe de investigación titulado: ***“Diseño de un Datamart para el apoyo en la toma de decisiones del Departamento de Admisión de la Universidad Peruana Unión, Filial Tarapoto”*** constituye la memoria que presenta el **Bachiller Guadaña Quiroz Ivan**; para aspirar al título Profesional de Ingeniero de Sistemas, que ha sido realizada en la Universidad Peruana Unión, bajo mi dirección.

Las opiniones y declaraciones en este informe son de entera responsabilidad del autor, sin comprometer a la institución.

Y estando de acuerdo, firmo la presente constancia en Morales, a los 07 días del mes de octubre del año 2016.



Jenson Daniel Chambi Aguilar

**Diseño de un Datamart para el apoyo en la toma de decisiones del
Departamento de Admisión de la Universidad Peruana Unión, Filial
Tarapoto**

TESIS

Presentada para optar el título profesional de Ingeniero de Sistemas

JURADO CALIFICADOR



.....
Presidente
D.V.

Mg. Danny Levano Rodriguez
Presidente



.....
Secretario

Ing. Joel Pérez Suárez
Secretario



.....
Vocal

Ing. Miguel Ángel Valles Coral
vocal



.....
Asesor

Ing. Jenson Daniel Chambi Aguilar
asesor

Morales, 07 de octubre del año 2016

Dedicatoria

iv

Dedico esta investigación a mis padres por confiar siempre en mí.

A mis hermanos: Sonia, Janeth, David, Talita y Dan, por su constante apoyo.

A mis amigos y compañeros, por su constante motivación.

A ti, amada Lesli, por ser mi gran bendición.

Agradecimientos

v

A Dios por su amor, sus bendiciones, contante protección y por estar conmigo a cada momento.

A mis padres: Victoria Quiros y Félix Guadaña, por dar lo mejor que tenían para lograr este sueño.

Al Mg. Esteban Tocto, por su amistad y asesoría en la planificación de esta investigación.

Al Ing. Jenson Chambi, por ser un gran amigo, excelente consejero y asesor del desarrollo de esta investigación.

A los ingenieros: Joel Pérez y Miguel Valles, por sus consejos y asesorías en clase y en cada trabajo de investigación.

A mi gran amigo Daniel Pachamora, por su apoyo y contribución en el desarrollo de esta investigación.

A la Universidad Peruana Unión, Filial Tarapoto y el Departamento de Admisión, por permitir el desarrollo de esta investigación.

Índice de tablas	viii
Índice de figuras.....	ix
Índice de anexos.....	xi
Glosario de términos	xii
Resumen.....	xiii
Abstract	xiv
Capítulo 1 El problema	1
Capítulo 2 Revisión de la literatura	5
2.1 Conceptos básicos.....	5
2.1.1 Inteligencia de Negocios.....	5
2.1.2 Datos, información y conocimiento.....	6
2.1.3 El ciclo de Inteligencia de Negocios.....	7
2.1.4 Arquitectura de un modelo de IN.....	8
2.1.5 Data Warehouse	9
2.1.6 Data Mart	11
2.1.7 OLAP.....	11
2.1.8 Herramientas OLAP.....	12
2.2 Características de la herramienta de IN	14
2.2.1 Proceso de extracción, transformación y carga (ETL).....	15
2.2.2 Repositorio de datos.....	16
2.2.3 Servidor OLAP	16
2.3 Arquitectura de datos	16
2.3.1 Esquema de estrella.....	16
2.3.2 Esquema copo de nieve.....	17
2.4 Herramientas de IN	17
2.4.1 Sagent Solution Plattform.....	18
2.4.2 Oracle Business Intelligence Solutions.....	18
2.4.3 MicroStrategy	18
2.4.4 Business Objects	18
2.4.5 Ascential Datastage.....	18
2.4.6 Jaspersoft.....	19
2.4.7 Microsoft Business Intelligence.....	19
2.4.8 Pentaho Business Analytics	19
2.5 Metodologías para el desarrollo de soluciones de IN	22
2.5.1 Metodología Ralph Kimbal.....	22
2.5.2 Metodología Inmon.....	22
2.5.3 Metodología HEFESTO.....	23
Capítulo 3 Materiales y métodos.	28
3.1 Tipo de investigación	28
3.2 Diseño de la investigación	28
3.3 Etapas de la investigación.....	30
3.3.1 Levantamiento de información	30
3.3.2 Identificación de requerimientos.....	30
3.3.3 Análisis de los requerimientos	30

	Tabla de contenidos	vi
3.3.4	Identificación de los orígenes de datos	31
3.3.5	Modelado multidimensional	31
3.3.6	Integración de los datos	31
3.3.7	Procesamiento de la información.....	32
3.3.8	Explotar información	32
	Capítulo 4 Resultados y discusión.	33
4.1	Elaboración del modelo de IN para Admisión.....	33
4.1.1	Proceso de selección de la información	33
4.1.2	Proceso de filtración de la data	37
4.1.3	Proceso de extracción y carga.....	39
4.1.4	Proceso de diseño del cubo de dimensiones	43
4.2	Validación de resultados	45
4.3	Estipulación de los indicadores del negocio	45
4.4	Discusiones frente a otras investigaciones.....	55
	Capítulo 5 Conclusiones y recomendaciones.....	75
	Lista de referencias	77
	Anexos	80
	Vita.....	94

Índice de tablas

viii

Tabla 1. Proceso de extracción, transformación y carga.....	15
Tabla 2. Metodologías de IN y sus componentes.	21
Tabla 3. Resumen del proceso ETL.....	26
Tabla 4. Comparativo entre reportes actuales y nuestro modelo de IN.	45

Figura 1. Postulantes a la UPeU-FT.	2
Figura 2. Ciclo de Inteligencia de Negocios.	7
Figura 3. Arquitectura de un modelo de IN.	9
Figura 4. Ambiente de un Data Warehouse	10
Figura 5. Arquitectura de un data warehouse	15
Figura 6. Esquema de estrella	17
Figura 7. Esquema copo de nieve	17
Figura 8. Pentaho Open Suite	20
Figura 9. Arquitectura Ralph Kimbal.	22
Figura 10. Arquitectura Inmon.	23
Figura 11. Diseño de la Investigación.	29
Figura 12. Data histórica de Admisión.	36
Figura 13. Datos filtrados.	37
Figura 14. Transformación de datos con Spoon.	38
Figura 15. Nuevo modelo de base de datos admisión.	39
Figura 16. Test de conexión a la nueva base de datos.	40
Figura 17. ETL Admisión y el tiempo de proceso utilizando Spoon.	42
Figura 18. Modelo de datos estrella.	43
Figura 19. Dimensiones Cubo Admisión.	44
Figura 20. Cantidad de ingresantes por religión y lugar de procedencia.	47
Figura 21. Cantidad de ingresantes según religión y procedencia.	48
Figura 22. Cantidad de estudiantes por carreras y lugar de procedencia.	50
Figura 23. Estudiantes por carreras y sectores de admisión (sectores de promotores).	51
Figura 24. Cantidad de ingresantes según campaña y lugar de procedencia.	52
Figura 25. Cantidad de ingresantes según campaña y modalidad de ingreso.	53
Figura 26. Información estadística en formato .xlsx y .pdf.	54
Figura 27. Cantidad de matrículas y top 10 de puntajes.	55
Figura 28. Porcentaje de Aprobación por Semestre en asignaturas de la carrera	57
Figura 29. Gráfico de barras según campaña y procedencia.	58
Figura 30. Caso de prueba para funcionalidad tipo comunicación.	59
Figura 31. Caso de prueba para funcionalidad tipo extracción.	60
Figura 32. Proceso de extracción utilizando Spoon.	60
Figura 33. Creación de cubo y sus dimensiones.	61
Figura 34. Diseño de Cubo utilizando Schema Workbench.	62
Figura 35. Reporte del pago por servicios por año de la MPC	63
Figura 36. Tareas de nuestra metodología, basada en HEFESTO 2.0.	64
Figura 37. Reportes utilizando Report Designer de Pentaho.	64
Figura 38. Diseño lógico del esquema de producción.	65
Figura 39. Idea de Modelo lógico para el DM Admisión.	66
Figura 40. Proceso de desarrollo del DW	67
Figura 41. Diseño físico del Datamart.	69
Figura 42. Diseño ETL de solución de inteligencia de negocios.	70
Figura 43. Esquema dimensional del DM.	71
Figura 44. Contenido de la tabla hecho_admision.	72

Figura 45. Modelo ETL del Kettle.....	x 73
---------------------------------------	---------

Índice de anexos

xi

Anexos 1. Consulta a la base de datos para extraer el backup Admisión.....	80
Anexos 2. Código SQL para crear la Base de datos en Postgresql.....	81
Anexos 3. Otros reportes.....	88
Anexos 4. Constancia de conformidad del Departamento de Admisión.	92
Anexos 5. Constancia del Centro de Idiomas.	93

BD	Base de datos
BI	Business Intelligence
BSC	Balanced Scorecard
CMI	Cuadro de Mando Integral
DSS	Sistemas de Soportes a la Decisión
DM	Almacén de datos
DW	Colección de datos
EIS	Sistemas de Información Ejecutiva
ETL	Extracción, transformación, carga
HEFESTO	Metodología de desarrollo BI
IN	Inteligencia de Negocios
PDF	Formato de Documento Portátil
PENTAHO	Suite BI basado en software libre
RUP	Rational Unified Process
SPOON	Integración de Pentaho
SQL	Lenguaje de Consulta Estructurado
UPEU-FT	Universidad Peruana Unión, Filial Tarapoto
XLS, XLSX	Extensión de archivo de Microsoft Excel
XML	Extensible Markup Language
TIC	Tecnologías de Información y Comunicación

La presente investigación tiene como propósito diseñar un Datamart (DM) para el apoyo en la toma de decisiones del Departamento de Admisión de la Universidad Peruana Unión, Filial Tarapoto. El cual sigue el proceso normal de una solución de Inteligencia de Negocios (IN): reuniendo información de las distintas fuentes de datos, diseñando un prototipo de extracción, transformación y carga de la data (ETL), para después mostrar la información a través de gráficos estadísticos. Este proceso se basó en la metodología flexible HEFESTO 2.0 que sirvió para bosquejar todas las tareas a realizar, las cuales se organizaron en las siguientes fases: levantamiento de la información, identificación de requerimientos, análisis de los requerimientos, identificación de los orígenes de datos, modelado multidimensional, integración de los datos, procesamiento y explotación de la información. Por tanto este trabajo es de tipo tecnológico y descriptivo, debido al énfasis de obtener conocimiento e interactuar con la información recolectada. Con este diseño, se pudo obtener los siguientes resultados: conocer la cantidad de ingresantes según lugar de procedencia y carrera profesional, cantidad de ingresantes según modalidad y campaña de admisión, cantidad de ingresantes según religión y carrera profesional, cantidad de ingresantes según campaña y sector de admisión. Gracias a esta investigación podemos concluir que: la Suite de Pentaho es de gran utilidad por su sencillez y su tecnología necesaria para realizar soluciones de IN, destacamos la herramienta Spoon que nos permitió realizar con éxito el proceso de ETL; por último nuestro cubo y generación de reportes tuvo éxito por la enorme utilidad de Schema Workbench y Analysis Report.

Palabras claves: IN, Datamart, Data Warehouse, ETL, HEFESTO, Pentaho.

The present research aims to design a Datamart for support in decision making in the office of undergraduate admissions of Peruvian Union University, Site Tarapoto. Which will follow a normal process of a Business Intelligence (BI) solution by gathering information from different data sources, designing an extraction prototype, transformation and loading of data (ETL), and then display the information through statistical graphics. This required to be based on the flexible methodology HEFESTO 2.0, which served as the basis for sketching all our process, which had as phases: Data collection, identification of requirements, analysis of requirements, identification of data sources, multidimensional modeling, data integration, data processing, and information exploitation. All this served to obtain the following results: To know the amount of entrants according to place of origin and professional career, amount of entrants according to admission type and campaign, amount of entrants according to religion and professional career, number of entrants according to campaign and admission sector. Therefore, this research type is technological and descriptive, because of the emphasis of obtaining knowledge and interacting with the information collected. Through this research we can conclude that : Pentaho Suite is useful for its simplicity and its technology solutions necessary for IN, highlight the Spoon tool that allowed us to successfully perform the ETL process ; Finally our cube and reporting succeeded by the enormous usefulness of Schema Workbench and Analysis Report.

Keywords: BI, Datamart, Data Warehouse, ETL, HEFESTO, Pentaho.

Capítulo 1

El problema

El Departamento de Admisión de la Universidad Peruana Unión, Filial Tarapoto (UPeU-FT) cumple con dos roles importante dentro de la institución: primero marketing y publicidad y segundo admisión propiamente dicho, por consiguiente tiene un objetivo fusionado de ambos roles, cuyas tareas son: captar la mayor cantidad de alumnos postulantes sostenedores de obra, verificar que las vacantes se hayan distribuido correctamente, e identificar a las carreras que le faltan vacantes, proporcionando ayuda en la documentación, consultas y otros. (Tarrillo, comunicación personal, 11 de mayo, 2015)

Cada campaña de admisión lleva consigo la distribución de los recursos para lograr una mayor cantidad de alumnos postulantes sostenedores de obra. Las campañas de admisión siguen ciertos parámetros para poder lograr las metas de postulantes establecidas. Cabe rescatar que al iniciar cada proceso se debe conocer información estadística detallada desde el promotor hasta la zona de impacto. Al final de cada campaña se obtiene información de suma importancia que sirve como input para próximas campañas (Tarrillo, comunicación personal, 11 de mayo, 2015). En la figura 1 propuesta por Tarrillo, se observa la cantidad general de nuevos postulantes a esta institución de los últimos cinco años, dando a mostrar que de cierta forma ha existido una baja en el número de postulantes.



Figura 1. Postulantes a la UPeU-FT.

En tal sentido, la dificultad para tener información explícita de ayuda para la toma de decisiones se da por las siguientes causas: Inadecuada definición y estructura de los reportes, complejos procesos para su generación, largos periodos para la generación y/o actualización de información estadística y problemas derivados de la deficiente confianza en la información generada; que como consecuencia genera: incertidumbre en la distribución de los recursos en las campañas de admisión, se desconoce los lugares que tienen la mayor cantidad de postulantes sostenedores de obra, dificultad en el uso de las Tecnologías de Información y Comunicación (TIC), incertidumbre sobre las zonas en que se desempeñan los promotores. Los principales afectados son: el área administrativa de la universidad, el jefe del departamento de admisión y el personal que labora en el área. Esto ocurre al planificar la distribución de los recursos, con una frecuencia trimestral o semestralmente, de acuerdo a las campañas programadas. Trayendo una magnitud de impacto alta.

Con este trabajo se comprobó la importancia del DM, que no ha sido muy aplicado en instituciones universitarias. De esta manera el área de admisión pudo obtener

información más detallada, siendo de gran beneficio para la toma de decisiones. Como guía para alcanzar los objetivos de estudio, se fusionó la metodología HEFESTO 2.0 y el ciclo normal de una solución de IN, dando prioridad a la metodología HEFESTO para tener un panorama claro al identificar y cubrir los requerimientos. Como resultado de esta investigación, los beneficiarios son la administración de la universidad, ya que les permite tomar decisiones basadas en hechos (representados por gráficos), la dirección de admisión definiendo el recurso humano según la ubicación geográfica, tipo de instituciones educativas y zonas de alto impacto, basadas en una data histórica. Además los usuarios directos serán el personal que labora en el área.

Gracias a este trabajo confirmo mi fe en Dios y mis creencias en la Santa Biblia, ya que para desarrollar una investigación se necesita razonar y descubrir nuevos conceptos o herramientas para resolver problemas, ese conocimiento solo proviene de Dios. “El principio de la sabiduría es el temor de Jehová” (Proverbios 1:7 versión Reina Valera, 1960).

Afirmo también, que una de mis convicciones es que Dios creó este mundo, así como también a nosotros. No hay duda de eso. La Biblia lo confirma así: “Los cielos cuentan la gloria de Dios, y el firmamento anuncia la obra de sus manos” (Salmos 19:1).

Este trabajo, va de acorde a mis creencias y principios, y me motiva a seguir investigando. Como dice la Biblia: “Las cosas secretas pertenecen a Dios; más las reveladas son para nosotros, a fin de que cumplir todas las palabras de esta ley” (Deuteronomio 29:29 versión Reina Valera, 1960). Así que siempre habrá misterios y nuevas cosas por descubrir.

La presente investigación tuvo como propósito diseñar un modelo de Inteligencia de Negocios, útil y necesaria en la toma de decisiones para el departamento de Admisión de la Universidad Peruana Unión, Filial Tarapoto. Para lo cual se investigó las herramientas tecnológicas de extracción, transformación y carga, para después proponer un modelo estándar de ETL. Luego se diseñó el modelo multidimensional de los indicadores, para su posterior explotación de la información. Siendo este modelo muy fácil de entender para los usuarios. En seguida se dio lugar a reportes específicos, donde el cliente, podrá analizarlos de distintas maneras según su perspectiva.

El capítulo 1 de esta investigación presenta la introducción a toda la tesis, compuesta por los objetivos, justificación y el planteamiento de la investigación. El capítulo 2 presenta toda la revisión científica y tecnológica utilizada para esta investigación, especificadas en: conceptos de IN, herramientas de BI y las características de cada una de ellas y las metodologías más comunes en el desarrollo de proyectos de IN. En el capítulo 3 mostramos la metodología que utilizamos en la investigación, destacando las tareas en cada fase del proyecto de IN, asimismo describimos algunos materiales utilizados en el proceso de construcción del modelo de IN. El capítulo 4 contiene toda la información de la herramienta tecnológica, su construcción y las pruebas realizadas con este modelo, además hacemos un análisis de nuestro modelo frente a otras investigaciones realizadas. El capítulo 5 presenta las conclusiones a las que se ha llegado con este modelo de IN y algunas recomendaciones que se deben tener en cuenta.

Capítulo 2

Revisión de la literatura

Este capítulo, presenta los conceptos primordiales de la tecnología para la IN, los componentes y su arquitectura de la misma. Del mismo modo se describe la parte de las partes o etapas del proyecto de IN. También tocamos el tema de los productos más utilizados en el mercado tecnológico para hacer IN, entre los que se destaca la Suite de Pentaho como software libre.

2.1 Conceptos básicos

2.1.1 Inteligencia de Negocios

Cárdenas (2011) define a la IN como: conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar, y transformar datos de los sistemas transaccionales e información desestructurada en información estructurada, para su explotación directa o para su análisis y conversión en conocimiento, dando así soporte a la toma de decisiones sobre el negocio. (p.15)

Del mismo modo “IN es el conjunto de tecnologías que reúne arquitecturas para almacenar datos, técnicas para analizar información, aplicando alguna metodología, con el objetivo de apoyar a la toma de decisiones” (Valladares, comunicación personal, 17 de mayo, 2015).

Considerando su concepto multifacético, la inteligencia de negocios actúa como un factor estratégico para una empresa u organización, generando una potencial ventaja competitiva, proporcionando información clasificada para responder a los problemas de

negocio: entradas a nuevos mercados, control financiero, análisis de perfiles de clientes, rentabilidad de un producto o servicio, etc. (Cárdenas, 2011, p.16)

Los componentes más importantes de orígenes de datos en la IN que se consideran en la actualidad son: Datamart y Data Warehouse (DW). Cruz (2003) afirma que un modelo de IN permite:

Observar ¿qué está ocurriendo?, comprender ¿por qué ocurre?, predecir ¿qué ocurrirá?, colaborar ¿qué debería hacer el equipo?, decidir ¿Qué camino se debe seguir? (p.13)

2.1.2 Datos, información y conocimiento

“Los datos, son la mínima unidad semántica, y se consideran como elementos primarios de información que por sí solos son irrelevantes. Es como una imagen en crudo, que no te emite informe. Ejemplo: un número de celular, un nombre, otros” (Gelinas, Sutton, & Federowicz, 2008, p.223).

La información se puede definir como un conjunto de datos procesado, con un significado, presentada de manera útil para la toma de decisiones, y es de gran ayuda para disminuir la incertidumbre. Algunas cualidades que contiene la información son: proporciona una salida útil de ayuda, describe cualidades, y su contenido es eficaz, oportuno y comprensible. (Gelinas et al., 2008, p.221)

Por otro lado, el conocimiento es una mezcla de información, experiencia y valores que sirve como marco para la incorporación de nuevas experiencias e información, y es útil para la acción. Su origen y aplicación se da en la mente de los pensadores. Para que la información se convierta en conocimiento es necesario: comparar

con otros elementos, buscar conexiones y conversar con otros portadores de conocimiento. (Cárdenas, 2011, p.17)

2.1.3 El ciclo de Inteligencia de Negocios

Bustos & Mosquera (2013) explican que IN es más que una actitud de negocios o una tecnología; de hecho es un marco de referencia para la gestión del rendimiento empresarial, a través de un ciclo mediante el cual se definen objetivos, se observan progresos, se obtiene conocimiento, se toma decisiones, se miden éxitos y nuevamente inicial el ciclo. (p.14)

A través de IN, se puede normalmente analizar datos provenientes de muchas fuentes de data histórica. El análisis conduce a ideas, algunas de ellas pequeñas, y otras grandes. Estas ideas indican maneras de mejorar el negocio cuando se actúa sobre ellas; estas ideas pueden ser medidas para ver si funcionan una y otra vez. A esto se llama ciclo de IN o BI (Cárdenas, 2011, p.12). A continuación en la Figura 2 podemos ver el ciclo tiene todo proceso de IN, iniciando siempre en un análisis para concluir con determinar resultados a través de medidas o indicadores.

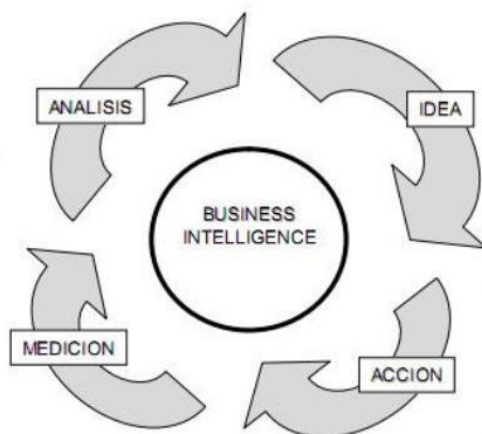


Figura 2. Ciclo de Inteligencia de Negocios.

2.1.4 Arquitectura de un modelo de IN

Toda solución de IN inicia por los sistemas de origen de una organización (archivos de texto, Excel, base de datos, etc.) sobre los cuales se debe aplicar una transformación estructural para mejorar su proceso analítico.

Todo ese proceso inicial se le conoce como fase de extracción, transformación y carga (ETL) de datos. Esta etapa con frecuencia se apoya en un almacén operacional intermedio de los datos (ODS), que actúa como conector entre los sistemas fuente y los sistemas destino (generalmente un DW), y cuyo fin es evitar la saturación de los servidores funcionales de la organización. (Oktavia, 2014, p.19)

Con este proceso se puede obtener información unificada, depurada y consolidada, para luego almacenarlo en un DM o DW general, que servirá de base para la construcción de distintos DM departamentales. Estos DM poseen una estructura óptima para el análisis de los datos de esa área de la empresa, ya sea mediante base de datos transaccionales (OLTP) o mediante base de datos analíticas (OLAP). (Ojeda, 2008, p17)

Zambrano, Rojas, Carvajal, & Acuña (2011) presentan en la figura 3 los diferentes datos almacenados en el DW o en cada DM se explotan utilizando herramientas comerciales de análisis, reporting, etc. Estas herramientas también se consideran en la construcción de un modelo de IN, como los sistemas de soporte a la decisión (DSS), los sistemas de información ejecutiva (EIS) y los cuadros de mando integral (CMI) o Balanced Scorecard (BSC). (p.16)

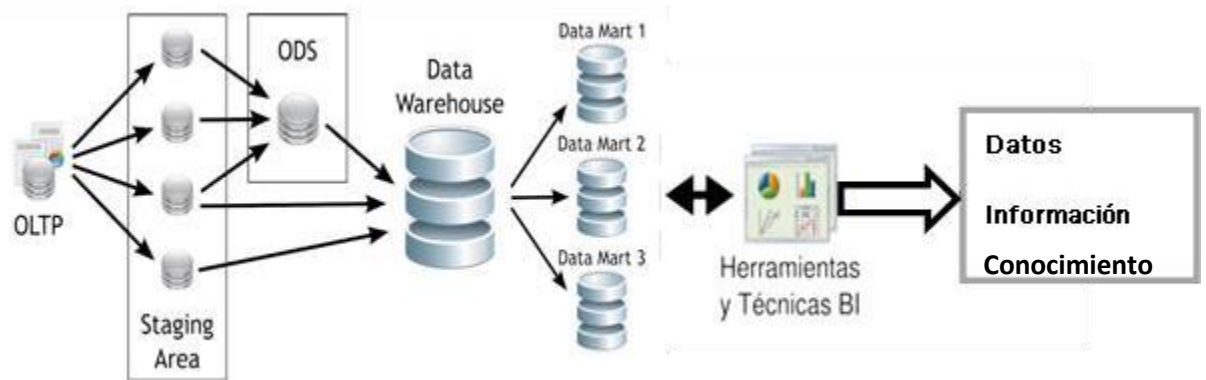


Figura 3. Arquitectura de un modelo de IN.

2.1.5 Data Warehouse

Bustos & Mosquera (2013) mencionan. “Un DW es una base de datos corporativa, integrada, estable, no volátil, de manera tal que al procesarla permita su análisis a gran velocidad de respuesta; dicho de otra forma almacena una gran cantidad de información histórica del negocio” (p. 15).

“Está compuesto de elementos básicos, entre los cuales están las dimensiones de análisis, los indicadores o variables de gestión y los hechos que representan los datos reales” (Zambrano et al., 2011, p.19).

En la Figura 4 se evidencia el ambiente del DW, formado por diversos elementos que interactúan entre sí y que cumplen una función específica dentro del sistema:

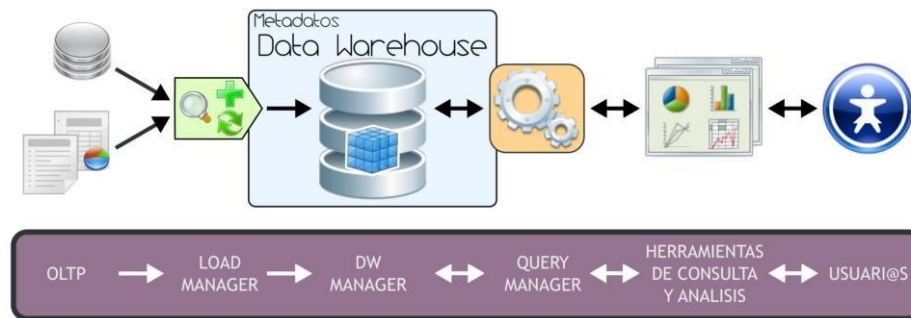


Figura 4. Ambiente de un Data Warehouse.

El término DW fue recalado por primera vez por Bill Inmon, y se traduce literalmente como almacén de datos. Sin embargo (Cárdenas, 2011) aclara este tema y menciona que es mucho más que esta definición, sino también por ser:

- Integrado: los datos integrados en el DW deben ser consistentes, para adecuarse a las distintas necesidades de los usuarios finales.
- Temático: los datos deben estar organizados por temas, para facilitar su acceso y entendimiento por parte de los usuarios finales. Ejemplo: los datos sobre clientes pueden estar en una única tabla del DW.
- Histórico: el tiempo es parte implícita de la información del DW. Esta información muchas veces servirá para realizar análisis de tendencias. En todo caso el tiempo permite comparaciones.
- No volátil: la información del almacén de un DW permite ser leída, pero no modificada. Por lo tanto la información es permanente.
- Contiene metadatos, es decir, datos sobre datos. Estos permiten saber la procedencia de la información, su periodicidad, fiabilidad, forma de cálculo, etc. Estos

permiten simplificar y automatizar la información de sistemas operacionales a sistemas informacionales.

Los metadatos contienen al menos: la estructura de los datos, los algoritmos usados para la esquematización y el mapeo desde el ambiente operacional al DW. (p.63)

2.1.6 Data Mart

Cárdenas (2011) se refiere a un DM como “una base de datos departamental, que almacena datos clasificados de un área específica de negocio. Proporciona la estructura óptima de datos para analizar la información al detalle, desde las diferentes perspectivas que afecten a los procesos de un departamento” (p.65).

Por consiguiente, para crear el DM de un área funcional de una organización es necesario determinar la estructura óptima para poder analizar la información, esta estructura puede ser realizada sobre una base de datos Online Transaction Processing (OLTP), como el propio DW, o sobre una base de datos OnLine Analytical Processing (OLAP). Todo esto dependerá de los datos, requisitos y características de cada departamento. (Bustamante, Galvis, & Gómez, 2013, p.34)

2.1.7 OLAP

Es una de las dos tecnologías de IN, que presenta un modelo de datos, sencillo e intuitivo para usuarios poco relacionados con el análisis. “Este modelo, le permite al usuario ver los datos a través de filtros o dimensiones. Dimensión viene a ser una vista de datos consistente y resumida” (Sánchez, 2009, p.49). Así mismo este autor considera que una de las ventajas del análisis multidimensional es ser rápido, ya que procesa una gran

cantidad de datos, los cuales pueden ser proveídos por sistemas operacionales o bases de datos operacionales (OLTP).

2.1.8 Herramientas OLAP

Dragoljub (citado por Sánchez, 2009) considera que la tecnología OLAP está distribuida en varios tipos de herramientas, las mismas que permiten la explotación de la información:

OLAP relacional (ROLAP): esta arquitectura sobresale en la posibilidad de trabajar sin límite en la cantidad de datos, pero su deficiencia es la velocidad de respuesta.

OLAP: maneja una cache para las consultas, pero sus datos físicamente están en una base de datos relacional.

MOLAP: Recibe consultas dimensionales (lenguaje MDX) y devuelve los datos en forma de un cubo, sólo que este cubo no es algo físico sino un conjunto de metadatos que definen como se han de “mapear” estas consultas.

HOLAP (instrumentos híbridos): este tipo de herramientas me permite usar el poder de las bases de datos relacionales (ROLAP), donde se encuentra el cubo, pero al momento de generar la consulta MDX, el reporte genera datos los cuales se almacenan en una cache temporal, haciendo más rápida la respuesta.

MDX (multidimensional expressions): lenguaje multidimensional creado por Microsoft. Es interpretado por el motor OLAP, este traduce las instrucciones al lenguaje SQL, que es entendido por cualquier base de datos relacional. Es decir para manejar

grandes volúmenes de datos y que las consultas sean rápidas y dinámicas se necesita de este lenguaje MDX.

Otros conceptos: Heidelberg (como se citó en Sánchez, 2009) piensa que ante un análisis o diseño dimensional, es necesario recordar algunos conceptos, los que se resumen a continuación:

Atributos: permiten el relacionamiento de las dimensiones hacia el dato numérico. Crea un tipo de mapeo al momento de las consultas.

Hecho: al modelar el cubo, se considera cual sería el valor numérico, con el cual navegaré por todo el cubo. Los hechos vienen a ser datos individuales que detallan las condiciones del área en estudio.

Jerarquías: es la relación que existe entre dos o más atributos, permitiendo navegar en los reportes.

Métricas: están basados en datos de los hechos, determinan el estado del negocio. Existen dos tipos de métricas:

Métrica de función de grupo. Ejemplo: total_cursos, se obtiene mediante consulta SQL a las tablas operacionales. Pero cuando se quiere manipular el cubo, se utiliza funciones de grupo AVG, SUM, COUNT, etc.

Métrica por fórmula. Ejemplo: total_ingresos = (cantidad vendida*(costo_venta – costo_compra)).

2.2 Características de la herramienta de IN

En toda solución de IN se inicia por los sistemas de origen de una organización (ficheros de texto, Excel, base de datos, etc.) sobre los cuales se debe aplicar una transformación estructural para mejorar su proceso analítico.

Todo ese proceso inicial se le conoce como fase de extracción, transformación y carga (ETL) de datos. Esta etapa con frecuencia se apoya en un almacén operacional intermedio de los datos (ODS), que actúa como conector entre los sistemas fuente y los sistemas destino (generalmente un DW), y cuyo fin es evitar la saturación de los servidores funcionales de la organización (Oktavia, 2014, p.5).

De este proceso se obtiene información unificada, depurada y consolidada, para luego almacenarlo en un DM o DW general, que servirá de base para la construcción de distintos DM departamentales. Estos DM poseen una estructura óptima para el análisis de los datos de esa área de la empresa, ya sea mediante base de datos transaccionales (OLTP) o mediante base de datos analíticas (OLAP).

Los diferentes datos almacenados en el DW o en cada DM se explotan utilizando herramientas comerciales de análisis, reporting, etc. Estas herramientas también se consideran en la construcción de un modelo de IN, como los sistemas de soporte a la decisión (DSS), los sistemas de información ejecutiva (EIS) y los cuadros de mando integral (CMI) o Balanced Scorecard (BSC). Gracias a (Zambrano et al., 2011,p.5) todo esto se evidencia en la figura 5.

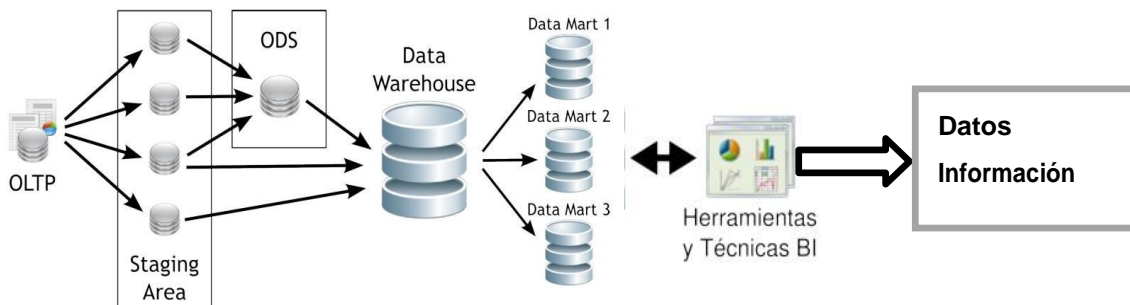


Figura 5. Arquitectura de un data warehouse.

2.2.1 Proceso de extracción, transformación y carga (ETL)

En este paso se procede a llenar el modelo con datos de calidad, utilizando el proceso extracción, transformación y carga (ETL) de tal manera que deposite los datos en el almacén de datos. Es probable que esta tarea a veces contenga una lógica un poco compleja, sin embargo, dependerá del correcto análisis, conocimientos y herramientas que se empleen con el fin de facilitar el trabajo (Bustamante, Galvis, Gonzáles, García, & Benavides, 2011).

Tabla 1. Proceso de extracción, transformación y carga.

<i>Componente</i>	<i>Entrada</i>	<i>Proceso</i>	<i>Salida</i>
Extracción	Base de datos, hojas de cálculo, archivos de texto, otros.	Selección	Datos crudos (cargados a memoria)
Transformación	Datos crudos (cargados a memoria)	Limpieza, transformación, personalización, realización de cálculos, aplicación de funciones	Datos formateados, estructurados y resumidos de acuerdo a las necesidades
Carga	Datos formateados, estructurados y resumidos de acuerdo a las necesidades (aún en memoria)	Inserción	Datos formateados, estructurados y resumidos con persistencia en el DW.

Nota: Según Bustamante et al. (2011).

Es recomendable primero cargar los datos de las dimensiones, y luego de la tabla hechos, siempre considerando que la tabla hechos contendrá el nivel de granularidad.

En el esquema estrella, varias tablas de dimensiones comparten una tabla de hechos.

En el esquema copo de nieve, las tablas dimensiones generan nuevas dimensiones.

Después de cargar el DW en su totalidad, se debe establecer políticas de actualización de datos (refrescar datos).

2.2.2 Repositorio de datos

“Contiene la definición de los objetos de análisis tales como: métricas, reportes, entre otros” (Rojas, 2014, p.40).

2.2.3 Servidor OLAP

El motor de base de datos relacional (ROLAP) interpreta las consultas hechos por el usuario (análisis multidimensional) en lenguaje MDX y las transforma en lenguaje SQL, entonces sucede el evento de consulta al data warehouse y devuelve información representada en gráficos estadísticos. (Rojas, 2014, p.42)

2.3 Arquitectura de datos

2.3.1 Esquema de estrella

En la figura 6 Bernabeu (2010) presenta el esquema estrella usado en el modelado de datos en IN, compuesto por un elemento o tabla central llamada tabla de hechos, la cual está relacionada con otras tablas (dimensiones) solo a nivel de descripción.

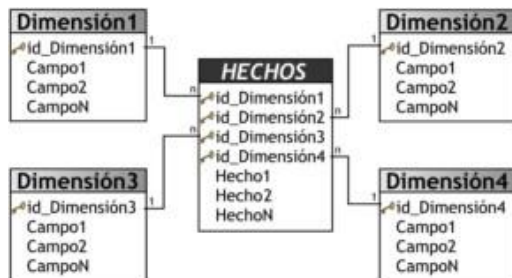


Figura 6. Esquema de estrella.

2.3.2 Esquema copo de nieve

La figura 7, muestra la forma de modelar este tipo de esquema, su estructura está por niveles (dimensiones) que puede contener más de una tabla, esto ocasiona que las dimensiones adquieran o se distribuyan en jerarquías. Existe una tabla de hechos central que está relacionada con una o más dimensiones que a su vez pueden estar relaciones con una o más dimensiones, esta representación es tomada de (Bernabeu, 2010).

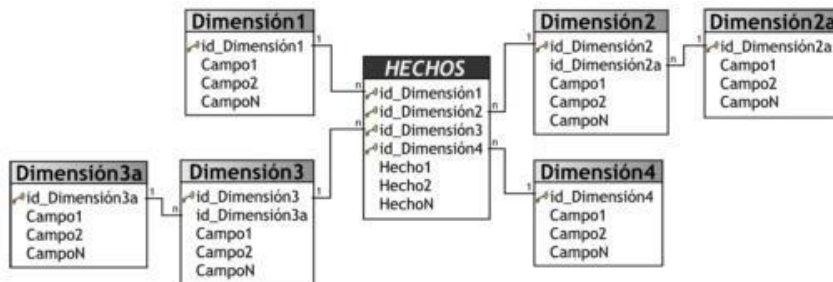


Figura 7. Esquema copo de nieve.

2.4 Herramientas de IN

Esta sección describe características de las herramientas de IN. Dicha información fue desarrollada y descrita por varios expertos:

2.4.1 Sagent Solution Plattform

“Sistema integrado que efectúa el proceso ETL, también permite definir el formato al exportar la información para la toma de decisiones de la organización” (Valdiviezo, Herrera, & Jáuregui, 2007, p.32).

2.4.2 Oracle Business Intelligence Solutions

“Oracle contiene un paquete completo para desarrollar inteligencia de negocios, siempre dependiendo de su base de datos Oracle” (Valdiviezo et al., 2007, p.33).

2.4.3 MicroStrategy

Esta herramienta permite diseñar y generar reportes a partir de la explotación de los cubos del DW. Es importante considerar que adicionalmente permite diseñar, documentos web basados en reportes elaborados. Otra gran ventaja de este software es la administración, desarrollo e implementación de soluciones de IN. (Gutiérrez, 2012, p.58)

2.4.4 Business Objects

Business Objects, admite a los usuarios el acceso, análisis y distribución de la información. Se caracteriza por ser una herramienta segura, escalable, extensible y sencilla de utilizar. Es muy avanzada, puesto que incluye consultas, generación de análisis e informes, broadcasting e importantes herramientas de administración. (Valdiviezo et al., 2007, p.35)

2.4.5 Ascential Datastage

Bringa una base para administrar gran cantidad, variedad y velocidad de datos. Ascential Enterprise Integration Suite, y sus frameworks que incluye, mejora la calidad de los datos siendo de beneficio a cada departamento u organización. Con su amplia

plataforma integrada, ayuda a las empresas a optimizar sus recursos, su análisis de estrategias, intercambio de información, comercio electrónico, y otros. (Gutiérrez, 2012, p.63)

2.4.6 Jaspersoft

Es una plataforma de IN utilizada para generar soluciones en pequeñas y medianas empresas. Cuenta con herramientas que permiten hacer el desarrollo de un modelo completa de IN, iniciando en la extracción de la data hasta su depósito en el DW, explotar la información, y finalmente ser visualizada por los usuarios del negocio (Gutiérrez, 2012, p.63).

2.4.7 Microsoft Business Intelligence

Torriente, Sentí, Hernández, & Ortega (2011) dicen que la plataforma Business Intelligence de Microsoft incluye poderosos sistemas de gestión de bases de datos, servicios de integración de SQL Server, los servicios de análisis del sistema SQL Server, servicios de comunicación de SQL Server y la posibilidad de la investigación de los datos del negocio de SQL Server. Aparte de que integre a la plataforma de desarrollo de Microsoft Visual Studio 2005. (p.8)

2.4.8 Pentaho Business Analytics

Fuentes & Valdivia (2010) se refieren a Pentaho como la plataforma que ofrece los servicios necesarios para realizar una soluciones de IN. Sus soluciones están centrada en procesos, brindando básicos que incluyen el registro, autenticación auditoría, flujo de trabajo, servicios web, y otros. Por si fuera poco Pentaho también contiene un motor de solución que integra la presentación de informes, análisis, dashboards y componentes de

minería de datos convirtiéndose en una plataforma completa y sofisticada. Se habla de completa y sofisticada puesto que contiene Kettle para el proceso de ETL, Mondrian para el análisis OLAP, Jpivot para los reportes y la community framework dashboard (CDF) para los dashboards. (p.7)

La figura 8, es tomada de Masaquiza (2011) para representar a Pentaho como una de las suites pioneras en cuanto a IN:

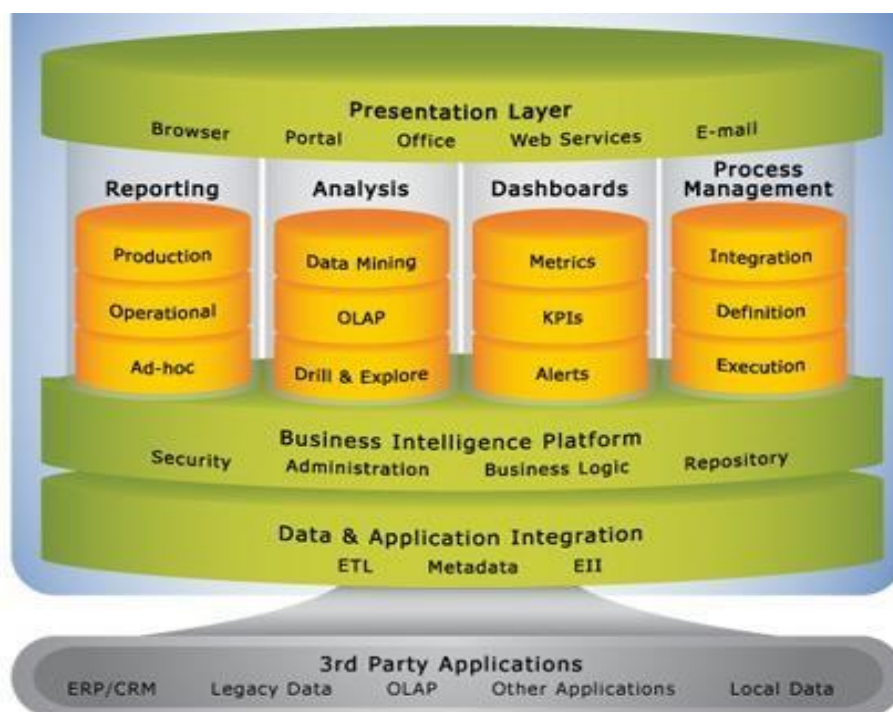


Figura 8. Pentaho Open Suite.

Se estima que la Suite o kit de Pentaho es de código abierto, mantenido por una comunidad de desarrolladores y lo podemos encontrar en la siguiente dirección electrónica: <https://sourceforge.net/projects/pentaho/files/?source=navbar>. Sin embargo en www.pentaho.com, podemos encontrar la misma suite licenciada y disponible en un solo instalador.

La suite o kit de pentaho está conformado por varias herramientas, las cuales se especifican a continuación. Pentaho (citado por Sánchez, 2009)

Spoon (antes Kettle): permite hacer procesos de ETL del BI, ahora esta herramienta es donde hemos hecho más pruebas, hasta poder comprender la lógica de extracción y transformación, que son los aspectos que más lo hemos visto complicado en esta tesis.

El spoon ofrece realizar en si dos tareas: Transformaciones y Job. Esta herramienta tiene distintas funcionalidades para hacer la carga y transformación de datos, en cuanto a la transformación se sabe que de la fuente al destino hay ciertas diferencias. Se observa en el sqlserver el tipo de dato datetime y se requiere cargar al gestor de base de datos Oracle que maneja el tipo de dato date (fecha) donde el formato es DD/MM/YYYY.

El job sincroniza la carga, es decir; se programa la hora, el día o la fecha que se desea hacer la transformación. Es software libre.

Schema Worbench: lo que hace esta herramienta la suite de pentaho se le llama diseñar esquemas o cubos capaces de brindar información relacionada al hecho. Esta herramienta permite genera un archivo XML capaz de ser exportado al Servidor Bussines Analytics de Pentaho. Es software libre.

Business Analytics: Es la Suite completa de Pentaho que provee todas las herramientas antes mencionadas, inclusive un servidor. Esta provee de hacer un análisis en tiempo real de todos los cubos ejecutados y presentados en reportes dinámicos

(gráficos), inclusive tiene la opción de generar un tipo de cubo, jalando datos del modelo multidimensional para luego ser explotado. Esta suite completa contiene licencia o también ofrece la Suite de prueba por 30 días.

2.5 Metodologías para el desarrollo de soluciones de IN

Existe una variedad de técnicas que se utilizan para un modelo de IN, sin embargo, “no existen consensos sólidos respecto de los modelos de procesos de desarrollo de software que puedan ser utilizados para construir este tipo de soluciones” (Bustamante et al., 2011, p.4).

En la siguiente tabla se muestra un comparativo sintetizado de las metodologías para el desarrollo de soluciones de IN donde se podrán visualizar el ciclo de vida, nivel de detalle en las actividades a realizar, y el alcance que tendrá el modelo de IN. Por su parte, en la misma tabla se presenta la metodología HEFESTO versión 2.0 que se propone utilizar, así mismo se resalta en negrita las metodologías más utilizadas para la elaboración de un DM o DW.

Tabla 2. Metodologías de IN y sus componentes.

<i>Metodologías y Técnicas / características</i>	<i>Modelo de Ciclo de vida</i>	<i>Nivel de Detalle</i>	<i>Alcance</i>
Ciclo de vida de Ralph Kimbal. Método de desarrollo de sistemas dinámicos para bodegas de datos	Cascada Iterativo	Alto Bajo	Bodega de datos
Metodología para la construcción de un DW (Hefesto)	Cascada	Medio	Bodega de datos
Metodología Inmon	Cascada	Medio	Bodega de datos
Metodología DWEP	Iterativo	Alto	Bodega de datos
MBD	Ágil	Bajo	Bodega de datos
SEMMA	Cascada	Medio	Minería de datos
CRISP-DM	Cascada	Medio	Minería de datos
BI RoadMap	Cascada	Bajo	Solución de BI
BIEP	Iterativa	Alto	Solución de BI
Metodología y modelo de ciclo de vida de un modelode IN (Gartner)	Iterativo	Bajo	Solución de BI

Nota: Tomada de Bustamante et al. (2011).

2.5.1 Metodología Ralph Kimbal

La Metodología Ralph Kimbal, proporciona un marco general que une distintas actividades de un DW e IN. En esta metodología un DW se considera como una copia de los datos transaccionales para realizar la consulta y análisis de los mismos. También conocida como Modelo Dimensional, constituyendo tablas y relaciones con el fin de mejorar la toma de decisiones, basadas en sentencias SQL realizadas en una base de datos relacional. (Bustos & Mosquera, 2013, p.39)

La Figura 9 fue tomada de Moss & Atre (2003) para mostrar la arquitectura de Kimball, y por lo que se observa se le denomina también modelo estrella:

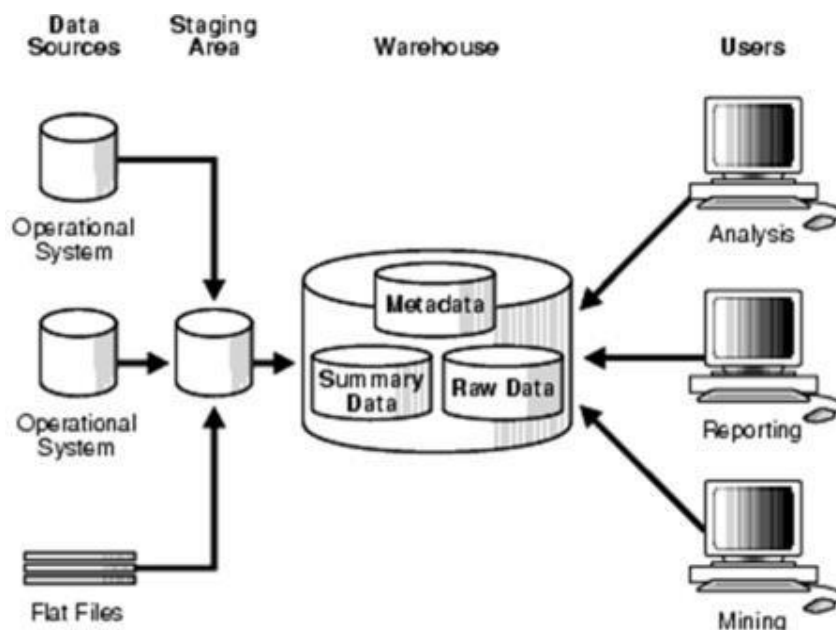


Figura 9. Arquitectura Ralph Kimbal.

2.5.2 Metodología Inmon

Metodología desarrollada en el año 1992 a cargo de Bill Inmon en su ejemplar “Building the Data Warehouse”. El objetivo de esta metodología es ser apoyo para la

toma de decisiones estratégicas. El detalle es que para almacenar información en los DM primero es importante normalizarla. Entonces debido a las grandes cantidades de datos que se almacenan, Inmon considera que el ambiente de origen de datos y el de acceso, deben estar físicamente separados, fijando así una estructura en capas del DW y DM, y forjando una dependencia del depósito central de datos. (Bustos & Mosquera, 2013, p.38)

La figura 10 propuesta por Moss & Atre (2003) evidencia esta arquitectura.

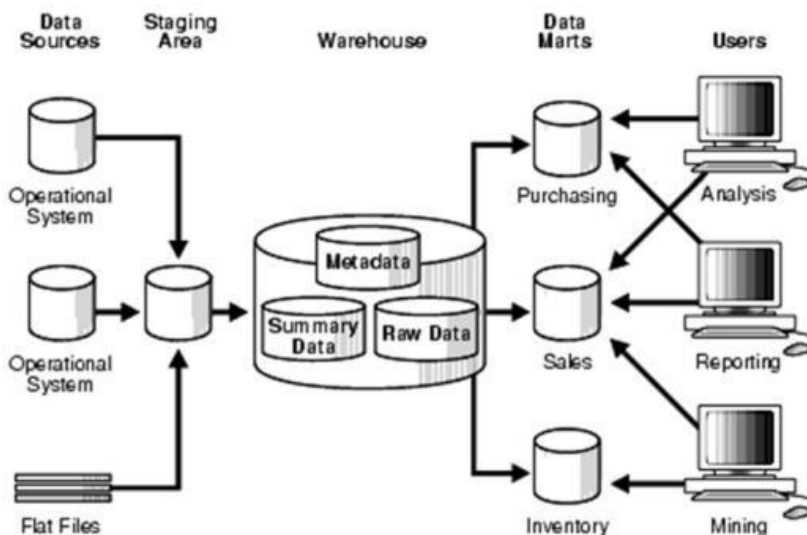


Figura 10. Arquitectura Inmon.

2.5.3 Metodología HEFESTO

HEFESTO es una metodología planteada por Bernabeu (2010) cuya propuesta está fundamentada en una investigación, comparando así metodologías existentes y experiencias propias del autor en el proceso de confección de almacenes de datos. Así mismo el autor manifiesta lo siguiente:

Se debe considerar que HEFESTO está en continua evolución.

Las fases de esta metodología se acoplan fácilmente al desarrollo y está enfocada en el análisis de requerimientos, identificación de indicadores y perspectivas, que

permitan realizar el análisis de datos específicos y obtener resultados de apoyo a la toma de decisiones de la organización. (p.85)

Para emplear esta metodología, se debe comprender cada paso a realizar, para no caer en la dificultad de tener que seguir un método al pie de la letra, sin saber que se está haciendo, ni por qué. Es así que la metodología HEFESTO puede ser adaptada en cualquier ciclo de vida que cumpla con la condición antes declarada.

Características de la metodología HEFESTO

Bernabeu (2010) propone las siguientes características:

- Los objetivos y resultados esperados en cada fase se distinguen fácilmente y son sencillos de comprender.
- Se basa en los requerimientos de los usuarios finales es capaz de adaptarse con facilidad y rapidez ante los cambios en el negocio.
- Reduce la resistencia al cambio, ya que involucra a los usuarios finales en cada etapa para que tome decisiones respecto al comportamiento y funciones del DW.
- Utiliza modelos conceptuales y lógicos, los cuales son sencillos de interpretar y analizar.
- Es independiente del tipo de ciclo de vida que se emplee para contener la metodología.
- Es independiente de las herramientas que se utilicen para su implementación.
- Es independiente de las estructuras físicas que contengan el DW y de su respectiva distribución.

- Cuando se culmina con una fase, los resultados obtenidos se convierten en el punto de partida para llevar a cabo el paso siguiente.

- Se aplica tanto para DW como para DM. (p.88)

Fases de la Metodología HEFESTO

Según Bernabeu (2010), la metodología HEFESTO, está conformada por:

Análisis de requerimientos: Lo primero que se debe realizar es identificar los requerimientos de los usuarios, a través de preguntas que determinen los objetivos de la organización. Por consiguiente, se analizarán las preguntas a fin de poder identificar cuáles serán los indicadores y perspectivas que serán consideradas para la construcción del DM o DW.

Análisis de los OLTP: Lo primero que se debe realizar es identificar los requerimientos de los usuarios, a través de preguntas que determinen los objetivos de la organización. Por consiguiente, se analizarán las preguntas a fin de poder identificar cuáles serán los indicadores y perspectivas que serán consideradas para la construcción del DM o DW.

Modelo lógico del DW: Aquí se establece el tipo de modelo que se desea utilizar (estrella, constelación o copo de nieve) y luego se lleva a cabo las acciones propias del caso, para diseñar las tablas de dimensiones y hechos. Al finalizar se hace la unión entre las tablas, dichas tablas contendrán un nombre que las identifique, un campo para la clave principal, se definirán los nombres de campos si es que no son intuitivos.

En este paso también se definirán la tabla o tablas de hechos, a través de los cuales se construirán los indicadores de estudio.

Dicho sea de paso, esta fase culmina con la unión entre tablas de dimensiones y las tablas de hechos.

Integración de datos: En este paso se procede a llenar el modelo con datos de calidad, utilizando el proceso extracción, transformación y carga (ETL) de tal manera que deposite los datos en el almacén de datos. Es probable que esta tarea a veces contenga una lógica un poco compleja, sin embargo, dependerá del correcto análisis, conocimientos y herramientas que se empleen con el fin de facilitar el trabajo. (p.90)

Tabla 3. Resumen del proceso ETL.

<i>Componente</i>	<i>Entrada</i>	<i>Proceso</i>	<i>Salida</i>
Extracción	Base de datos, hojas de cálculo, archivos de texto, otros.	Selección	Datos crudos (cargados a memoria)
Transformación	Datos crudos (cargados a memoria)	Limpieza, transformación, personalización, realización de cálculos, aplicación de funciones	Datos formateados, estructurados y resumidos de acuerdo a las necesidades
Carga	Datos formateados, estructurados y resumidos de acuerdo a las necesidades (aún en memoria)	Inserción	Datos formateados, estructurados y resumidos con persistencia en el DW.

Nota: Tomada de (Bustamante et al., 2011).

Es recomendable primero cargar los datos de las dimensiones, y luego de la tabla hechos, siempre considerando que la tabla hechos contendrá el nivel de granularidad.

En el esquema estrella, varias tablas de dimensiones comparten una tabla de hechos.

En el esquema copo de nieve, las tablas dimensiones generan nuevas dimensiones.

Después de cargar el DW en su totalidad, se debe establecer políticas de actualización de datos (refresco de datos).

Creación de cubos multidimensionales

Bernabeu (2010) también describe que el cubo multidimensional debe estar basado en el modelo lógico propuesto, y debe contener lo siguiente:

Mostrar la correcta distinción entre hechos de una tabla e indicadores de un cubo.

Mostrar de manera adecuada la distinción entre campos de una tabla de dimensiones y atributos del cubo: Indicadores, atributos, jerarquías. (p.92)

Capítulo 3

Materiales y métodos.

Este capítulo marca un hecho importante en nuestra investigación, ya que describe cómo se desarrolló la investigación, definiendo el tipo de investigación y el diseño de la misma, representado por medio de una gráfica el total del proyecto desde la concepción del proyecto hasta su finalización, asimismo se explica en qué consistió cada etapa.

3.1 Tipo de investigación

Según Rincón (2013) el tipo de investigación es tecnológica, ya que se propone un modelo para ayudar en la toma de decisiones al departamento de admisión, a través de descubrimientos ya realizados, utilizando tecnología. A su vez (Bello, 2006) menciona que la investigación tecnológica comprende con mayor énfasis la transformación, cuyo fin es obtener conocimiento para lograr modificar la realidad en estudio.

Por su parte Hernández, Fernández, & Baptista (2010) mencionan que esta investigación es de tipo descriptiva, debido a que implica una interacción con la información recolectada (data histórica) o determinar su comportamiento a lo largo del tiempo.

3.2 Diseño de la investigación

La figura 11, es de elaboración propia; muestra la metodología usada en el desarrollo del proyecto, que tuvo de base a la metodología HEFESTO versión 2.0, utilizada para el proceso de proyectos en IN y contiene las fases de: Análisis de requerimientos, análisis de los OLTP, modelo lógico del DM o DWH, integración de datos. Con todo esto tenemos una idea clara de cómo desarrollar el proyecto de IN.

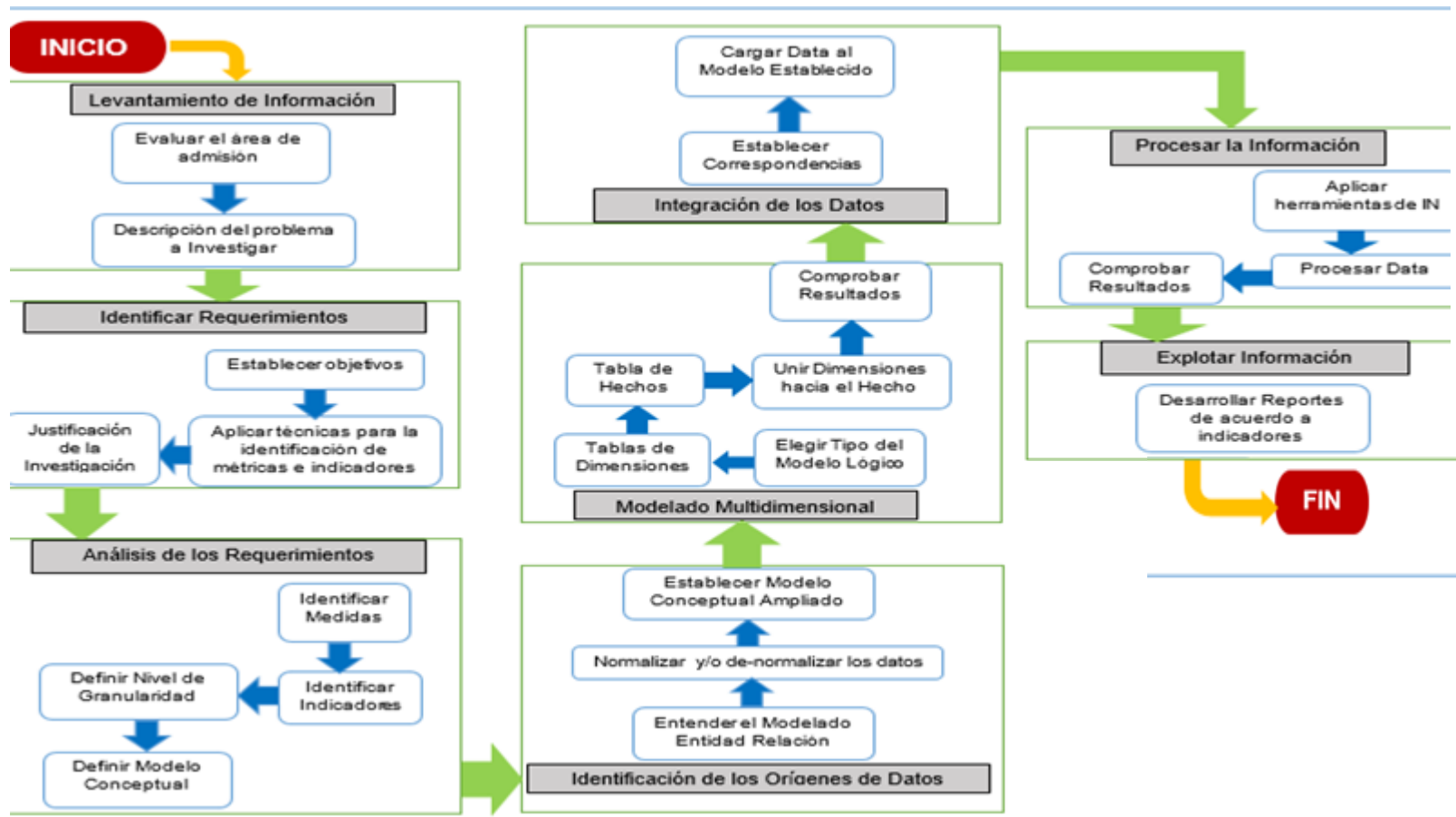


Figura 11. Diseño de la Investigación.

3.3 Etapas de la investigación

3.3.1 Levantamiento de información

Evaluar el departamento de Admisión

El levantamiento de información se hizo a través de reuniones pactadas con el coordinador del departamento de admisión, dando lugar a: entrevistas y lluvia ideas. Esto permitió detectar el problema al que nos enfrentábamos e iniciar con esta investigación.

Descripción del problema

Se describió la dificultad de la que adolecía el departamento de Admisión, tal y como se describe en el capítulo 1, el problema.

3.3.2 Identificación de requerimientos

En esta fase se procedió a identificar los requerimientos del departamento y transformarlos en objetivos los cuales permitieron seleccionar las principales ideas y necesidades del cliente. También se aplicó el método log files (análisis de archivos) para identificar las métricas e indicadores y tener lo necesario para justificar la investigación.

Algunos de los requerimientos obtenidos fueron: conocer las zonas o lugares de los que provenían más ingresantes, saber qué sector de admisión es el que más registra ingresantes y tener un comparativo entre las campañas de admisión.

3.3.3 Análisis de los requerimientos

Esta fase permitió identificar las medidas e indicadores con la información que se extrajo de la fase anterior (Ingresantes según religión, modalidad, etc.). Asimismo fue de gran ayuda para definir el nivel de granularidad del modelo de IN (por campañas), por otro lado también se definió el modelo conceptual a utilizar.

3.3.4 Identificación de los orígenes de datos

En esta fase se estudió y analizó el origen de los datos (base de datos de Admisión) que fue provista por el Área de Desarrollo de la Dirección General de Tecnologías de Información (DIGETI). También se normalizó y denormalizó toda la data necesaria para el diseño del modelo de IN. Al final se estableció el modelo lógico ampliado apto para cumplir los propósitos.

3.3.5 Modelado multidimensional

En esta fase teniendo ya el modelo lógico bien establecido se procedió al modelado y/o armado del modelo multidimensional tipo estrella, creando la tabla principal (hecho_admision) y las siguientes tablas de dimensiones: adm_carrera, adm_modalidad, adm_postulante, adm_procedencia (procedencia del postulante), adm_sectoradm (sector donde hay un promotor de admisión), adm_religion, adm_tiempo cada cual con sus campos respectivos. Teniendo las dimensiones listas se unió hacia el hecho. El gestor de base de datos que se utilizó fue Postgresql, que es libre.

3.3.6 Integración de los datos

Lo primero que se realizó en esta fase es establecer correspondencias, que entendemos como un tipo de vista o manera de generar una consulta de todos los datos que serán cargados al hecho_admision. En seguida pudimos cargar con normalidad los hechos al modelo dimensional establecido, utilizando una herramienta del Kit de Pentaho (Spoon, antes Kettle).

3.3.7 Procesamiento de la información

En esta fase se aplicó las herramientas de la Suite de Pentaho como (Schema Workbench) para la construcción del cubo. Cabe mencionar que hicimos un solo cubo para todas las dimensiones establecidas e importamos un tipo de archivo XML.

3.3.8 Explotar información

En esta fase se utilizó la Suite de Pentaho Business Analytics (versión 6.0.1 de prueba) para explotar toda la información y empezar a generar los reportes necesarios que muestren los resultados esperados de acuerdo a indicadores. Aquí se dio lugar a los reportes a través de gráficos estadísticos.

Capítulo 4

Resultados y discusión.

Este capítulo todo el proceso de diseño del modelo de IN, de acuerdo a las necesidades (indicadores) manifestados por el cliente y su validación correspondiente.

Asimismo se detalla la explotación del modelo propuesto a través de gráficos estadísticos, de acuerdo a cada indicador. Para todo este proceso se utilizó el Business Analytics de Pentaho y su servidor.

4.1 Elaboración del modelo de IN para Admisión

4.1.1 Proceso de selección de la información

Para obtener la información (data operacional) y manipularla, se hizo una consulta SQL (ver anexo 01) a la misma base de datos de Admisión y se exportó en formato xlsx, solo los datos que sean útiles para el diseño de nuestro modelo, ya que la estructura de su base de datos depende del modelo de base de datos del sistema académico. De esta manera no se toca para nada los datos fuente.

La figura 12 (fuente: Microsoft Excel), muestra la data exportada solo con los datos necesarios para nuestro modelo de IN.

Backup_TESIS.xlsx - Excel

HERRAMIENTAS DE TABLA

ARCHIVO INICIO INSERTAR DISEÑO DE PÁGINA FÓRMULAS DATOS REVISAR VISTA DISEÑO

Calibri 11 Fuente Alineación Número

Normal Buena Incorrecto Neutral Cálculo

Celda de co... Celda vincul... Entrada Notas Salida

DOCUMENTOS_CODUNIV

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	DOCUMENTO	MATRI	ALUMNO	DISTRITO_PRO	PROVINCIA_PRO	DEPARTAMEN	SEXO	FECHA_N/	EDAD	RELIGION	SECTOR_ADMSIS	EP_POSTULACION	SEMESTRE	FECHA_EXAMEN	COLEGIO	MODALIDAD
2	201410776	2014-1	VALERA ARO, JAIR JOEL	RAIMONDI	ATALAYA	UCAYALI	M	06/02/95	18	Adventista	Central Tarapoto	E.P. de Contabilidad y Gestio	No Tribut: 2013-3	11/17/13	Otros	Pre Ingreso
3	201410829	2014-1	MARTINEZ AHUANARI, LADYS ESTHE	RAIMONDI	ATALAYA	UCAYALI	F	02/14/92	21	Adventista	Central Tarapoto	E.P. de Psicología - Filial Tarapoto	2013-3	11/17/13	Otros	Examen de Admisión
4	201310699	2014-1	CARRION ABARCA, FERNANDO	RAIMONDI	ATALAYA	UCAYALI	M	01/07/95	19	Adventista	Central Tarapoto	E.P. de Psicología - Filial Tarapoto	2014-1	03/06/14	Otros	Examen de Admisión
5	201121384	2011-1	ZELADA ESCOBEDO, ADESMIRO	LEIMEBAMBA	CHACHAPOYAS	AMAZONAS	M	03/21/88	22	Adventista	Central Tarapoto	E.P. de Ingeniería de Sistemas - Filial	2011-1	03/07/11	Otros	Examen de Admisión
6	201121570	2011-1	GARRO DIAZ, KEYLY TATIANA	LEIMEBAMBA	CHACHAPOYAS	AMAZONAS	F	03/30/89	21	Otros	Central Tarapoto	E.P. de Ingeniería Ambiental - Filial	Ta 2011-1	02/22/11	Otros	Examen de Admisión
7	201521041	2015-1	QUISPE HUAMAN, RUTH ARMINDA	ALTO SELVA ALEG	AREQUIPA	AREQUIPA	F	04/23/98	16	Adventista	Central Tarapoto	E.P. de Arquitectura - Filial Tarapoto	2015-1	02/25/15	MANUEL GC	Examen de Admisión
8	200611401	2009-2	MEZA TORRES, SAUL MELCHOR	ANDAMARCA	CONCEPCION	JUNIN	M	07/25/76	37	Adventista	Central Tarapoto	E.P. de Psicología - Filial Tarapoto	2014-1	03/02/14	Otros	Cambio de Carrera
9	201521167	2015-1	VIVAS RAMIREZ, KIMBERLY ANDREA	SATIPO	SATIPO	JUNIN	F	02/14/98	17	Adventista	Central Tarapoto	E.P. de Marketing y Negocios Internai	2015-1	03/05/15	Otros	Examen de Admisión
10	201521073	2015-1	HUILLCA CORDOVA, LUIS ANGEL	SATIPO	SATIPO	JUNIN	M	12/24/96	18	Adventista	Central Tarapoto	E.P. de Ingeniería Ambiental - Filial	Ta 2015-1	03/02/15	Otros	Examen de Admisión
11	201420654	2014-1	VIVAS RAMIREZ, BRAYAM MARDIN	SATIPO	SATIPO	JUNIN	M	08/07/96	17	Adventista	Juanjui	E.P. de Ingeniería Ambiental - Filial	Ta 2014-1	03/02/14	Otros	Examen de Admisión
12	201610856	2016-1	ROSALES LAZO, JOSUE ISMAEL	SATIPO	SATIPO	JUNIN	M	11/17/98	16	Adventista	Chanchamayo	E.P. de Ingeniería Ambiental - Filial	Ta 2015-3	11/14/15	Satipo	Pre Ingreso
13	201420604	2014-1	PARADO AQUINO, GINA NALDA	SATIPO	SATIPO	JUNIN	F	06/19/95	18	Adventista	Tacna	E.P. de AdministraciNo - MencNo : G	2014-1	03/03/14	Otros	Examen de Admisión
14	201121470	2011-1	MORETO RUIZ, DANNAE MORAYA	CHANCHAMAYO	CHANCHAMAYO	JUNIN	F	12/28/93	17	Adventista	Central Tarapoto	E.P. de Psicología - Filial Tarapoto	2011-1	03/04/11	Otros	Examen de Admisión
15	201210665	2012-1	GUTIERREZ FLORES, KATHERINE KIAF	CHANCHAMAYO	CHANCHAMAYO	JUNIN	F	03/01/94	17	Adventista	Central Tarapoto	E.P. de AdministraciNo - MencNo : G	2011-3	11/20/11	Otros	Examen de Admisión
16	201410774	2014-1	AQUINO ESPINOZA, MARIMAR	CHANCHAMAYO	CHANCHAMAYO	JUNIN	F	03/06/96	17	Adventista	Central Tarapoto	E.P. de Contabilidad y Gestio	No Tribut: 2013-3	11/17/13	Otros	Pre Ingreso
17	201323407	2013-2	GASTELO QUIROZ, ASTRID EMPERAT	CHICLAYO	CHICLAYO	LAMBAYEQUE F	12/12/92	20	Adventista	Central Tarapoto	E.P. de AdministraciNo - MencNo : G	2013-2	08/07/13	Otros	Traslado Externo	
18	201420632	2014-1	CANGO RAMIREZ, DEYSSI GUILLERM	CHICLAYO	CHICLAYO	LAMBAYEQUE F	08/11/96	17	Adventista	Central Tarapoto	E.P. de Contabilidad y Gestio	No Tribut: 2014-1	02/05/14	Jose de San	Pre Ingreso	
19	201520947	2015-1	DIAZ SANCHEZ, LEONOR NOEMI	CHICLAYO	CHICLAYO	LAMBAYEQUE F	03/21/97	17	Adventista	Central Tarapoto	E.P. de AdministraciNo - MencNo : G	2015-1	03/01/15	Jose de San	Examen de Admisión	
20	201122969	2012-1	NEIRA MIJA, EDWIN DAVID	CHICLAYO	CHICLAYO	LAMBAYEQUE M	07/02/94	17	Adventista	Central Tarapoto	E.P. de Contabilidad y Finanzas - Filial	2011-2	08/08/11	Otros	Examen de Admisión	
21	201320896	2013-1	GUERRERO ZAMORA, SHARON JENN	CHICLAYO	CHICLAYO	LAMBAYEQUE F	11/13/94	18	Adventista	Central Tarapoto	E.P. de AdministraciNo - MencNo : G	2013-1	03/06/13	Otros	Examen de Admisión	
22	201320927	2013-1	OLIVERA SANCHEZ, MIGUEL ANGEL	CHICLAYO	CHICLAYO	LAMBAYEQUE M	03/04/95	18	Adventista	Central Tarapoto	E.P. de Contabilidad y Gestio	No Tribut: 2013-1	03/04/13	Otros	Examen de Admisión	
23	201520921	2015-1	VILLANUEVA RAMOS, EVA ENI	CHICLAYO	CHICLAYO	LAMBAYEQUE F	05/01/98	16	Adventista	Alto Mayo	E.P. de AdministraciNo - MencNo : G	2015-1	03/01/15	Otros	Examen de Admisión	
24	201522289	2015-2	CRUZ MUNOZ, JERY JELEN	CHICLAYO	CHICLAYO	LAMBAYEQUE F	10/05/98	16	Catolico	Central Tarapoto	E.P. de Ingeniería de Sistemas - Filial	2015-2	07/12/15	Otros	Examen de Admisión	
25	201510574	2015-1	ROJAS CAMPOS, HELTON JHON	CHICLAYO	CHICLAYO	LAMBAYEQUE M	12/18/97	16	Catolico	Central Tarapoto	E.P. de Arquitectura - Filial Tarapoto	2014-3	11/16/14	Otros	Examen de Admisión	
26	201121477	2011-1	VARGAS PEREZ, ALDO	CHICLAYO	CHICLAYO	LAMBAYEQUE M	11/10/93	17	Otros	Chiclayo	E.P. de Psicología - Filial Tarapoto	2011-1	03/07/11	Otros	Examen de Admisión	
27	201121457	2011-1	PALACIOS GUERRERO, ARANZA JIME	CHICLAYO	CHICLAYO	LAMBAYEQUE F	06/19/94	16	Adventista	Central Tarapoto	E.P. de Psicología - Filial Tarapoto	2011-1	03/08/11	Chiclayo	Examen de Admisión	
28	200921016	2009-1	BECERRA HERRERA, JULIO CESAR	CHICLAYO	CHICLAYO	LAMBAYEQUE M	09/18/88	20	Adventista	Central Tarapoto	E.P. de Psicología - Filial Tarapoto	2009-1	03/12/09	KARL WIESS	Examen de Admisión	
29	200820372	2008-2	OLAYA OROSCO, KATHERIN DIINA	CHICLAYO	CHICLAYO	LAMBAYEQUE F	07/28/90	17	Adventista	Central Tarapoto	E.P. de Marketing y Negocios Internai	2008-2	07/06/08	ABILIA OCAI	Examen de Admisión	
30	201420645	2014-1	TINEO SANCHEZ, SANDRA QUEEN YA	CHICLAYO	CHICLAYO	LAMBAYEQUE F	08/12/96	17	Catolico	Central Tarapoto	E.P. de Contabilidad y Gestio	No Tribut: 2014-1	02/05/14	Ofelia Velas	Pre Ingreso	
31	200610788	2007-1	CORONEL SANTA MARIA, JESNS OBD	CHICLAYO	CHICLAYO	LAMBAYEQUE M	12/11/86	20	Adventista	Central Tarapoto	E.P. de Marketing y Negocios Internai	2007-1	02/22/07	ADVENTISTA	Examen de Admisión	
32	200810133	2010-0	QUISPE COBOS, JOHANY	CHONGOYAPE	CHICLAYO	LAMBAYEQUE F	01/11/80	28	Adventista	Central Tarapoto	E.P. de Psicología - Filial Tarapoto	2008-1	03/11/08	JOSE JIMENI	Examen de Admisión	
33	201612513	2016-1	MIRANDA CUEVA, DARWIN SEGUND	MONSEFU	CHICLAYO	LAMBAYEQUE M	07/08/98	17	Adventista	Cajamarca	E.P. de AdministraciNo - MencNo : G	2016-1	02/28/16	Otros	Examen de Admisión	
34	201320926	2013-1	DIAZ SANCHEZ, JOSUE ABRAHAM	PATAPO	CHICLAYO	LAMBAYEQUE M	04/29/95	17	Adventista	Central Tarapoto	E.P. de Contabilidad y Gestio	No Tribut: 2013-1	03/04/13	Otros	Examen de Admisión	

Export Worksheet BACKUP SQL

Figura 12. Data histórica de Admisión.

4.1.2 Proceso de filtración de la data

Después de tener la información en .xlsx, se puede manipular y filtrar con facilidad. La figura 13 (fuente: Microsoft Excel), prototipa solo las columnas que se utilizaron, que son 15 para cumplir a cabalidad con nuestro modelo.

ID DocuMat	ID Alumno	APELLIDOS	NOMBRES	ID Procc	DISTRITO_PRO	PROVINCIA_PRO	DEPARTAM	FECHA_ENT	ID Recl	RELIGION	ID Sectm	SECTOR ADMISION	ID Carr	EP POSTULACION				
2444	201220771	2012-1	201220771	UMBO YANAYACO	YESICA	PRO152	PARINAS	TALARA	PIURA	F	08/09/93	16	RELO01	ADVENTISTA	SEC008	CENTRAL TARAPOTO	CP13	E.P. DE PSICOLOGIA - FI
2447	2013110488	2013-1	2013110488	DIAZ QUISEP	MANASES ELIU	PRO138	NUEVA CAJAMARCA	RIQUIA	SAN MARTIN	M	06/24/96	16	RELO01	ADVENTISTA	SEC008	CENTRAL TARAPOTO	CP13	E.P. DE PSICOLOGIA - FI
2449	201310500	2013-1	201310500	SIAZ CARDENAS	MILAGROS	PRO156	VARGAS PROGRESO	TOCACHE	SAN MARTIN	F	11/23/96	16	RELO01	ADVENTISTA	SEC008	CENTRAL TARAPOTO	CP13	E.P. DE PSICOLOGIA - FI
2449	201410824	2014-1	201410824	SHUNA CASTRO	ANNIE CAROLINE	PRO156	VARGAS GUERRA	UCAYALI	LORETO	F	05/22/97	16	RELO01	ADVENTISTA	SEC008	CENTRAL TARAPOTO	CP13	E.P. DE PSICOLOGIA - FI
2450	201410831	2014-1	201410831	CALLE TARRILLO	ALBERTO	PRO009	BELLAVISTA	BELLAVISTA	SAN MARTIN	M	13-JAN-97	16	RELO01	ADVENTISTA	SEC008	CENTRAL TARAPOTO	CP13	E.P. DE PSICOLOGIA - FI
2451	201410842	2014-1	201410842	QUIEVEDO HUANCAS	EDITRIZA SADITH	PRO000	NO REGISTRA	NO REGISTRA	NO REGISTRA	F	09/01/96	17	RELO01	ADVENTISTA	SEC008	CENTRAL TARAPOTO	CP13	E.P. DE PSICOLOGIA - FI
2452	201410847	2014-1	201410847	Lopez IGNACIO	DEIVI ISAC	PRO071	SAN MARTIN	EL DORADO	SAN MARTIN	M	10/22/95	18	RELO01	ADVENTISTA	SEC008	CENTRAL TARAPOTO	CP13	E.P. DE PSICOLOGIA - FI
2454	201410848	2014-1	201410848	FLORES SALAZAR	ROSA LUZ	PRO142	TARAPOTO	SAN MARTIN	SAN MARTIN	F	08/04/97	16	RELO01	ADVENTISTA	SEC008	CENTRAL TARAPOTO	CP13	E.P. DE PSICOLOGIA - FI
2454	201420721	2014-1	201420721	RAMIREZ IRENE	MERYLY AYDEE	PRO140	HUARANGO	SAN IGNACIO	CAJAMARCA	F	02/03/97	17	RELO01	ADVENTISTA	SEC008	CENTRAL TARAPOTO	CP13	E.P. DE PSICOLOGIA - FI
2455	201420737	2015-1	201420737	SALAS TUANAMA	GEDEON	PRO087	LAMAS	LAMAS	SAN MARTIN	M	03/28/96	17	RELO01	ADVENTISTA	SEC008	CENTRAL TARAPOTO	CP13	E.P. DE PSICOLOGIA - FI
2456	201420741	2014-1	201420741	SANTOS GOMEZ	YOLANDA	PRO071	SAN MARTIN	EL DORADO	SAN MARTIN	F	02/05/97	16	RELO01	ADVENTISTA	SEC008	CENTRAL TARAPOTO	CP13	E.P. DE PSICOLOGIA - FI
2457	200910398	No Reg	200910398	SALAS DEL CASTILLO	CELSO	PRO142	TARAPOTO	SAN MARTIN	SAN MARTIN	M	04/02/92	16	RELO02	CATOLICO	SEC008	CENTRAL TARAPOTO	CP13	E.P. DE PSICOLOGIA - FI
2458	200910418	2009-1	200910418	MUNOZ DEL AGUILA	GIAN CARLO	PRO142	MORALES	SAN MARTIN	SAN MARTIN	M	01/04/92	16	RELO02	CATOLICO	SEC008	CENTRAL TARAPOTO	CP13	E.P. DE PSICOLOGIA - FI
2459	201110820	2011-1	201110820	PINEDO VELA	KIARA VANESSA	PRO142	TARAPOTO	SAN MARTIN	SAN MARTIN	F	08/24/94	16	RELO02	CATOLICO	SEC008	CENTRAL TARAPOTO	CP13	E.P. DE PSICOLOGIA - FI
2460	201220768	2012-1	201220768	ZEGARRA YUMBATO	GUMERCINDO ELISEO	PRO001	YURIMAGUAS	ALTO AMAZONAS	LORETO	M	01/10/91	21	RELO02	CATOLICO	SEC008	CENTRAL TARAPOTO	CP13	E.P. DE PSICOLOGIA - FI
2461	2012121004	2013-1	2012121004	REATEGUI MACAHUACHI	GRESY	PRO142	TARAPOTO	SAN MARTIN	SAN MARTIN	F	06/05/96	16	RELO02	CATOLICO	SEC008	CENTRAL TARAPOTO	CP13	E.P. DE PSICOLOGIA - FI
2462	201410850	2014-1	201410850	CUZQUE COTRINA	GEORGE STEEB	PRO142	TARAPOTO	SAN MARTIN	SAN MARTIN	M	05/20/97	16	RELO02	CATOLICO	SEC008	CENTRAL TARAPOTO	CP13	E.P. DE PSICOLOGIA - FI
2463	201420729	2015-1	201420729	DAMIAN AREVALO	GIANNY NATHALIE	PRO142	TARAPOTO	SAN MARTIN	SAN MARTIN	F	06/25/98	15	RELO02	CATOLICO	SEC008	CENTRAL TARAPOTO	CP13	E.P. DE PSICOLOGIA - FI
2464	201420736	2014-1	201420736	VELA RAMIREZ	ANGELICA SOLANGE	PRO155	TRUJILLO	TRUJILLO	LA LIBERTAD	F	12/24/96	17	RELO02	CATOLICO	SEC008	CENTRAL TARAPOTO	CP13	E.P. DE PSICOLOGIA - FI
2465	201420739	2014-1	201420739	LINARES ROJAS	CLAUDIA GRACIELA	PRO000	NO REGISTRA	NO REGISTRA	NO REGISTRA	F	04/17/97	16	RELO02	CATOLICO	SEC008	CENTRAL TARAPOTO	CP13	E.P. DE PSICOLOGIA - FI
2466	201420742	2014-1	201420742	SANCHEZ GONZA	DIEGO ALONZO	PRO001	YURIMAGUAS	ALTO AMAZONAS	LORETO	M	12/28/95	18	RELO02	CATOLICO	SEC008	CENTRAL TARAPOTO	CP13	E.P. DE PSICOLOGIA - FI
2467	200720555	No Reg	200720555	BOCANEGRA CALDERON	LIRIA MICHELIN	PRO142	TARAPOTO	SAN MARTIN	SAN MARTIN	F	05/28/90	17	RELO03	EVANGELICO	SEC008	CENTRAL TARAPOTO	CP13	E.P. DE PSICOLOGIA - FI
2468	200910420	2009-1	200910420	AGUILAR MACHUCA	ROXANA	PRO123	MOYOBAMBA	MOYOBAMBA	SAN MARTIN	F	08/26/90	18	RELO03	EVANGELICO	SEC008	CENTRAL TARAPOTO	CP13	E.P. DE PSICOLOGIA - FI
2469	201220781	2012-1	201220781	VILCA FLORES	SHARON PRICILA	PRO142	MORALES	SAN MARTIN	SAN MARTIN	F	08/27/94	17	RELO03	EVANGELICO	SEC008	CENTRAL TARAPOTO	CP13	E.P. DE PSICOLOGIA - FI
2470	200720562	2010-1	200720562	TORRES VELA	DOUBETH	PRO142	TARAPOTO	SAN MARTIN	SAN MARTIN	F	01/30/90	17	RELO04	OTROS	SEC008	CENTRAL TARAPOTO	CP13	E.P. DE PSICOLOGIA - FI
2471	201420726	2014-1	201420726	GERMAN OROSCO	LUZ GEMITA	PRO142	MORALES	SAN MARTIN	SAN MARTIN	F	04/24/96	17	RELO04	OTROS	SEC008	CENTRAL TARAPOTO	CP13	E.P. DE PSICOLOGIA - FI
2472	201612588	2016-1	201612588	REQUEJO SHAPIAMA	SARITA	PRO142	CHAZUTA	SAN MARTIN	SAN MARTIN	F	05/12/97	18	RELO01	ADVENTISTA	SEC008	CHANCHAMAYO	CP13	E.P. DE PSICOLOGIA - FI
2473	200910415	2009-1	200910415	MILIAN RODAS	RUBI	PRO021	LLAMA	CHOTA	CAJAMARCA	F	08/07/92	16	RELO01	ADVENTISTA	SEC009	CHICLAYO	CP13	E.P. DE PSICOLOGIA - FI
2474	201410835	2015-1	201410835	CANAQUIRI CHOTA	ANNIE MARIBI	PRO122	IQUITOS	MAYNAS	LORETO	F	09/06/96	17	RELO01	ADVENTISTA	SEC014	IQUITOS	CP13	E.P. DE PSICOLOGIA - FI
2475	201310487	2013-1	201310487	VALLES PINA	IRIS BERTHA	PRO119	NAUTA	LORETO	LORETO	F	09/07/96	16	RELO04	OTROS	SEC014	IQUITOS	CP13	E.P. DE PSICOLOGIA - FI
2476	201210610	2012-1	201210610	FRIAS GONZALES	JOSE ELVIS	PRO085	JAEN	JAEN	CAJAMARCA	M	08/16/93	18	RELO01	ADVENTISTA	SEC015	JAEN	CP13	E.P. DE PSICOLOGIA - FI
2477	201122915	2012-1	201122915	TELLO GALLEGO	LEYDI ESTHER	PRO157	BAGUA GRANDE	UTCUBAMBA	AMAZONAS	F	11/09/95	15	RELO01	ADVENTISTA	SEC018	NUEVA CAJAMARCA	CP13	E.P. DE PSICOLOGIA - FI
2478	201110428	2012-1	201110428	CORDOVA NOLE	ELIZABETH LEITZY	PRO000	NO REGISTRA	NO REGISTRA	NO REGISTRA	F	01/22/94	18	RELO01	ADVENTISTA	SEC018	PIURA	CP13	E.P. DE PSICOLOGIA - FI
2479	201110468	2012-1	201110468	CORDOVA NOLE	ELIZABETH LEITZY	PRO150	SULLANA	SULLANA	PIURA	F	01/22/94	17	RELO01	ADVENTISTA	SEC018	PIURA	CP13	E.P. DE PSICOLOGIA - FI
2480	201110789	2013-1	201110789	SALAZAR NAVARRRO	REINA	PRO051	YARINACUCHA	CORONEL PORTILLO	UCAYALI	F	02/28/92	18	RELO01	ADVENTISTA	SEC019	PUCALLPA	CP13	E.P. DE PSICOLOGIA - FI
2481	201220751	2012-1	201220751	MENDOZA SALAS	NOEMI	PRO123	MOYOBAMBA	MOYOBAMBA	SAN MARTIN	F	01/08/95	17	RELO01	ADVENTISTA	SEC020	RIQUIA	CP13	E.P. DE PSICOLOGIA - FI
2482	201220737	2012-1	201220737	VERA CALDERON	LOIDA NORALI	PRO141	EDUARDO VILLANUEVA	SAN MARCOS	CAJAMARCA	F	05/19/94	17	RELO01	ADVENTISTA	SEC027	TRUJILLO	CP13	E.P. DE PSICOLOGIA - FI
2483	201510538	2015-1	201510538	LEIVA LEYVA	DAYSY LISETH	PRO073	CASCAS	GRAN CHIMU	LA LIBERTAD	F	07/20/98	16	RELO01	ADVENTISTA	SEC027	TRUJILLO	CP13	E.P. DE PSICOLOGIA - FI

Figura 13. Datos filtrados.

Donde:

Se eliminan los campos que sean de valor NULL, para que no cree conflicto en el momento de la carga del ETL

Los tipos de campo: DATE, TIMESTAMP y FLOAT, muchas veces ocasionan conflicto, por lo que deben ser configurados según la base de datos que estemos utilizando, ver figura 14 (fuente: Spoon).

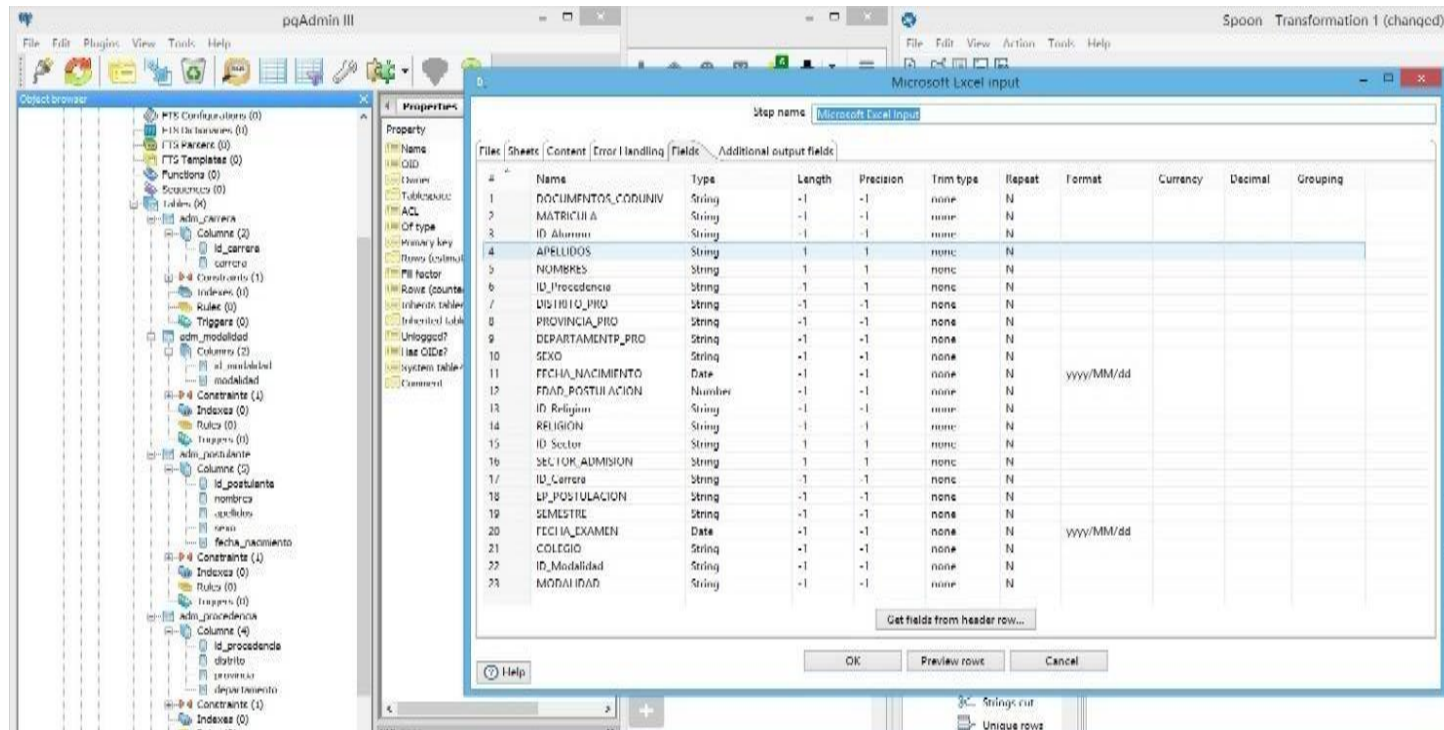


Figura 14. Transformación de datos con Spoon.

En nuestro caso el problema fue la fecha con formato DD/MM/YYYY, al formato de la base de datos YYYY/MM/DD.

4.1.3 Proceso de extracción y carga

Un detalle antes de hacer el ETL, se utilizó a postgres como motor de base de datos, y creamos allí un modelo dimensional para nuestro DM (ver código en el anexo 02), que reciba toda la carga del xlsx a postgresql. Observemos la figura 15 (fuente: Postgresql).

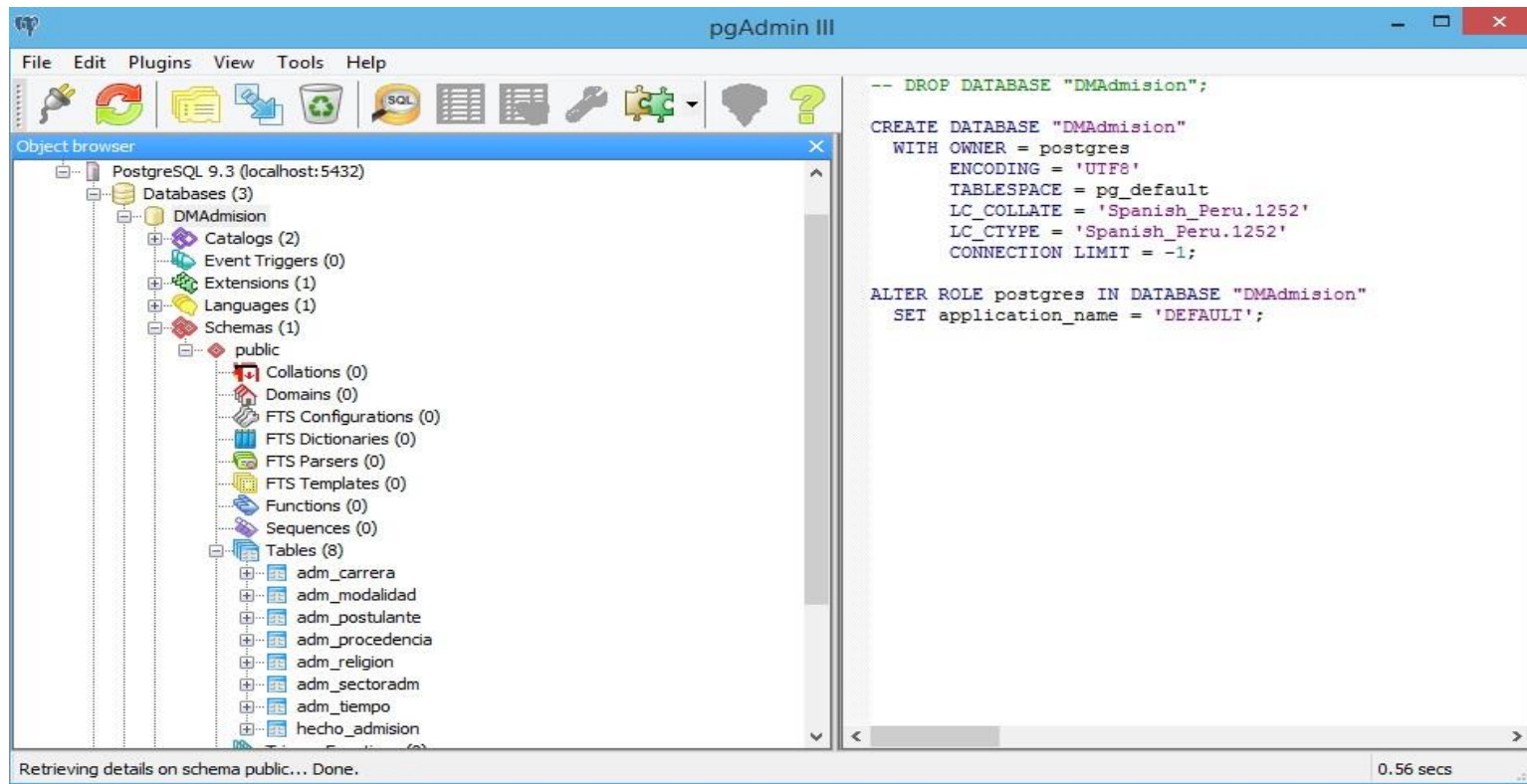


Figura 15. Nuevo modelo de base de datos admisión.

Al tener listo la nueva base de datos para la carga se hace la conexión respectiva a la base de datos desde el Spoon, como se observa en la figura 16 (fuente: Spoon).

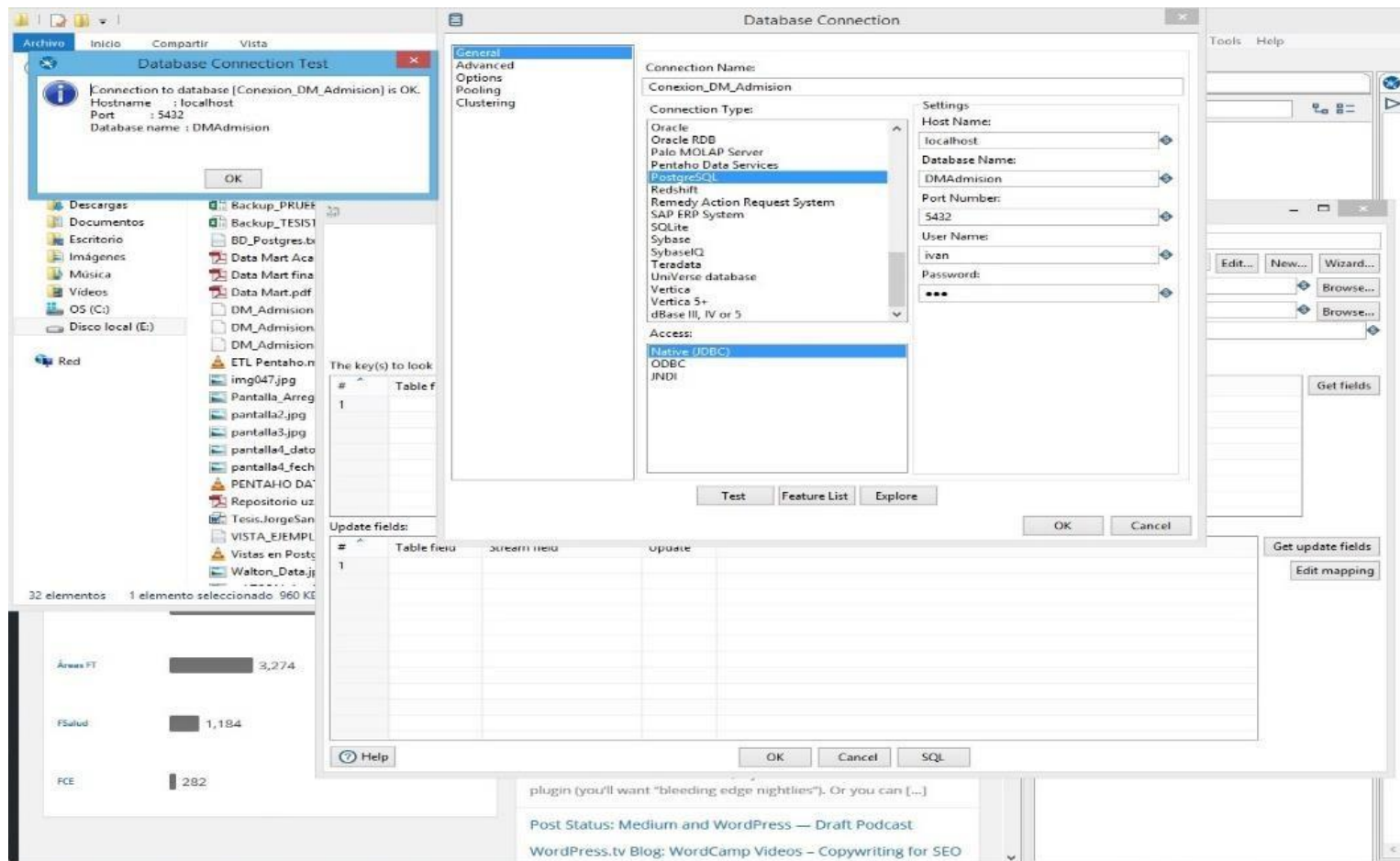


Figura 16. Test de conexión a la nueva base de datos.

Para realizar el ETL, se utilizó dos herramientas del Spoon: la primera Select values que sirve para seleccionar solo los campos que queremos, transformar algunos si es necesario, la segunda herramienta fue el Insert/Update que toma los datos seleccionados del Select values y los carga al nuevo modelo.

En la figura 17 (fuente: Spoon), se aprecia todo el ETL y carga de data del .xlsx hacia el nuevo modelo de datos (datamart). De la misma forma en la parte inferior se observa el tiempo que tarda en migrar.

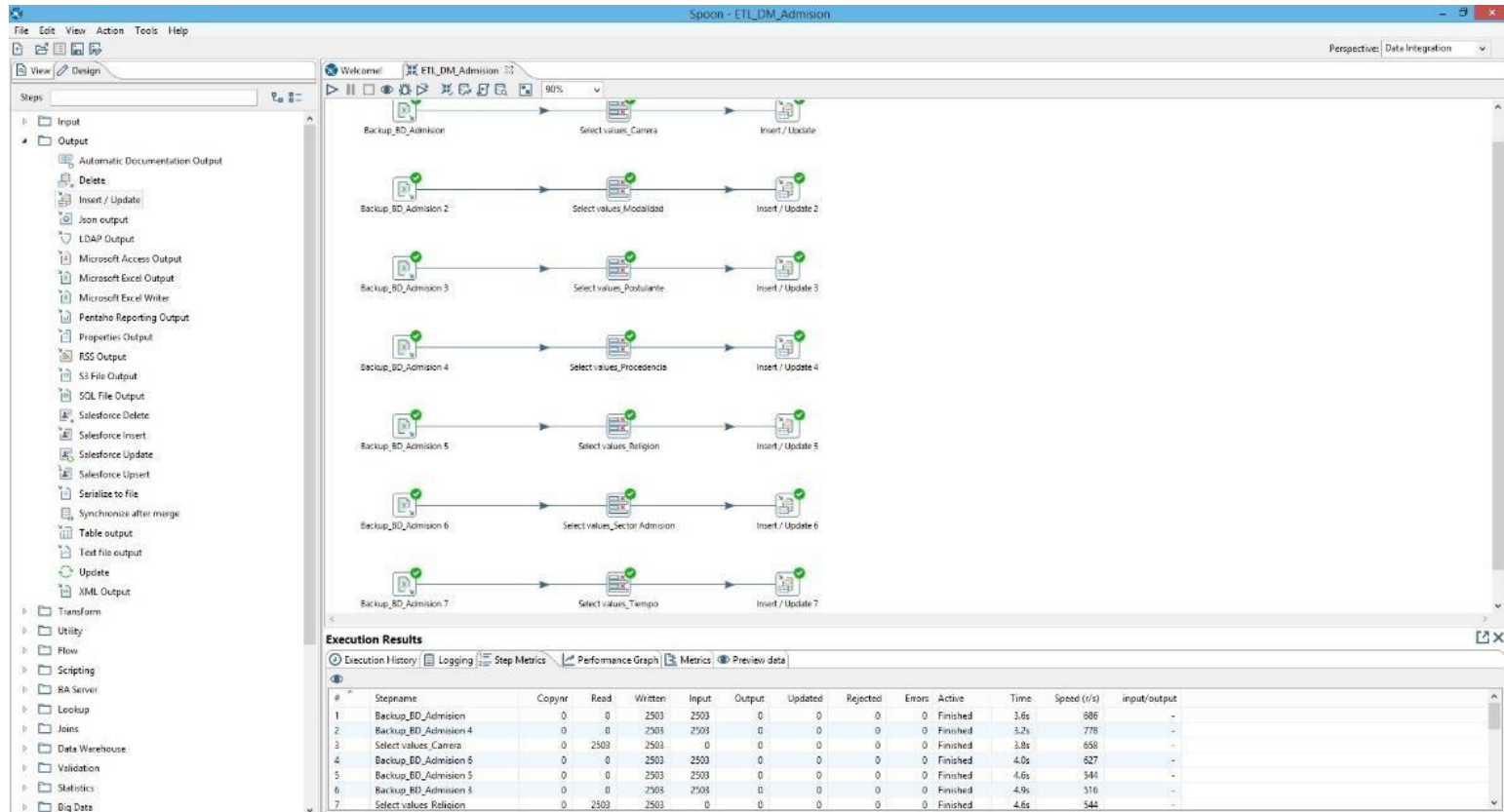


Figura 17. ETL Admisión y el tiempo de proceso utilizando Spoon.

4.1.4 Proceso de diseño del cubo de dimensiones

La figura 18 detalla el modelo lógico de datos tipo estrella, el cual ha sido creado en el gestor de base de datos PostgreSQL y trabajado con la herramienta Pentaho Report Designer con su extensión SQL Query Designer; a su vez la figura 19 muestra las dimensiones utilizadas a través de la herramienta de Pentaho Schema Workbench.

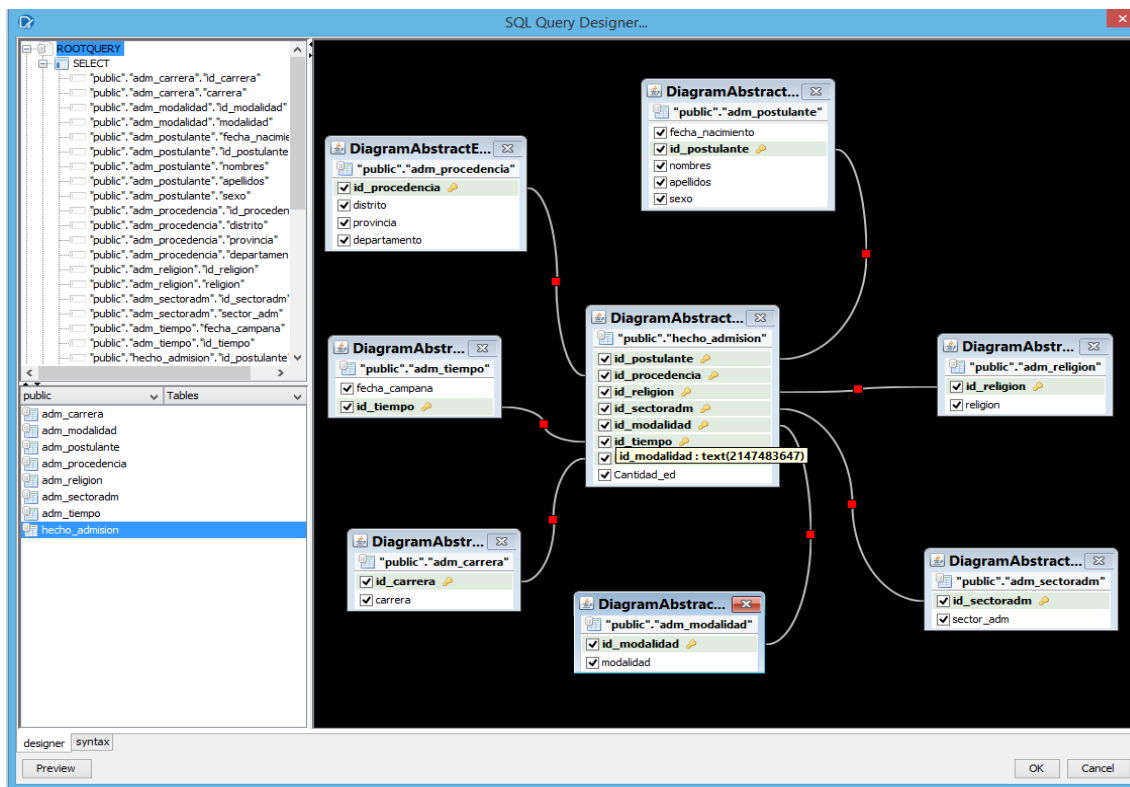


Figura 18. Modelo de datos estrella.

Schema Workbench

File Edit View Options Windows Help

Schema - PruebaAdmision (Cubo_Admision.xml)*

Table for Hierarchy

Attribute	Value
schema	public
name	adm_tiempo
alias	

The screenshot displays the Schema Workbench interface for a cube named 'ADMISION'. The left pane shows a hierarchical tree of dimensions and their associated tables:

- ADMISION
 - Table: hecho_admision
 - Carrera
 - default
 - Carrera
 - Table: adm_carrera
 - Modalidad
 - default
 - Modalidad
 - Table: adm_modalid
 - Postulante
 - default
 - Nombres
 - Apellidos
 - Table: adm_postular
 - Procedencia
 - default
 - Distrito
 - Provincia
 - Departamento
 - Table: adm_procede
 - Religion
 - default
 - Religion
 - Table: adm_religion
 - Sector
 - default
 - Sector
 - Table: adm_sectorad
 - Tiempo
 - default
 - Campania
 - FechaCampana
 - Table: adm_tiempo

The right pane shows the 'Table for Hierarchy' window, which displays the attributes and values for the selected dimension:

Attribute	Value
schema	public
name	adm_tiempo
alias	

Figura 19. Dimensiones Cubo Admisión.

El cubo Admisión está en función a un solo hecho (hecho admisión), el cual determinará: cantidad de ingresantes por religión y procedencia, ingresantes por carreras y procedencia, cantidad de ingresantes por carrera y sectores de admisión (sector de promotores), cantidad de ingresantes por semestres de la distintas carreras y zonas de procedencia y cantidad de ingresantes por semestre, modalidad y procedencia.

4.2 Validación de resultados

La validación de resultados se realizó en función a los reportes actuales de admisión comparados con los reportes arrojados por el DM. Para lo cual se hizo un comparativo en el tiempo que demora en obtener tal o cual reporte.

Tabla 4. Comparativo entre reportes actuales y nuestro modelo de IN.

	<i>Reportes actuales</i>	<i>Reportes Datamart Admisión</i>
Tiempo por consulta a la base de datos	1 minuto	15 segundos
Tiempo en hacer un reporte	3 horas	1 minuto
<i>Beneficios según tipos de sistemas</i>		
<i>Sistema Actual</i>	<i>Modelo de IN</i>	
Reportes no aptos para la toma de decisiones	Reportes según la necesidad del Departamento	
Realizados en función de indicadores generales	Realizados en función a indicadores específicos	

Nota: Elaboración propia.

4.3 Estipulación de los indicadores del negocio

Después de considerar los objetivos y analizar la base de datos operacional, se determinaron los siguientes indicadores:

Nuestro primer indicador se hizo con la finalidad de determinar la cantidad de ingresantes por religión y procedencia, el cual permite conocer a detalle, de que lugares

proviene más estudiantes sostenedores de obra. Para hacer este análisis nos centramos en provincias, ya que los distritos son más numerosos y sería de poco impacto.

El segundo ayuda a conocer la preferencia de los postulantes a las distintas carreras profesionales según lugar de procedencia, lo cual sería muy útil para captar estudiantes para alguna carrera en particular.

El indicador tercero, permite saber la cantidad de ingresantes por carrera profesional o sector de admisión, este sector de admisión tiene que ver con las sedes establecidas para los promotores. Este indicador ayudará a tomar la decisión de la existencia de un promotor en los sectores establecidos.

Nuestro cuarto indicador está basado en la cantidad de ingresantes por semestre académico, carrera profesional o zona de procedencia. Este indicador es importante para saber que campañas de admisión fueron más efectivas y en qué lugares tuvieron mayor impacto.

Nuestro quinto y último determina la cantidad de ingresantes por modalidad de ingreso, semestre o lugar de procedencia. El cual permitirá saber bajo que modalidad prefieren ingresar los postulantes a la universidad.

En las figuras 20 y 21 (fuente: elaboración propia, utilizando Pentaho Analysis Report) revelamos los resultados según indicadores:

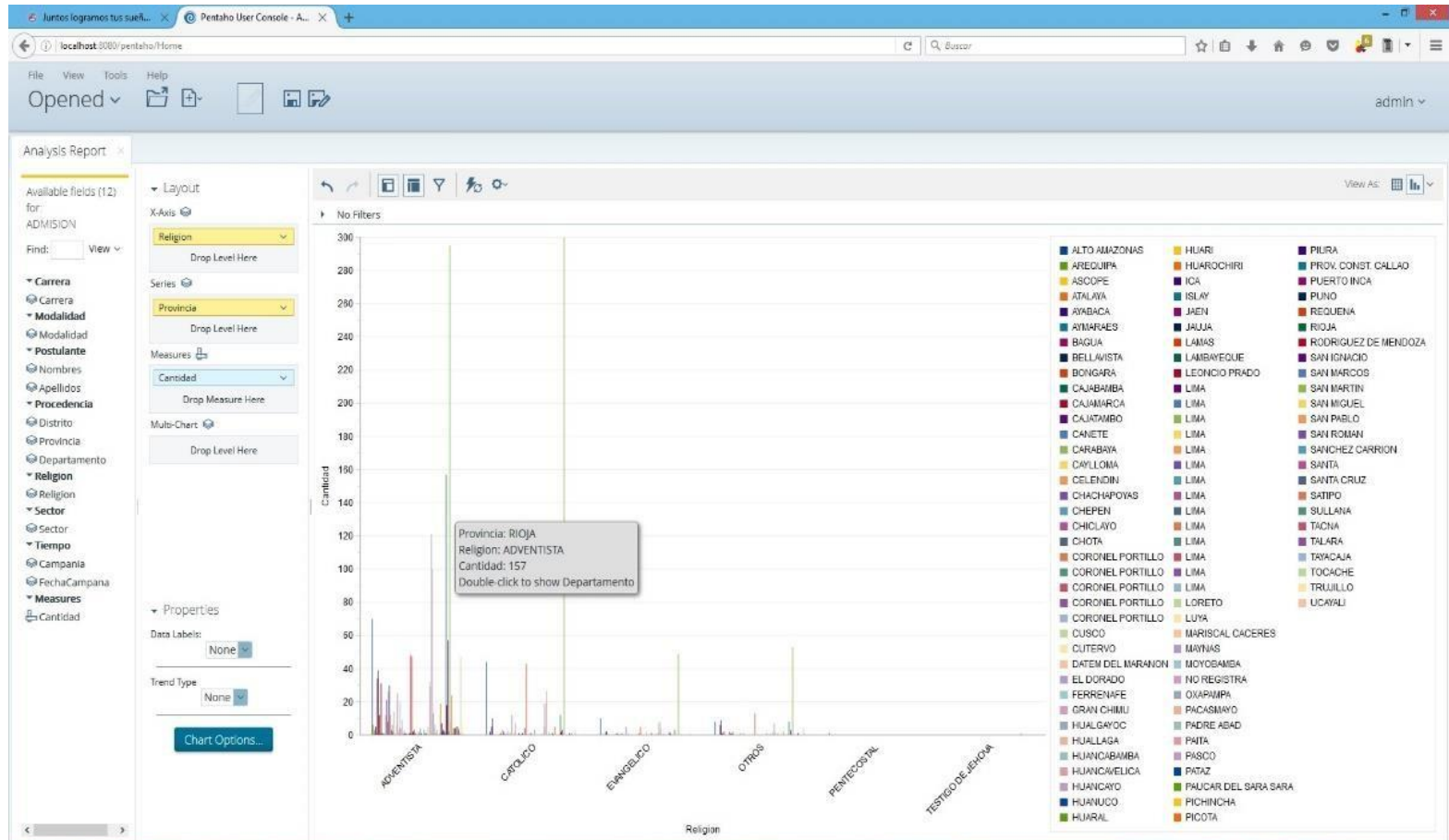


Figura 20. Cantidad de ingresantes por religión y lugar de procedencia.

Haciendo un breve análisis de los dos gráficos anteriores, se concluye que los alumnos sostenedores de obra y los católicos hacen la mayoría de ingresantes (más reportes, ver anexo 3). Esto efectivamente servirá para la toma de decisiones.

En las figuras 22 y 23 (fuente: elaboración propia), también se observa el resultado del modelo de IN. Los reportes serán posibles si y solo si, el modelo ETL está correctamente ejecutado, el diseño del cubo está importado o publicado en el servidor de Pentaho y si existe una métrica (medida), que en nuestro caso fue cantidad de ingresantes. La herramienta Analysis Report de Pentaho, fue muy útil en este proceso, porque permitió explorar y analizar toda la data.

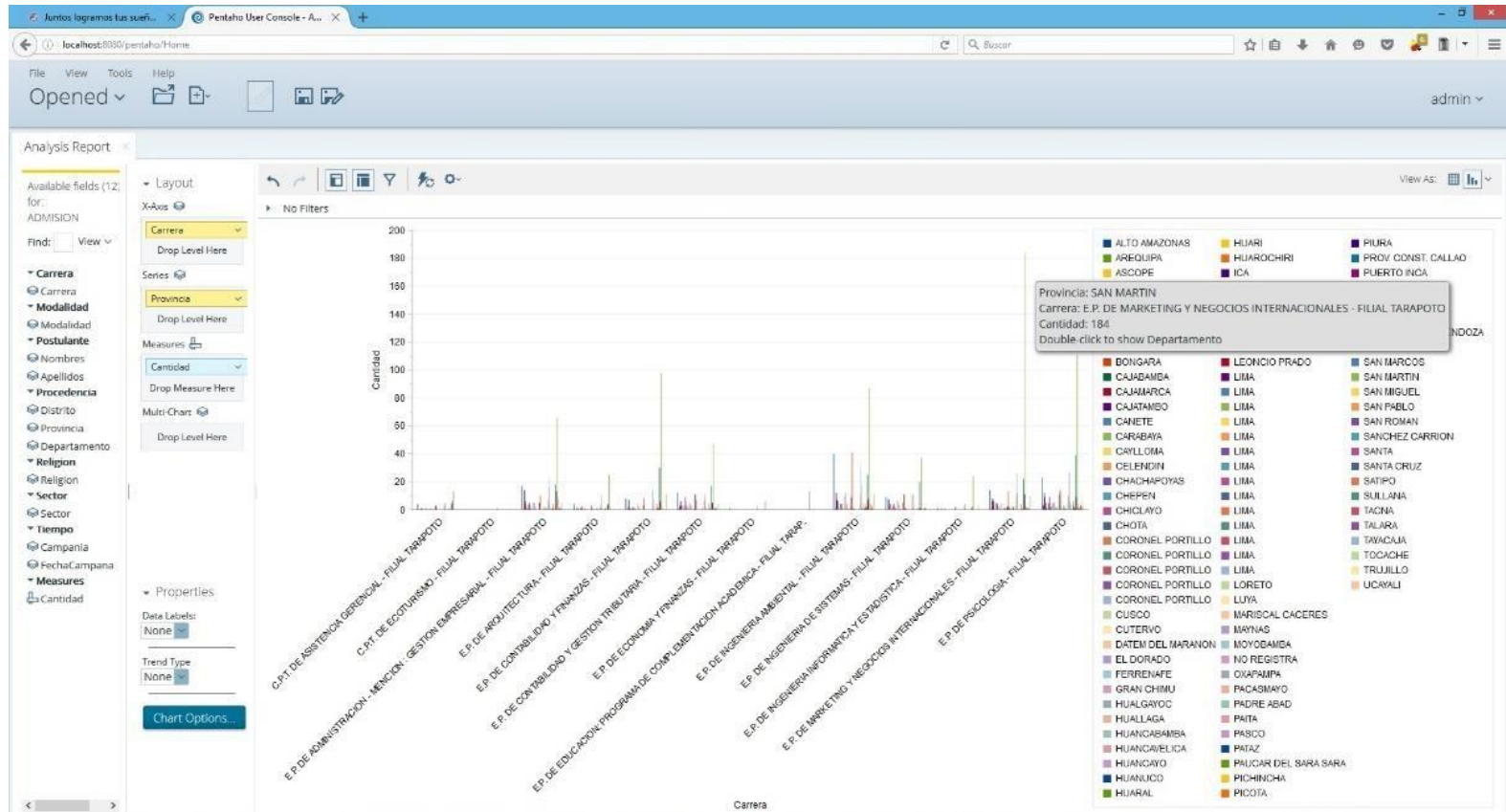


Figura 22. Cantidad de estudiantes por carreras y lugar de procedencia.

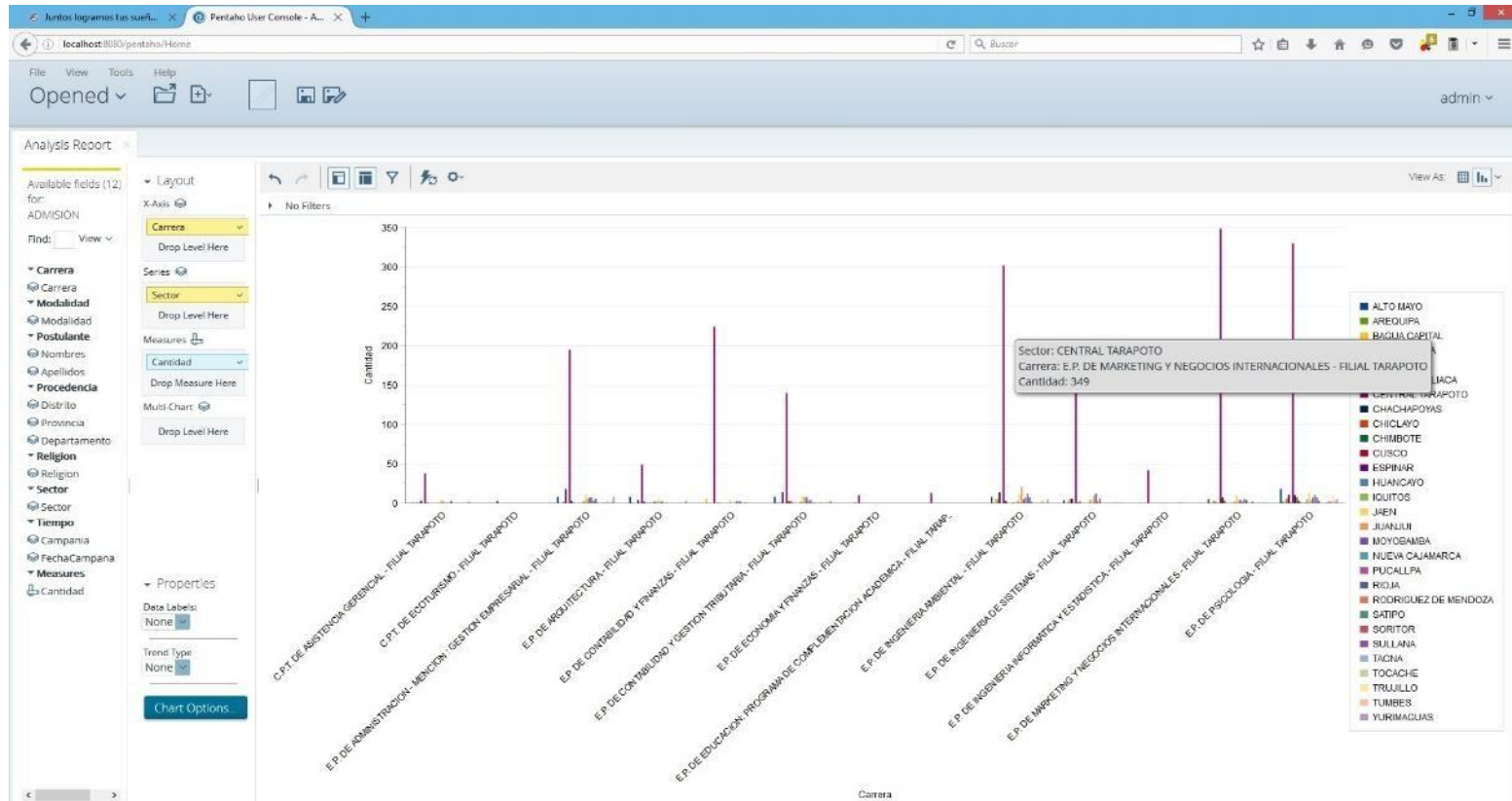


Figura 23. Estudiantes por carreras y sectores de admisión (sectores de promotores).

Las figuras 22 y 23, mostradas anteriormente, permiten conocer en todo existe mayor impacto del trabajo realizado por el Departamento de Admisión, y claramente se observa que es Central Tarapoto, provincia San Martín.

Es necesario mencionar que nuestro cubo cubre el tipo de reportes específicos y necesarios para el departamento.

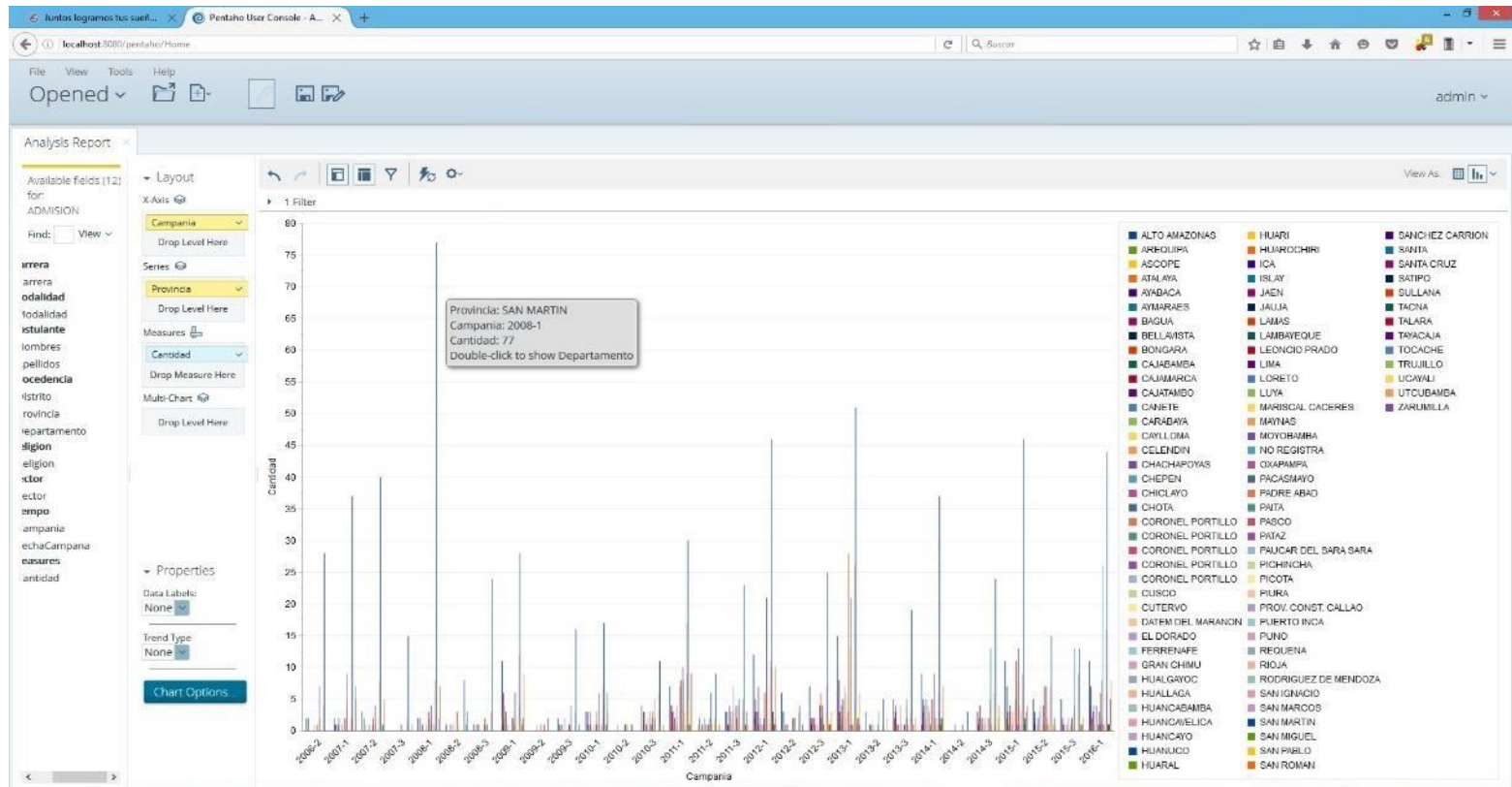


Figura 24. Cantidad de ingresantes según campaña y lugar de procedencia.

La figura 24 (fuente: elaboración propia), toma la dimensión tiempo. Por lo que queda demostrado que nuestra dimensión funcionó correctamente gracias al Schema Workbench y al Analysis Report.

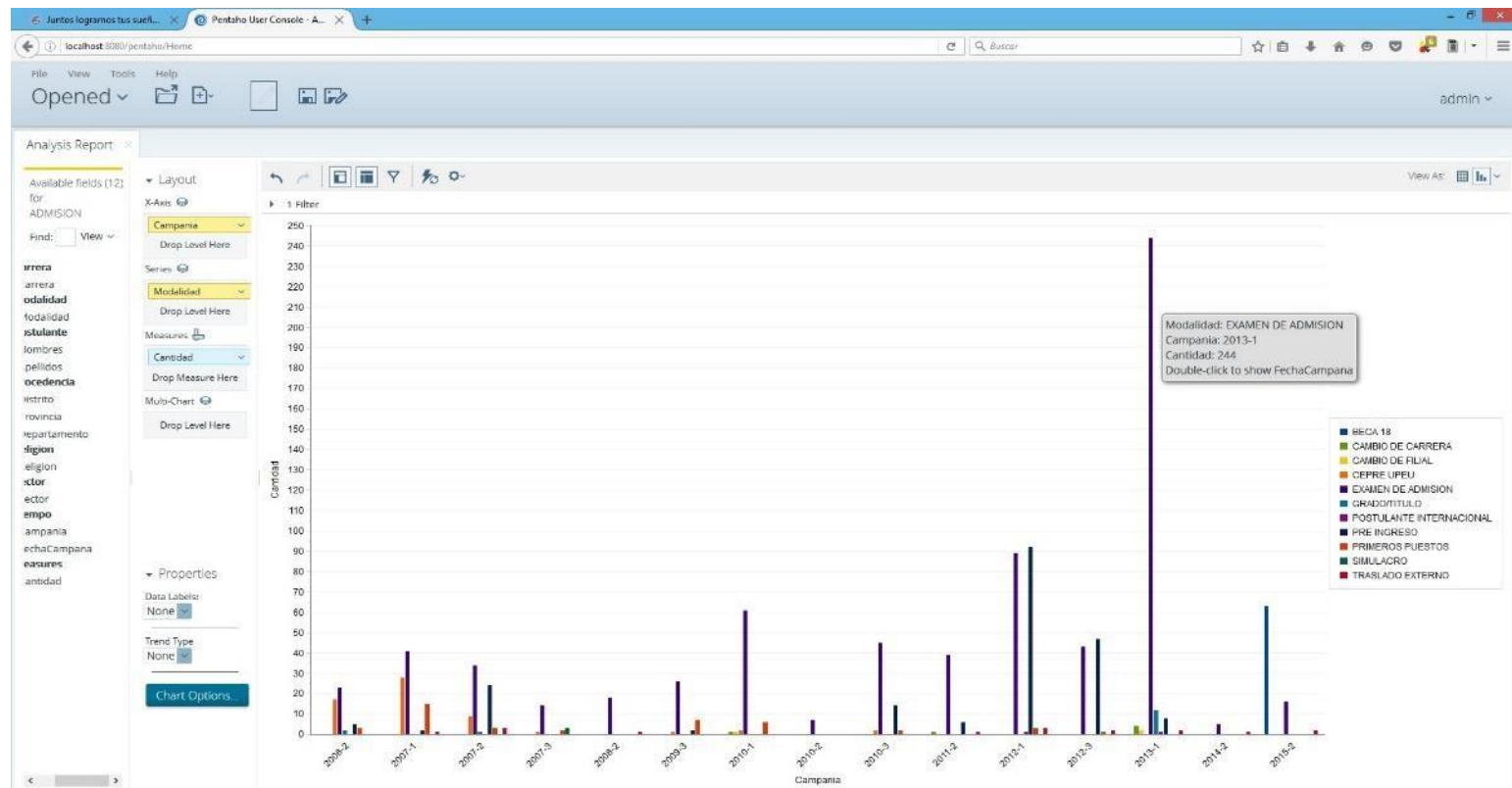


Figura 25. Cantidad de ingresantes según campaña y modalidad de ingreso.

La figura 25 (fuente: elaboración propia), muestra los ingresantes de cada campaña según modalidad de ingreso.

El modelo de IN, que realizamos permite a través del Analysis Report exportar todos los cuadros estadísticos que deseemos en los formatos: .xlsx, .pdf o .xml. Observemos la figura 26 (fuente: elaboración propia).

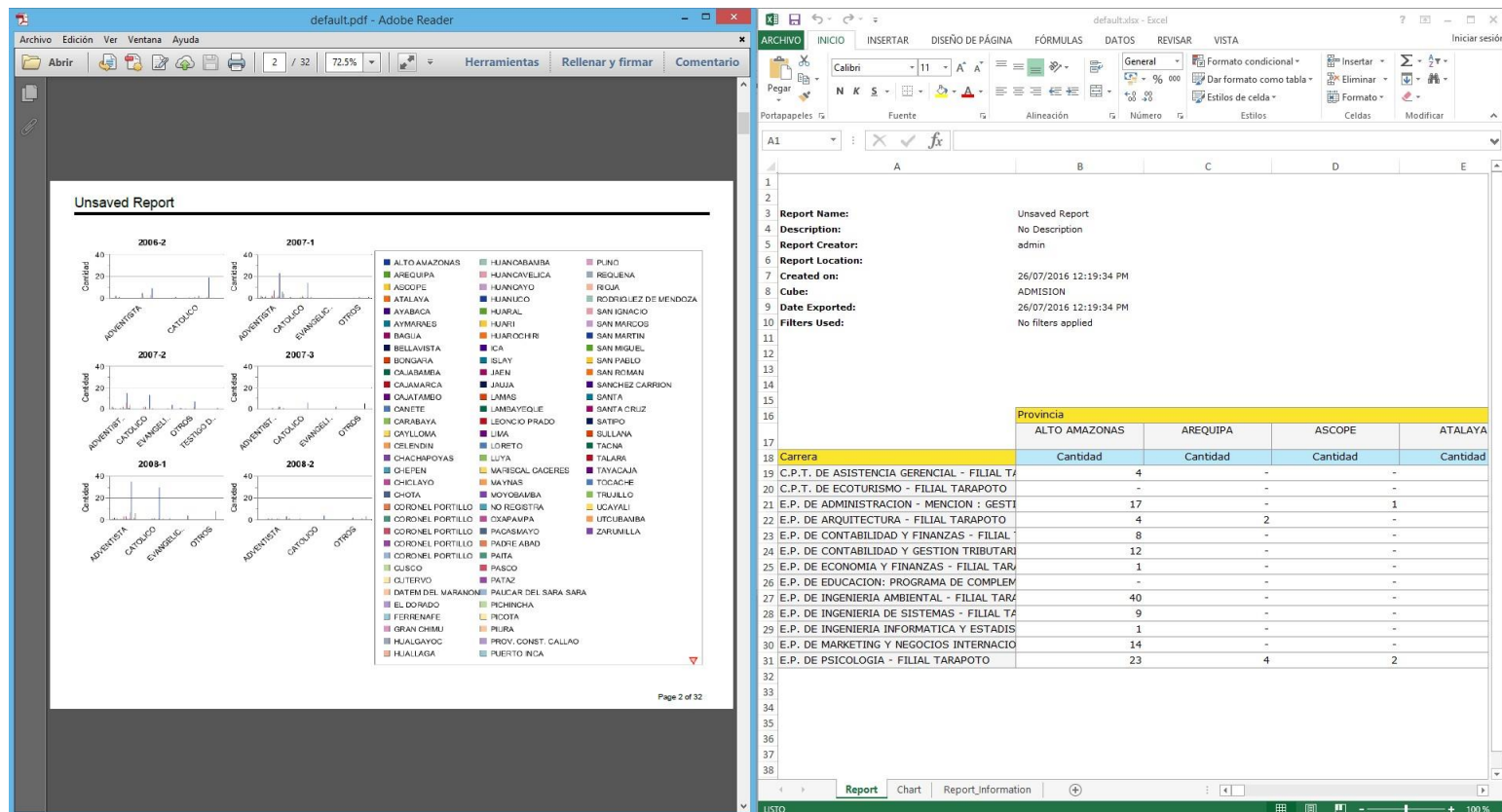


Figura 26. Información estadística en formato .xlsx y .pdf.

4.4 Discusiones frente a otras investigaciones

Caso 1: Fuentes & Valdivia (2010) describen una investigación realizada en la Universidad de Tarapacá, Arica, Chile, que incorporan la IN como recurso para crear una plataforma que permita la visualización y el análisis de información relacionados con el proceso de admisión y matrícula. Como resultados significativos para admisión se obtuvo gráficos comparativos (figura 27) de la cantidad de matrículas por sedes desde el 2006 hasta 2009; otro resultado fue la comparación de matrículas por facultades, también se mostró un tipo de reporte indicando la procedencia de los alumnos, de acuerdo a su etnia, dependiendo en gran manera de las regiones y procesos previamente seleccionados. También se consideró puntajes bajos, medios y altos según resultados de ambos procesos. Por lo tanto el Datamart (DM) mejoró el proceso de toma de decisiones y sirvió como una única fuente de información para el análisis de las actividades correspondientes al área de admisión. (p.9)

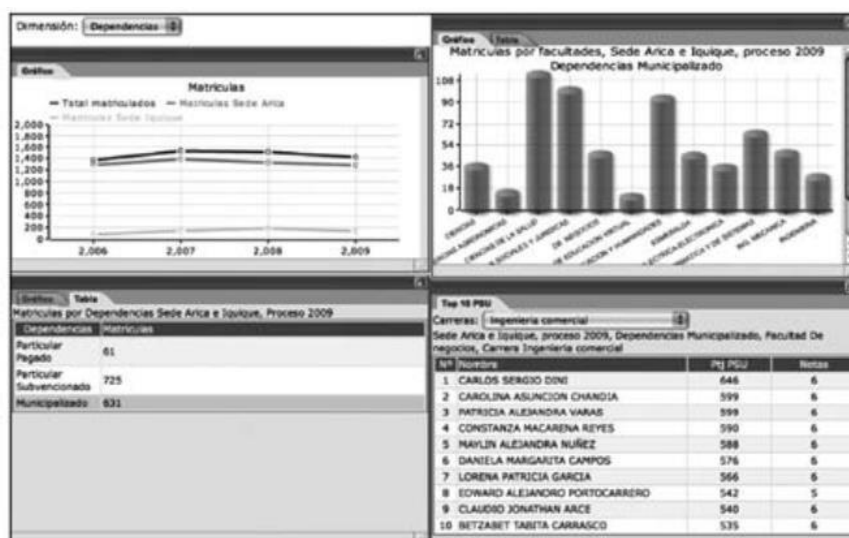


Figura 27. Cantidad de matrículas y top 10 de puntajes.

Discusión referente al caso 1: Con esta investigación queda demostrado la importancia de analizar la data histórica de una institución. Siempre se ha dicho que una solución de IN se realiza a una entidad comercial, sin embargo, conforme van pasando los años se hace más necesario y útil este tipo de soluciones para ayudar en la toma de decisiones. Con ese fin, se realizó este proyecto. Pero lo que no compartimos con este proyecto es que no dan importancia a la metodología ni a la herramienta tecnológica que utilizaron, las cuales son muy necesarias para no dejar a sospechas. Por lo que nosotros motivamos a todos los desarrolladores de proyectos de IN, a utilizar una metodología y software libre, salvo se haga soluciones para entidades del estado o particulares que demandan del uso de licencia. Por lo tanto si queremos tener reportes más exactos, debemos obtener a través del análisis de la data histórica, en otras palabras con soluciones de IN.

Caso 2: Zambrano, Rojas, Carvajal, & Acuña (2011) presentan el análisis del rendimiento académico estudiantil usando DM para analizar información de sistemas que producen varias operaciones que son almacenadas en bases de datos transaccionales, con el fin de predecir en base a la información histórica el desempeño académico de los estudiantes y las asignaturas aprobadas. Gracias a la arquitectura descrita se obtuvo la salida deseada, es decir, se puede obtener a través de gráficos desde la cantidad de asignaturas aprobadas por alumno, en un semestre determinado (figura 28). Es preciso agregar que el beneficio de contar con un DM radicó en la facilidad de análisis exploratorio de los datos para la creación de reportes, en tanto esto mejoró la asertividad en la predicción del rendimiento académico.

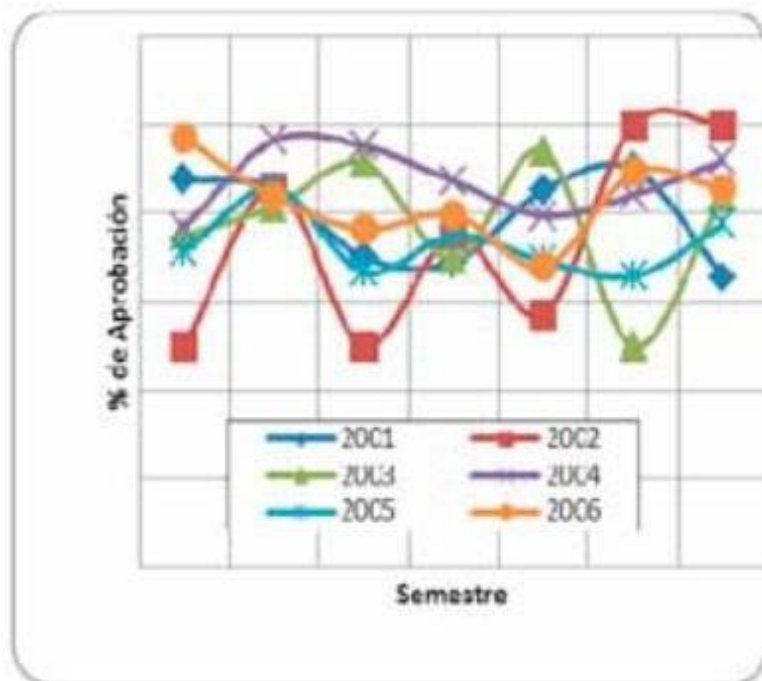


Figura 28. Porcentaje de Aprobación por Semestre en asignaturas de la carrera.

Discusión respecto al caso 2: Como se puede observar en la figura anterior, si todo el proceso de ETL de una solución de IN está bien hecho, entonces generar gráficos no será un problema, sino una manera de interpretación sencilla y entendible para el usuario final. Sin embargo los gráficos no tienen que ser tan confusos, de preferencia se deben usar gráficos de barras que son más ordenados o en todo caso gráficos circulares, salvo que los datos sean las descripciones de los datos sean inmensos.

En la figura 29 elaborada en nuestra investigación, tenemos un modelo de reporte de nuestro proyecto, sobre las campañas realizadas por admisión, según zonas de procedencia, utilizando gráfico de barras, el cual es muy sencillo de interpretar.

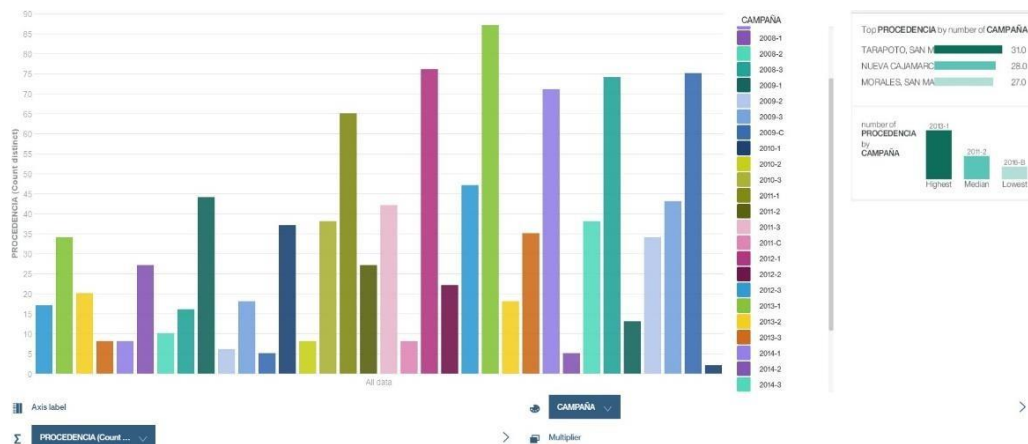


Figura 29. Gráfico de barras según campaña y procedencia.

Caso 3: Dall’Orto & Wu (2006) presentan un modelo de extracción para el desarrollo de un proyecto de IN, utilizando Framework Struts y JSF, que según su criterio están hechos de manera que pueden ser manipulados fácilmente por un programador. A la vez utilizaron la metodología tradicional RUP, respetando sus fases: concepción, elaboración, construcción y transición.

En la figura 30 propia de los autores, se puede observar toda la parte de programada y configurada para la conexión a la base de datos:

Configurar y Conectar Fuente de Datos	
ID Prueba:	BE01
Objetivo :	Configurar una fuente de datos mediante el ingreso de datos válidos para una base de datos SQL Server.
Precondición:	El usuario ha ingresado al sistema. El usuario ha abierto un paquete.
Clases :	JSFServlet, BEConfigurarBDBean, BEGestorFuenteDatos, BEControlFuenteDatos, BEFuenteDatos, IdFactory, Empresa, Proyecto, BEPaquete.
Proceso:	<ol style="list-style-type: none"> 1. En el menú seleccione la opción "Paquete" y "Agregar Fuente de Datos". 2. Seleccione "Conexión a Base de datos". 3. Ingrese en el campo "Nombre" el valor "Fuente de prueba". 4. Ingrese en el campo "Descripción" el valor "Fuente creada para caso de prueba". 5. Elija del combobox de "Tipo" el valor "SQL Server" de la lista. 6. Ingrese en el campo "Conexión" el valor "Inti". 7. Ingrese en el campo "Base Datos" el valor "Bipucp". 8. Ingrese en el campo "Usuario" el valor "biextrausr". 9. Ingrese en el campo "Contraseña" el valor "2006biusr". 10. Seleccione la opción "Grabar".
Resultado Esperado:	Se regresa al área de trabajo inicial y se muestra la estructura de la Fuente de Datos.

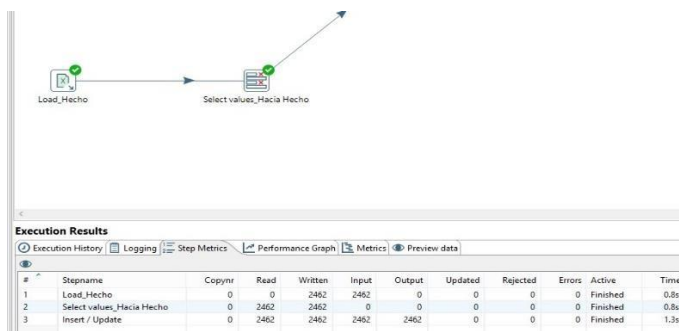
Figura 30. Caso de prueba para funcionalidad tipo comunicación.

La figura 31 también elaborada por los mismos autores, presenta el proceso de extracción realizada por los mencionados investigadores:

Crear flujo de transformación (continúa)	
ID Prueba:	BE 13
Objetivo :	Crear el flujo de transformación mediante la selección de datos válidos.
Precondición:	El usuario ha ingresado al sistema. El usuario ha abierto un job. Las fuentes de datos a utilizar han sido configuradas correctamente.
Clases :	AppletExtraccion, DiagramaDibujoo, BEControlDibXML, DiagramaDibujooBean, AppletServlet, BEControlPaquete, Empresa, Proyecto, BEJob, BEPaquete,
Proceso:	<ol style="list-style-type: none"> 1. Seleccione del árbol de navegación de objetos el paquete con nombre "Paquete Prueba". 2. Elija de la lista de tablas de la Fuente de Datos con nombre "Fuente Prueba" la tabla "Persona". 3. Click derecho sobre la tabla y seleccione la opción "Agregar tabla al flujo". 4. Elija de la lista de tablas de la Fuente de Datos con nombre "Fuente Destino - DW" la tabla "Fact_Persona". 5. Click derecho sobre la tabla y seleccione la opción "Agregar tabla al flujo". 6. Presione el botón que corresponde al componente "Proceso" 7. Presione en el punto donde se insertará el componente.
Crear flujo de transformación (continuación)	
Proceso:	<ol style="list-style-type: none"> 8. Presione el botón que corresponde al componente "Conector" 9. Seleccione la tabla "Persona" para indicar el origen. 10. Seleccione el componente "Proceso" para indicar el destino. 11. Presione el botón que corresponde al componente "Conector" 12. Seleccione el componente "Proceso" para indicar el origen. 13. Seleccione la tabla "Fact_Persona" para indicar el destino. 14. Seleccione la opción "Grabar".
Resultado Esperado:	Se graba exitosamente el flujo y se muestra el área de trabajo sin ningún mensaje de error.

Figura 31. Caso de prueba para funcionalidad tipo extracción.

Discusión del caso 3: Esta investigación no utilizó una metodología para mega proyectos de desarrollo de software, por lo que no comparto la idea de su proceso de extracción, ya que utiliza frameworks que solo son entendidos por programadores y utilizados por ellos, lo cual solo sería de beneficio para ellos y no una herramienta de ayuda para otros usuarios. La figura 32 propia de nuestro trabajo, muestra un proceso de extracción sencillo, utilizando software libre y especialmente para este tipo de procesos, en un tiempo muy bueno.



#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time
1	Load_Hecho	0	0	2462	2462	0	0	0	0	Finished	0.8s
2	Select values_Hacia Hecho	0	2462	2462	0	0	0	0	0	Finished	0.8s
3	Insert / Update	0	2462	2462	2462	2462	0	0	0	Finished	1.3s

Figura 32. Proceso de extracción utilizando Spoon.

Caso 4: Falcón, Palma, & Celi (2006), diseñaron una herramienta de IN capaz de explotar la data histórica, tomando en cuenta nuevamente la metodología tradicional RUP y los Frameworks como: JPivot, Mondrian, JSF y Apache como servidor; soportando a los motores de base de datos como: Microsoft SQL Server, Oracle, MySQL, Microsoft Acces. Esta herramienta permitió realizar cubos y por ende reportes. El proceso del cubo se puede apreciar en la figura 33 realizada por los investigadores.

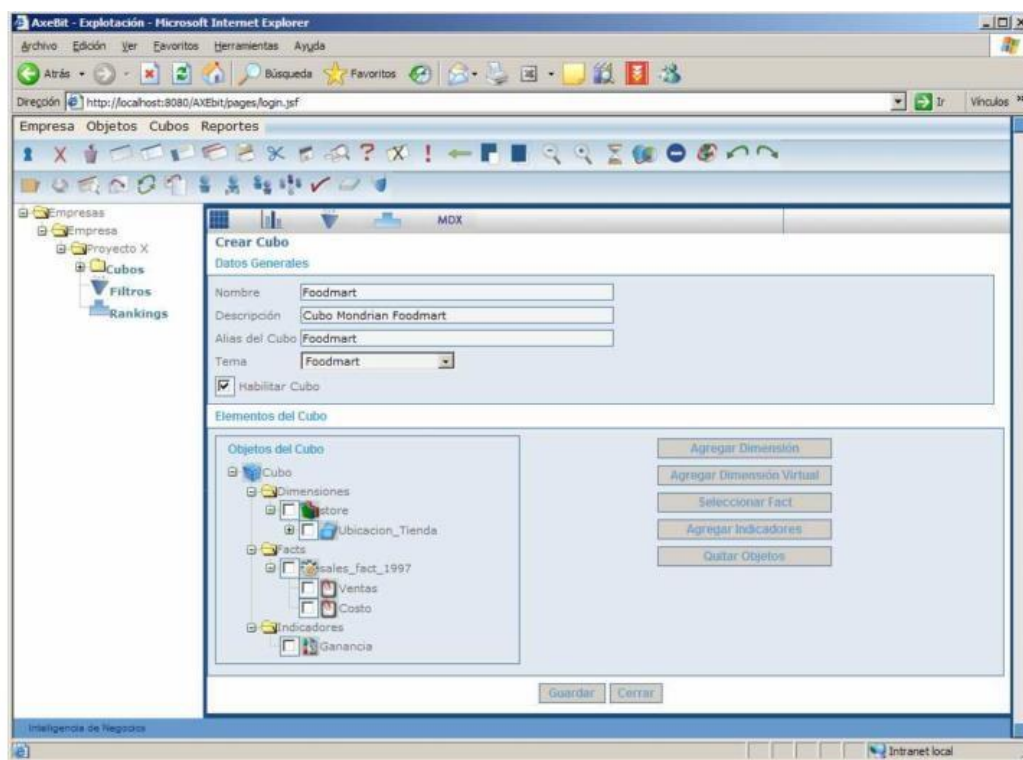


Figura 33. Creación de cubo y sus dimensiones.

Discusión del caso 4: Respecto a este caso, se distintas herramientas para hacer un cubo, aunque por la antigüedad del proyecto se entiende. Sin embargo, como investigador estoy en contra de esta manera de trabajo, ya que existe una herramienta capaz de hacer todo este proceso que es el Schema Workbench de Pentaho, que es libre y no se le necesita añadir nada, además soporta los principales motores de bases de datos. Dicha evidencia lo podemos observar a continuación en la figura 34 que es de elaboración propia.

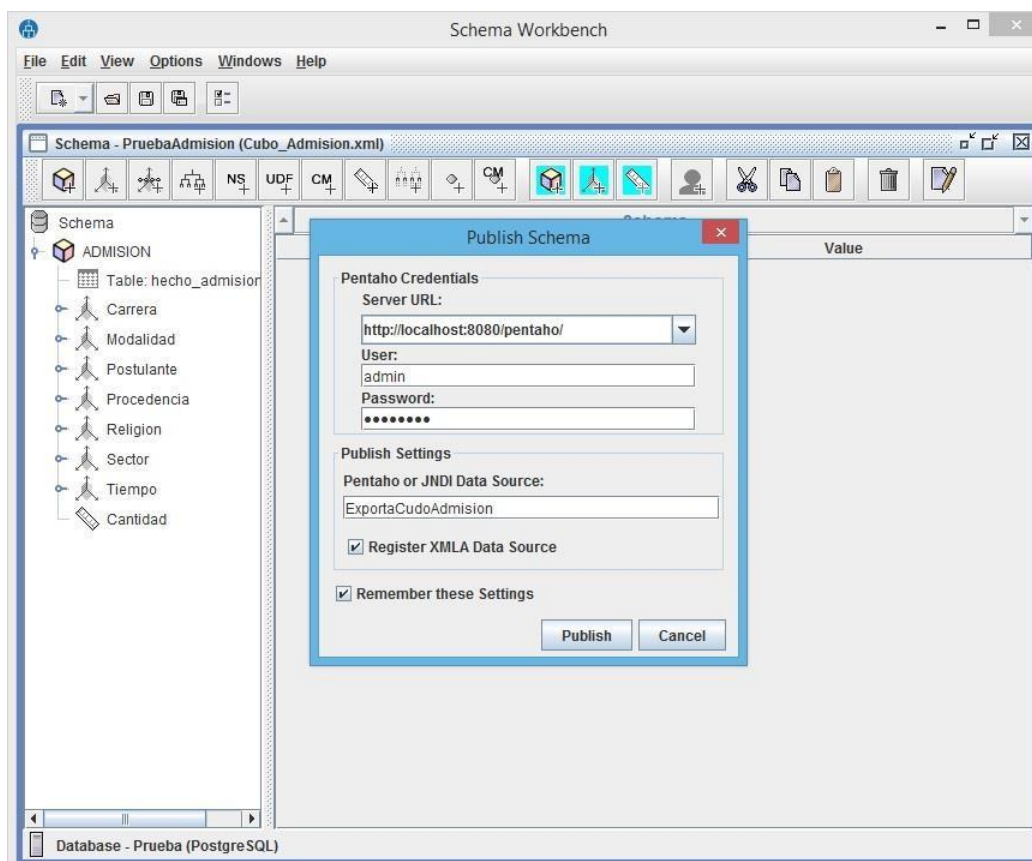
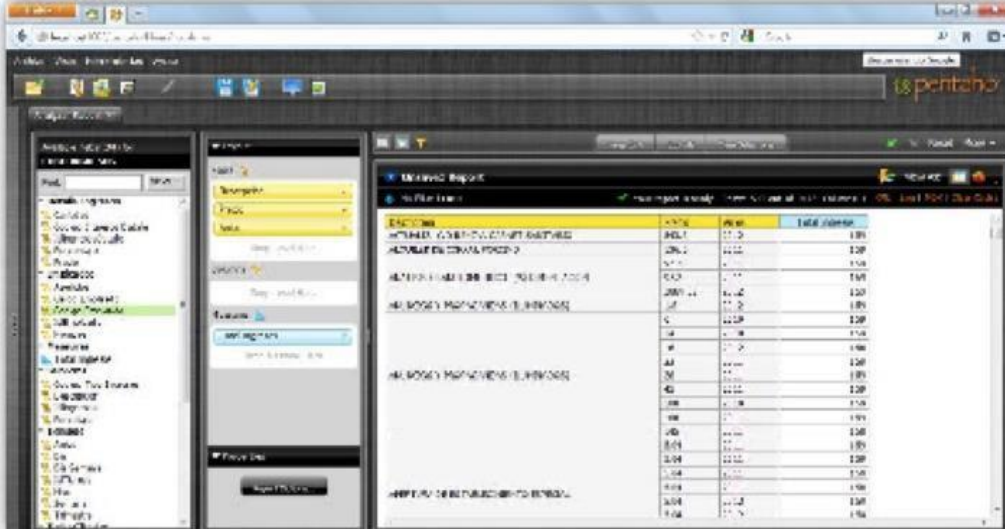


Figura 34. Diseño de Cubo utilizando Schema Workbench.

Caso 5: Guillén (2012) desarrolló un DM para mejorar la toma de decisiones en el área de Tesorería de la Municipalidad Provincial de Cajamarca (MPC), para lo cual utilizó Pentaho Business Analytics, la metodología Ralph Kimball y motor de base de datos MySQL, en todo el proceso de construcción. Como resultado obtuvo reportes importantes y específicos para el área.

La figura 35 muestra un tipo de reporte utilizando Business Analytics de Pentaho versión 5.0.



The screenshot shows the Pentaho Business Analytics interface. On the left, there is a navigation tree with folders like 'Data Sources', 'Views', and 'Reports'. The main window displays a report titled 'Delayed Report' with a table of data. The table has columns for 'CATEGORIA', 'ANIO', 'MUNICIPIO', and 'VALOR'. The data is grouped by 'ANIO' and 'MUNICIPIO'.

CATEGORIA	ANIO	MUNICIPIO	VALOR
RENTAS	2011	01	1.81
RENTAS	2011	02	1.28
RENTAS	2011	03	1.54
RENTAS	2011	04	1.20
RENTAS	2011	05	1.81
RENTAS	2011	06	1.28
RENTAS	2011	07	1.54
RENTAS	2011	08	1.20
RENTAS	2011	09	1.81
RENTAS	2011	10	1.28
RENTAS	2011	11	1.54
RENTAS	2011	12	1.20
RENTAS	2011	13	1.81
RENTAS	2011	14	1.28
RENTAS	2011	15	1.54
RENTAS	2011	16	1.20
RENTAS	2011	17	1.81
RENTAS	2011	18	1.28
RENTAS	2011	19	1.54
RENTAS	2011	20	1.20
RENTAS	2011	21	1.81
RENTAS	2011	22	1.28
RENTAS	2011	23	1.54
RENTAS	2011	24	1.20
RENTAS	2011	25	1.81
RENTAS	2011	26	1.28
RENTAS	2011	27	1.54
RENTAS	2011	28	1.20
RENTAS	2011	29	1.81
RENTAS	2011	30	1.28
RENTAS	2011	31	1.54
RENTAS	2011	32	1.20
RENTAS	2011	33	1.81
RENTAS	2011	34	1.28
RENTAS	2011	35	1.54
RENTAS	2011	36	1.20
RENTAS	2011	37	1.81
RENTAS	2011	38	1.28
RENTAS	2011	39	1.54
RENTAS	2011	40	1.20
RENTAS	2011	41	1.81
RENTAS	2011	42	1.28
RENTAS	2011	43	1.54
RENTAS	2011	44	1.20
RENTAS	2011	45	1.81
RENTAS	2011	46	1.28
RENTAS	2011	47	1.54
RENTAS	2011	48	1.20
RENTAS	2011	49	1.81
RENTAS	2011	50	1.28
RENTAS	2011	51	1.54
RENTAS	2011	52	1.20
RENTAS	2011	53	1.81
RENTAS	2011	54	1.28
RENTAS	2011	55	1.54
RENTAS	2011	56	1.20
RENTAS	2011	57	1.81
RENTAS	2011	58	1.28
RENTAS	2011	59	1.54
RENTAS	2011	60	1.20
RENTAS	2011	61	1.81
RENTAS	2011	62	1.28
RENTAS	2011	63	1.54
RENTAS	2011	64	1.20
RENTAS	2011	65	1.81
RENTAS	2011	66	1.28
RENTAS	2011	67	1.54
RENTAS	2011	68	1.20
RENTAS	2011	69	1.81
RENTAS	2011	70	1.28
RENTAS	2011	71	1.54
RENTAS	2011	72	1.20
RENTAS	2011	73	1.81
RENTAS	2011	74	1.28
RENTAS	2011	75	1.54
RENTAS	2011	76	1.20
RENTAS	2011	77	1.81
RENTAS	2011	78	1.28
RENTAS	2011	79	1.54
RENTAS	2011	80	1.20
RENTAS	2011	81	1.81
RENTAS	2011	82	1.28
RENTAS	2011	83	1.54
RENTAS	2011	84	1.20
RENTAS	2011	85	1.81
RENTAS	2011	86	1.28
RENTAS	2011	87	1.54
RENTAS	2011	88	1.20
RENTAS	2011	89	1.81
RENTAS	2011	90	1.28
RENTAS	2011	91	1.54
RENTAS	2011	92	1.20
RENTAS	2011	93	1.81
RENTAS	2011	94	1.28
RENTAS	2011	95	1.54
RENTAS	2011	96	1.20
RENTAS	2011	97	1.81
RENTAS	2011	98	1.28
RENTAS	2011	99	1.54
RENTAS	2011	100	1.20

Figura 35. Reporte del pago por servicios por año de la MPC.

Discusión del caso 5: Gracias al desarrollo de este proyecto pude comprender la importancia de utilizar una metodología sencilla para soluciones de IN, tal es el caso de HEFESTO vs. 2.0 y la facilidad que Pentaho en su última versión, que permite hacer todo el desarrollo de un DM o DW. En las figuras 36 y 37 de elaboración propia, mostramos la metodología descrita en un cronograma de actividades y un tipo de reporte para el área de admisión.

Nombre de tarea	Duración		Duración
Desarrollo del Proyecto (Data Mart - Admisión UPeU-FT)	195 días		
Levantamiento de Información	27 días	Modelado Multidimensional	16 días
Evaluar el Área de Admisión	20 días	Elegir Tipo de Modelo Lógico	2 días
Descripción del Problema a Investigar	7 días	Establecer Tabla de Dimensiones	6 días
Fin Fase I	0 días	Establecer Tabla de Hechos	4 días
Identificar Requerimientos	15 días	Unir Dimensiones hacia el Hecho	1 día
Establecer Objetivos	5 días	Comprobar Resultados	3 días
Aplicar Técnicas para la Identificación de Métricas e Indicadores	5 días	Fin de Fase V	0 días
Justificación de la Investigación	5 días	Integración de los Datos	6 días
Fin Fase II	0 días	Establecer Correspondencias	3 días
Análisis de Requerimientos	18 días	Cargar Data al Modelo Establecido	3 días
Identificar Medidas	7 días	Fin de Fase VI	0 días
Identificar Indicadores	5 días	Procesamiento de Información	34 días
Definir Niveles de Granularidad	2 días	Aplicar Herramientas de IN	20 días
Definir Modelo Conceptual	4 días	Procesar data	7 días
Fin de Fase III	0 días	Comprobar Resultados	7 días
Identificación de los Orígenes de Datos	33 días	Fin de Fase VII	0 días
Entender el Modelado Entidad Relación	15 días	Explotar Información	30 días
Normalizar y/o De-normalizar los datos	15 días	Desarrollo de Reportes de Acuerdo a Indicadores	30 días
Establecer Modelo Conceptual Ampliado	3 días	Fin de Fase VIII	0 días
Fin de Fase IV	0 días		

Figura 36. Tareas de nuestra metodología, basada en HEFESTO 2.0.

Report Name:	Unsuaved Report
Description:	No Description
Report Creator:	admin
Created on:	26/07/2016 12:19:34 PM
Cuba:	ADMISSION
Date Exported:	26/07/2016 12:19:34 PM
Filters Used:	No filters applied

Carrera	Provincia											Cant
	ALTO AMAZONAS	AREQUIPA	ASCOPE	ATALAYA	AYABACA	AYMARAES	BAGUA	BELLAVISTA	BONGARA	CAJAMA	CAJAMA	
C.P.T. DE ASISTENCIA GERENCIAL - FILIAL TAJ	4	-	-	-	-	-	-	-	-	-	-	1
C.P.T. DE ECOTURISMO - FILIAL TARAPOTO	-	-	-	-	-	-	-	-	-	-	-	-
E.P. DE ADMINISTRACION - MICHON - SESTE	17	-	1	-	-	-	-	4	14	-	-	3
E.P. DE ARQUITECTURA - FILIAL TARAPOTO	4	2	-	-	-	-	-	-	1	-	-	-
E.P. DE CONTABILIDAD Y FINANZAS - FILIAL T	8	-	-	-	-	-	2	7	-	-	1	-
E.P. DE CONTABILIDAD Y GESTION TRIBUTARI	12	-	-	1	2	-	3	6	-	-	-	-
E.P. DE ECONOMIA Y FINANZAS - FILIAL TAR	1	-	-	-	-	-	-	-	-	-	-	-
E.P. DE EDUCACION: PROGRAMA DE COMPLEM	-	-	-	-	-	-	-	-	-	-	-	-
E.P. DE INGENIERIA AMBIENTAL - FILIAL TARJ	40	-	-	-	1	-	12	7	-	-	1	-
E.P. DE INGENIERIA DE SISTEMAS - FILIAL TFI	9	-	-	-	-	-	8	-	-	-	3	-
E.P. DE INGENIERIA INFORMATICA Y ESTADIS	1	-	-	-	-	-	-	1	-	-	-	-
E.P. DE MARKETING Y NEGOCIOS INTERNACIO	14	-	-	-	1	-	6	8	-	-	2	-
E.P. DE PSICOLOGIA - FILIAL TARAPOTO	23	4	2	2	3	1	12	9	-	-	2	-

Figura 37. Reportes utilizando Report Designer de Pentaho.

Caso 6: Chavez (2013) analiza, diseña e implementa una solución de inteligencia de negocios orientada a controlar los procesos de generación y emisión del DNI en el RENIEC. En su investigación detalla la metodología que siguió y como fue el desarrollo de la solución de IN. También menciona como realizó el modelado de procesos, la arquitectura lógica, los procesos ETL y como realizó los diferentes reportes gráficos. Para la construcción de su solución utilizó la Suite de Pentaho.

En la figura 38 elaborada por el mismo autor, se presenta el modelo dimensional tipo estrella que propuso para dar la solución a la exigencias del usuario.

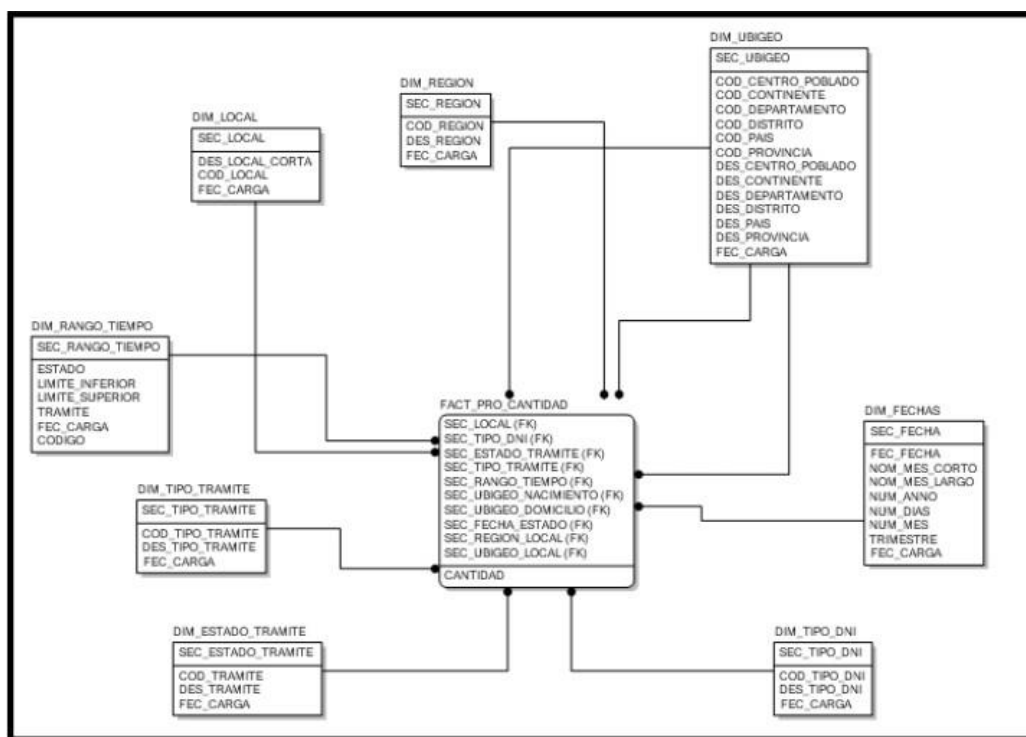


Figura 38. Diseño lógico del esquema de producción.

Discusión del caso 6: Esta investigación realizada por Chávez, tiene una peculiaridad, que todo el desarrollo de su solución de IN, utiliza la Suite de Pentaho, apoyado de otras herramientas para estructurar su idea del modelo dimensional de base de datos para su DM. En una primera impresión nosotros lo hicimos en hoja, ya que este modelo se da bajo una primera impresión, el cual puede ser modificado en la siguiente etapa al pulir el modelo bajo indicadores, y dimensionados por los mismos.

En la figura 39 de elaboración propia, muestra nuestra primera propuesta de modelo lógico para nuestro DM, el cual tiene una peculiaridad, está hecho en hoja, ya que este se dio al analizar la base de datos origen, y compararlos con los requerimientos hechos por el cliente.

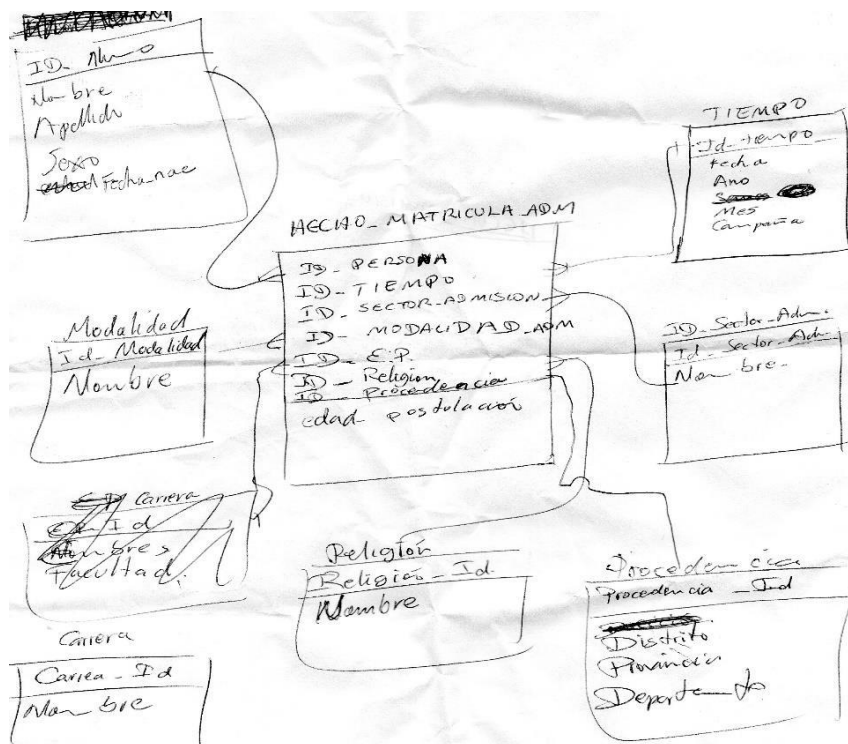


Figura 39. Idea de Modelo lógico para el DM Admisión.

Caso 7: Cruz (2003) resalta la importancia de la tecnología Data Warehouse y el Datamart en instituciones educativas. El caso de estudio que presenta Cruz, es al Instituto Hidalguense de Educación Media Superior y Superior (IHEMSyS). Lo curioso de esta investigación es que utiliza dos metodologías para DW que son poco utilizadas: Metodología Big Bang y Metodología Rapid Weurehousing, las cuales se concentran muchos en evidencias y no en resultados. La figura 40 propia de la autora, nos muestra en resumen la metodología de desarrollo para su DW.

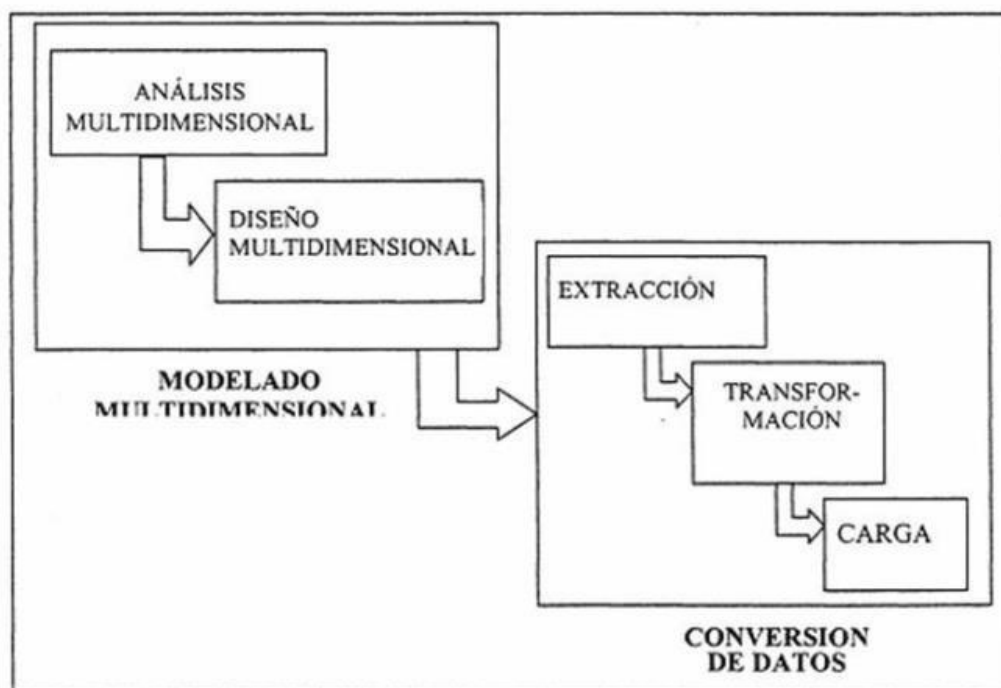


Figura 40. Proceso de desarrollo del DW.

Discusión del caso 7: En relación a la figura anterior, el proceso de su metodología nos parece más para proyectos de clase, o primeras prácticas con IN, pero no para acompañar el desarrollo de una solución que permita tomar decisiones. Además, esta metodología se centra en el modelado multidimensional y en la conversión de datos, más no en la información que se pueda obtener de este proceso, por lo que sería necesario aplicar otra metodología que acompañe en el proceso de explotación de la data.

A la par está nuestra metodología que tuvo como base HEFESTO 2.0 que describe cada una de las actividades y aconseja que se debe y no debe hacer en cada fase de construcción de una solución de IN. Es necesario entender que recomendamos nuestra metodología para proyectos educativos y no para comerciales, ya que los comerciales tienden a manejar más números (dinero) que es lo que mueve al negocio.

Caso 8: Rojas (2014) implementó un DM como solución para la toma de decisiones en el Departamento de Finanzas de la Contraloría General de la República, bajo la metodología Ralph Kimball. Dicho sea de paso Rojas también utiliza el modelo dimensional tipo estrella, sin embargo, utilizó SQL Server Business Intelligence Development 2008, para el desarrollo de su solución de IN.

En las figuras 41 y 42 diseñadas por el propio investigador, podemos observar su modelo dimensional y su diseño ETL respectivamente.

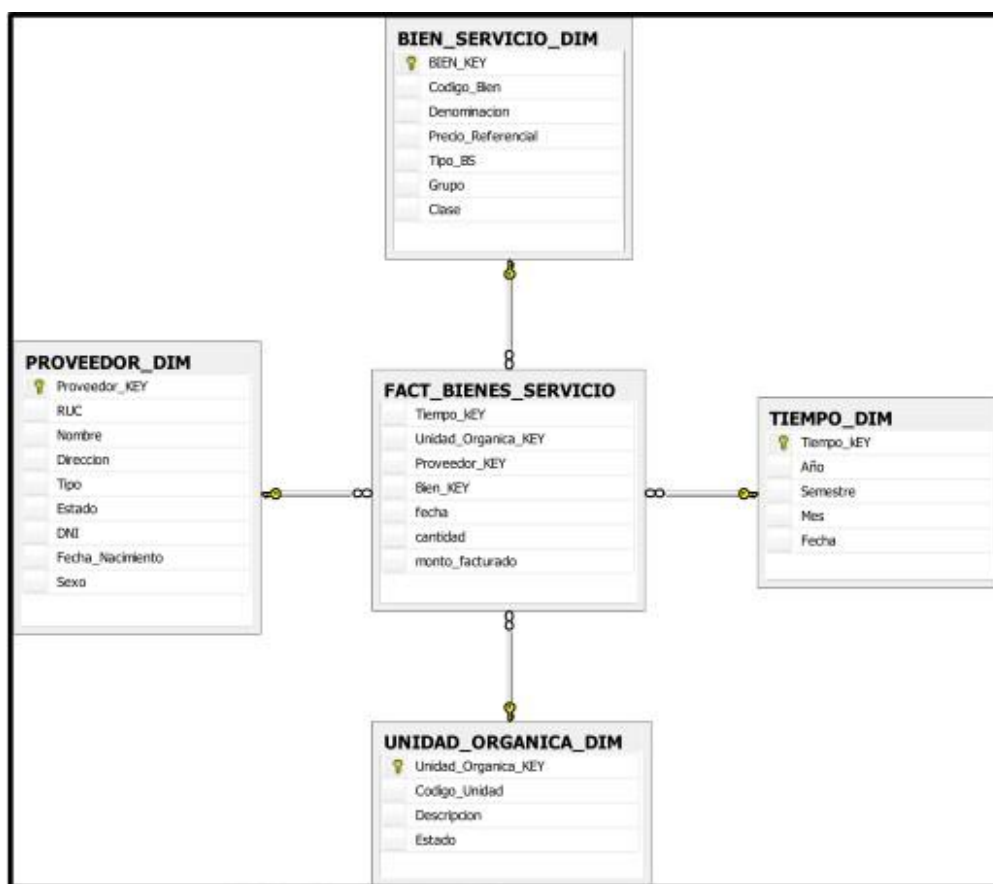


Figura 41. Diseño físico del Datamart.

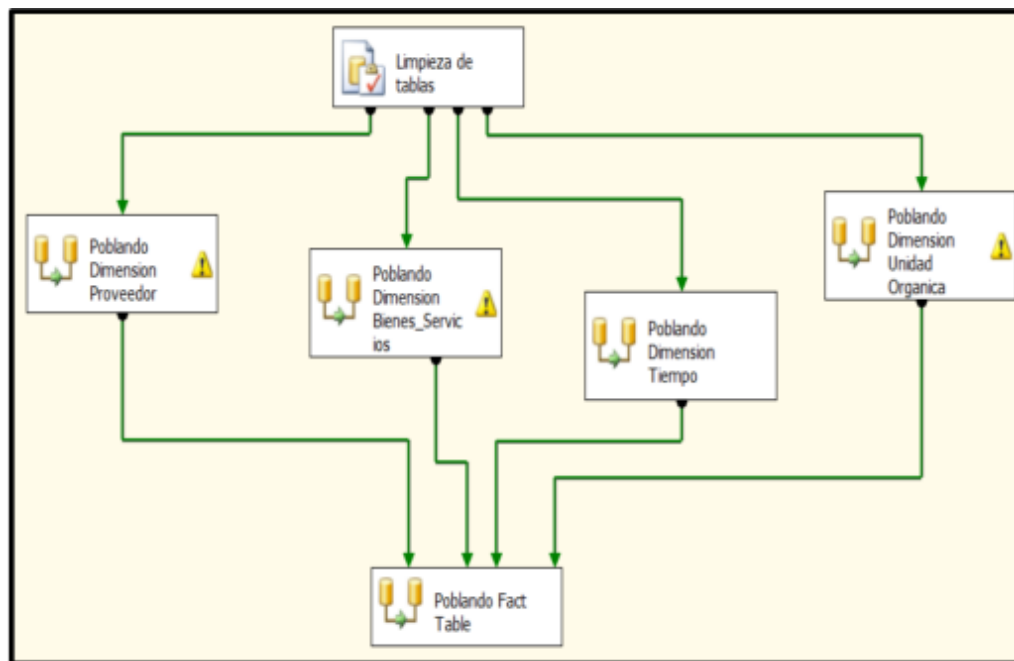


Figura 42. Diseño ETL de solución de inteligencia de negocios.

Discusión del caso 8: Cuando se tiene un proyecto de esta envergadura, es recomendable utilizar software especializado y licenciado. Tal es el caso del proyecto realizado para el gobierno. Por las figuras se entiende que este proyecto maneja mucha data histórica, por lo que es necesario utilizar una metodología como Ralph Kimball y SQL Server Business Intelligence. Sin embargo también podríamos utilizar la versión licenciada de Pentaho Business Analytics para proyectos similares, que lo podemos encontrar en: <http://www.pentaho.com> o comunicarnos a: mcalderon@cognus.cl de Marcela Calderón, Gerente Comercial de Cognus, firma representante de Pentaho en Latinoamérica.

Caso 9: Tapia (2006) en su investigación, presenta una metodología para sectorizar pacientes en el consumo de medicamentos aplicando Datamart en un Hospital Nacional. Lo interesante de esta solución es que aplicó su propia metodología que es parecida a HEFESTO y también desarrollo sus reportes DM a través de consultas. Lo que nos llama la atención es el modelo dimensional que utilizan para realizar su solución.

En la figura 43 del mismo autor, se presenta el esquema dimensional de su DM.

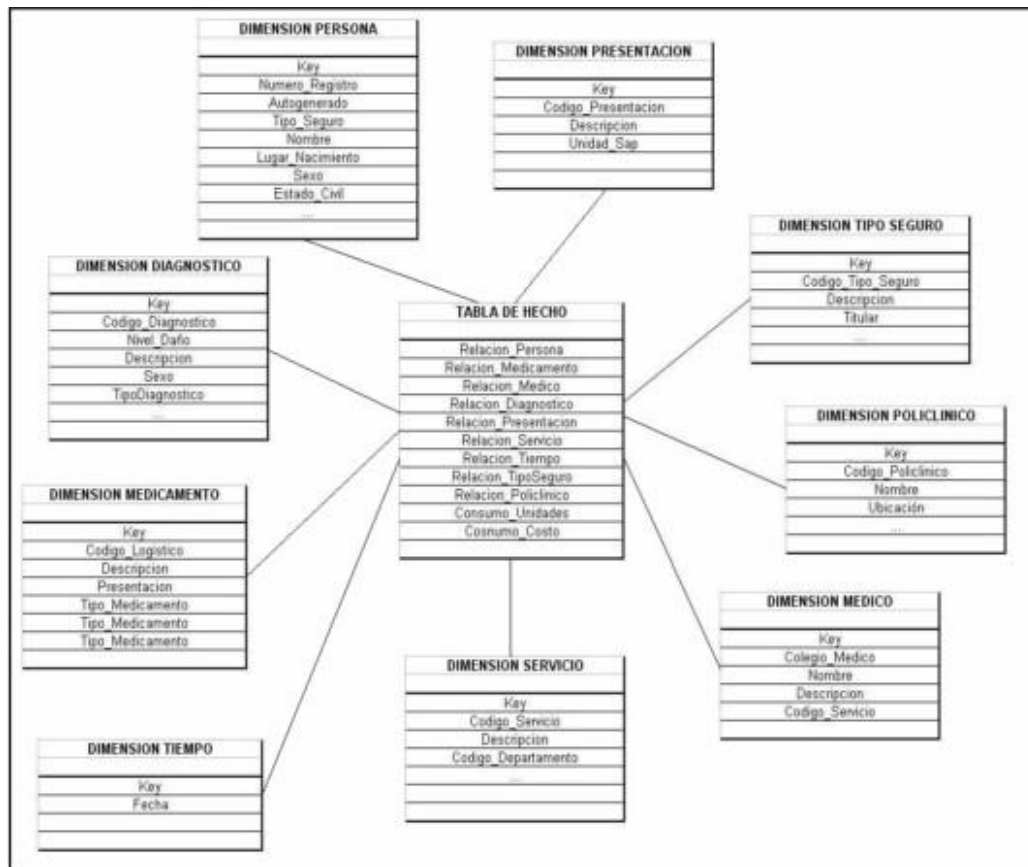
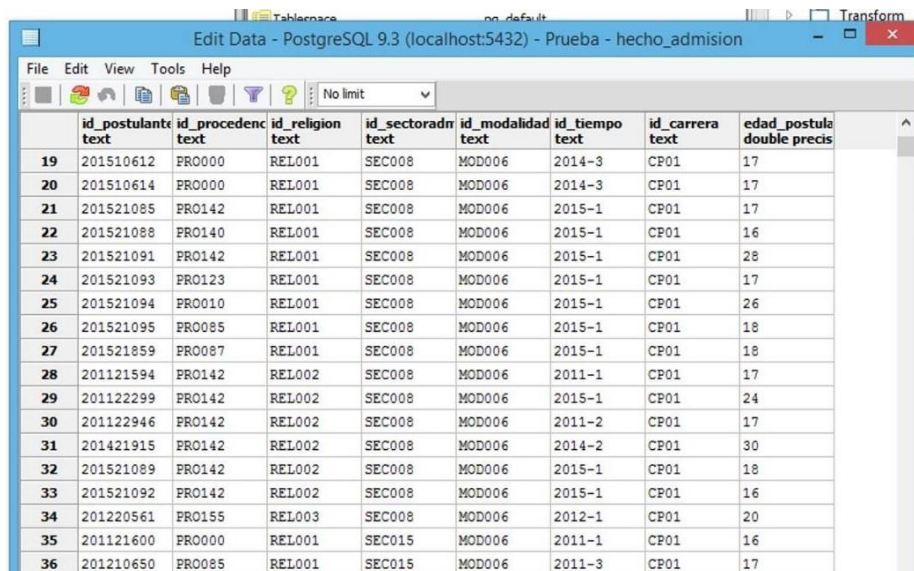


Figura 43. Esquema dimensional del DM.

Discusión del caso 9: Coincidimos con la idea del esquema dimensional del proyecto desarrollado por Tapia, ya que cuando los indicadores están bien establecidos, las dimensiones se harán con más facilidad y si todos están relacionados, entonces no será necesario utilizar un modelo dimensional copo de nieve, sino un modelo de estrella, como el que utilizamos en nuestro modelo. A su vez resaltamos que el hecho debe contener toda la información que se desea obtener para generar correctamente los reportes.

En la figura 44 elaborada en esta investigación, revelamos toda la información que contiene el hecho para cubrir con las exigencias propuestas por el cliente (admisión), basado en indicadores.



	id_postulante text	id_procedencia text	id_religion text	id_sectoradm text	id_modalidad text	id_tiempo text	id_carrera text	edad_postula double precis
19	201510612	PRO000	REL001	SEC008	MOD006	2014-3	CP01	17
20	201510614	PRO000	REL001	SEC008	MOD006	2014-3	CP01	17
21	201521085	PRO142	REL001	SEC008	MOD006	2015-1	CP01	17
22	201521088	PRO140	REL001	SEC008	MOD006	2015-1	CP01	16
23	201521091	PRO142	REL001	SEC008	MOD006	2015-1	CP01	28
24	201521093	PRO123	REL001	SEC008	MOD006	2015-1	CP01	17
25	201521094	PRO010	REL001	SEC008	MOD006	2015-1	CP01	26
26	201521095	PRO085	REL001	SEC008	MOD006	2015-1	CP01	18
27	201521859	PRO087	REL001	SEC008	MOD006	2015-1	CP01	18
28	201121594	PRO142	REL002	SEC008	MOD006	2011-1	CP01	17
29	201122299	PRO142	REL002	SEC008	MOD006	2015-1	CP01	24
30	201122946	PRO142	REL002	SEC008	MOD006	2011-2	CP01	17
31	201421915	PRO142	REL002	SEC008	MOD006	2014-2	CP01	30
32	201521089	PRO142	REL002	SEC008	MOD006	2015-1	CP01	18
33	201521092	PRO142	REL002	SEC008	MOD006	2015-1	CP01	16
34	201220561	PRO155	REL003	SEC008	MOD006	2012-1	CP01	20
35	201121600	PRO000	REL001	SEC015	MOD006	2011-1	CP01	16
36	201210650	PRO085	REL001	SEC015	MOD006	2011-3	CP01	17

Figura 44. Contenido de la tabla hecho_admision.

Caso 10: Sánchez (2009) desarrolló un DM para el apoyo en la toma de decisiones en los procesos académicos de la Universidad Peruana Unión, utilizando la metodología definida para desarrollo de software. Usó Pentaho para desarrollar su solución y la extensión Mondrian para agregarlo a su diseño web, teniendo un impacto positivo en el Departamento para el que lo realizó.

En las figuras 45 y 46 propuestas por el mismo investigador, podemos observar el modelo ETL que utilizó y sus reportes respectivamente.

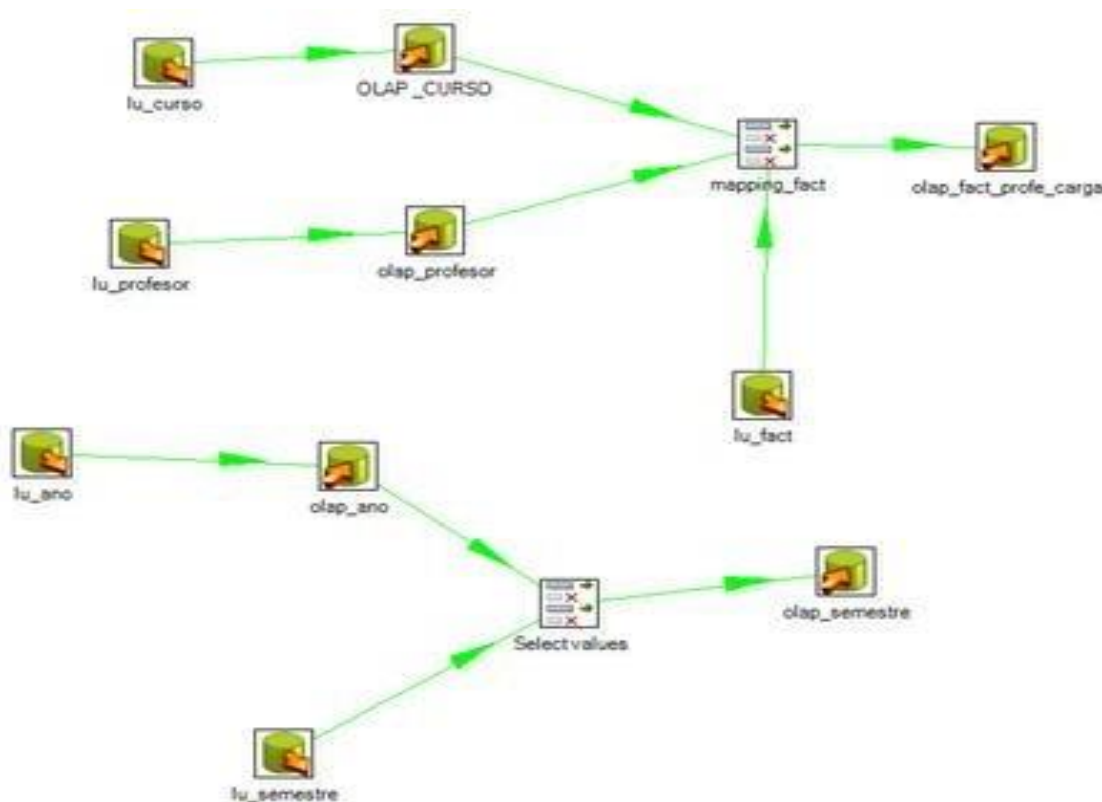


Figura 45. Modelo ETL del Kettle.

Discusión del caso 10: Respecto al proceso de ETL realizado por Sánchez, no coincidimos en que realiza doble trabajo en pasar toda la data de su .xls a las tablas designadas sin relación alguna y luego las extrae y transforma, para luego pasarlas al modelo dimensional propuesto por el autor. En contraposición nuestro modelo de ETL, propone seleccionar los datos necesarios de la fuente que es .xls y pasarlos al modelo dimensional estrella que propusimos, utilizando Select Values e Insert/Update que está dentro de Transformation en la herramienta Spoon del Pentaho.

Capítulo 5

Conclusiones y recomendaciones.

Conclusiones

Después del trabajo realizado, podemos destacar la importancia de un análisis y diseño inteligentes como Datamart para Admisión, el cual permitió convertir toda su data en información estratégica confiable de beneficio en las decisiones para sus campañas.

Esta investigación permitió tener una mejor experiencia en cuanto al análisis, exploración, transformar y carga de la fuente de datos, permitiendo detectar diferentes errores durante el proceso y ser corregidos a tiempo.

Nos llena de alegría haber tenido como base de ejecución a la metodología HEFESTO versión 2.0, que es flexible, ya que permitió agregar o modificar algunas fases para centrarse más en los requerimientos y los resultados.

Gracias a este trabajo, se pudo tener un mejor manejo de la suite Pentaho; que es muy útil para este tipo de proyectos tecnológicos, ya que contiene herramientas como: Spoon, Schema Workbench, Report Designer y Pentaho Analysis Report; que ayudan en toda la construcción de una solución de IN.

Recomendaciones

Recomendamos al momento de manipular la data histórica de una institución o empresa, se debe extraer un backup con extensión .xls, debido a que este tipo de extensión es leída por cualquier sistema operativo que cuente con Microsoft Excel, y es de fácil manejo al momento de filtrar los datos.

Con respecto a nuestra investigación faltó generar Dashboards, que serían útiles para la continuación de este modelo. Recomendamos continuar usando Pentaho para ampliar nuestro proyecto y si es posible obtener una licencia para adquirir la Suite instalable.

Nuestro modelo comprende solo el Departamento de Admisión, y se recomienda ingresar al sistema de admisión la mayor cantidad de información de las campañas de admisión, para que no existan datos nulos, lo que ocasiona perder algunos registros.

En relación a la recomendación anterior, es necesario estandarizar los datos de admisión, para que se pueda obtener un mejor resultado a futuro.

Recomiendo también adquirir el Pentaho última versión, que es licenciada y permite utilizar todas las herramientas de la Suite incluido su servidor online, y toda su fase de reportes y dashboards.

Por ultimo recomendamos usar nuestra metodología, para desarrollar importantes soluciones a las áreas importantes de la universidad como: área financiera, área académica, y área contable; o para soluciones de BI a instituciones educativas u otras que no tengan que ver con giros comerciales.

Lista de referencias

- Bello, F. (2006). REFLEXION:La investigación tecnológica: o cuando la solución es el problema. *Revista FACES*, 13, 14. Retrieved from <http://servicio.bc.uc.edu.ve/faces/revista/a6n13/6-13-3.pdf>
- Bernabeu, R. (2010). Hefesto. Córdoba, Argentina. Retrieved from <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/hefesto-metodologia-propia-para-la-construccion-un-data-wareh><http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/ii-hefesto-metodologia-propia-para-la-construccion-un-data-wa>
- Bustamante, A., Galvis, E., Gonzáles, M., García, A., & Benavides, L. (2011). Soluciones de inteligencia de negocios en la práctica: apoyo a la toma de decisiones en proyectos educativos para población infantil vulnerable en el caribe Colombiano. (Spanish). *UIS Ingenierías*, 10(2), 123–135. JOUR. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=zbh&AN=86166562&lang=es&site=ehost-live>
- Bustamante Martínez, A., Galvis Lista, E. A., & Gómez Flórez, L. C. (2013). Técnicas de modelado de procesos de ETL: una revisión de alternativas y su aplicación en un proyecto de desarrollo de una solución de BI. (Spanish). *ETL Processes Modeling Techniques: An Alternatives Review and Its Application in a BI Solution Development Project. (English)*, 18(1), 185–191. JOUR. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=eih&AN=89861721&lang=es&site=ehost-live>
- Bustos, S., & Mosquera, V. (2013). *Análisis, diseño e implementación de una solución BUSINESS INTELLIGENCE para la generación de indicadores y control de desempeño, en la empresa OTECEL*. Retrieved from <http://repositorio.espe.edu.ec/handle/21000/6305>
- Cárdenas, J. (2011). Implementación de business intelligence para el área de satisfacción al cliente de una empresa de telecomunicaciones utilizando control de inventarios.
- Chavez, J. D. (2013). *Análisis, Diseño E Implementación Una Solución De Inteligencia De Negocios Orientada a Controlar Los Procesos De Generación Y Emisión Del Dni (Documento Nacional De Identidad) En El Reniec (Registro Nacional De Identificación Y Estado Civil)*. Universidad Católica del Perú. Retrieved from <http://tesis.pucp.edu.pe/repositorio/handle/123456789/5314>
- Cruz, G. R. (2003). *Utilización de la tecnología Data Warehouse en Instituciones Educativas*. Universidad Autónoma del Estado de Hidalgo. Retrieved from <http://www.uaeh.edu.mx/docencia/Tesis/icbi/maestria/documentos/Utilizacion de la>

tecnologia.pdf

- Dall'Orto, L., & Wu, R. (2006). *Construcción y Pruebas de una herramienta de desarrollo de soluciones para inteligencia de Negocios Módulo de Extracción. Test*. Pontificia Universidad Católica del Perú. <http://doi.org/10.1017/CBO9781107415324.004>
- Falcón, S., Palma, R., & Celi, R. (2006). *Construcción de una Herramienta de Desarrollo de Soluciones para Inteligencia de Negocios - Explotación. Test*. Universidad Católica del Perú. <http://doi.org/10.1017/CBO9781107415324.004>
- Fuentes, L., & Valdivia, R. (2010). Incorporation of Business Intelligence Elements in the Admission. *Revista Chilena de Ingeniería*, 18, 383–394.
- Gelinas, U., Sutton, S., & Federowicz, J. (2008). *Business Processes and Information Technology. Information Systems*.
- Guillén, F. (2012). *Desarrollo de un Datamart para mejorar la toma de decisiones en el Área de Tesorería de la Municipalidad Provincial de Cajamarca*. Universidad Privada del Norte. <http://doi.org/10.1017/CBO9781107415324.004>
- Gutiérrez, P. (2012). *Metodología de uso de herramientas de inteligencia de negocios como estrategia para aumentar la productividad y competitividad de una PyME*. Retrieved from <http://148.204.210.201/tesis/1359572993732PamelaGutirre.pdf>
- Hernández, R., Fernández, C., & Baptista, P. (2010). *Metodología de la investigación*.
- Masaquiza, C. (2011). *Facultad De Ingeniería En Sistemas , Electrónica E Industrial. Estudent*. Retrieved from http://repo.uta.edu.ec/bitstream/123456789/8595/1/Tesis_t953si.pdf
- Moss, L., & Atre, S. (2003). *Business Intelligence Roadmap. Communication*.
- Ojeda Villanueva, Á. (2008). *Analisis, Diseño E Implementacion De Un Datawarehouse De Soporte De Desiciones Para Un Hospital Del Sistema De Salud Publico. Sistemas de Informacion*. Universidad Católica del Perú. Retrieved from <http://tesis.pucp.edu.pe/repositorio/handle/123456789/333>
- Oktavia, T. (2014). IMPLEMENTING DATA WAREHOUSE AS A FOUNDATION FOR DECISION SUPPORT SYSTEM (PERSPECTIVE : TECHNICAL AND NONTECHNICAL FACTORS). *Journal of Theoretical & Applied Information Technology*, 60(3), 476–482. JOUR. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=iih&AN=94866637&lang=es&site=ehost-live>

- Rincón, I. (2013). Investigación científica y tecnológica como factores de innovación, 127–148. Retrieved from <http://www.eumed.net/entelequia/pdf/2012/e14a09.pdf>
- Rojas, A. (2014). *Implementación de un Data Mart como Solución de Inteligencia de Negocios, bajo la Metodología de Ralph Kimball para optimizar la toma de decisiones en el Departamento de Finanzas de la Contraloría General de la República*. Universidad San Martín de Porres. Retrieved from http://www.repositorioacademico.usmp.edu.pe/bitstream/usmp/1061/1/rojas_a.pdf
- Sánchez, J. A. (2009). *Tesis*. Universidad Peruana Unión.
- Tapia, I. (2006). Una Metodología para sectorizar pacientes en el consumo de medicamentos aplicando Datamart y Datamining en un hospital nacional, 111. Retrieved from <http://cybertesis.unmsm.edu.pe/xmlui/handle/cybertesis/273>
- Torriente, L. T., Sentí, V. E., Hernández, W. G., & Ortega, R. R. (2011). Instrumento y herramienta informática para guiar, controlar y evaluar las interacciones de los estudiantes en foros virtuales. *Revista Cubana de Educacion Medica Superior*, 25(2), 59–70.
- Valdiviezo, M., Herrera, I., & Jáuregui, G. (2007). *Análisis y Diseño de una herramienta de desarrollo de soluciones para inteligencia de negocios - Análisis dimensional. Test*. Retrieved from http://tesis.pucp.edu.pe/repositorio/bitstream/handle/123456789/327/VALDIVIEZO_MANUEL_AN%C3%81LISIS_Y_DISE%C3%91O_DE_UNA_HERRAMIENTA_DE_DESARROLLO_DE_SOLUCIONES_PARA_INTELIGENCIA_DE_NEGOCIOS_AN%C3%81LISIS_DIMENSIONAL.pdf?sequence=1
- Zambrano, C., Rojas, D., Carvajal, K., & Acuña, G. (2011). Análisis de rendimiento académico estudiantil usando data warehouse y redes neuronales. Analysis of students ' academic performance using data warehouse and neural networks. *Revista Chilena de Ingeniería*, 19, 369–381.

Anexos

Anexos 1. Consulta a la base de datos para extraer el backup Admisión.

```
UPEUADMISION_DISTRITO di, UPEUADMISION
_INSTITUCION_PROCEDENCIA ie, UPEUADMISION _FILIACION_RELIGIOSA
re, UPEUADMISION _SEDE se, UPEUADMISION _MODALIDAD mo
where a.ID_DP = b.ID_DP
and a.ID_DD= di.ID_D
and di.ID_PRO = pr.ID_PRO
and pr.ID_DEP = de.ID_DEP
and a.ID_INSTITUCION = ie.ID_INSTITUCION
and a.ID_FR= re.ID_FR
and b.SEDE_ID = se.SEDE_ID
and b.ID_MOD = mo.ID_MOD
and length(a.DOC_CODUNI) = 9
and b.SEC_ID1 like '5%'
```

Anexos 2. Código SQL para crear la Base de datos en Postgresql.

```
CREATE TABLE public.adm_carrera
(
  id_carrera text NOT NULL,
  carrera text,
  CONSTRAINT id_carrera PRIMARY KEY (id_carrera)
)
WITH (
  OIDS=FALSE
);
ALTER TABLE public.adm_carrera OWNER TO postgres;
CREATE INDEX idx_adm_carrera_lookup
ON public.adm_carrera
USING btree
(id_carrera, carrera);

-----
CREATE TABLE public.adm_modalidad
(
  id_modalidad text NOT NULL,
  modalidad text,
  CONSTRAINT id_modalidad PRIMARY KEY (id_modalidad)
)
```

```
WITH (  
    OIDS=FALSE  
);  
ALTER TABLE public.adm_modalidad OWNER TO postgres;  
CREATE INDEX idx_adm_modalidad_lookup  
    ON public.adm_modalidad  
    USING btree  
    (id_modalidad, modalidad);  
  
-----  
CREATE TABLE public.adm_postulante  
(  
    fecha_nacimiento date,  
    id_postulante text NOT NULL,  
    nombres text,  
    apellidos text,  
    sexo text,  
    CONSTRAINT id_postulante PRIMARY KEY (id_postulante)  
)  
WITH (  
    OIDS=FALSE  
);  
ALTER TABLE public.adm_postulante OWNER TO postgres;
```



```
CREATE INDEX idx_adm_postulante_lookup
ON public.adm_postulante
USING btree
(id_postulante, nombres, apellidos, sexo, fecha_nacimiento);

-----

CREATE TABLE public.adm_procedencia
(
id_procedencia text NOT NULL,
distrito text,
provincia text,
departamento text,
CONSTRAINT id_procedencia PRIMARY KEY (id_procedencia)
)
WITH (
OIDS=FALSE
);
ALTER TABLE public.adm_procedencia OWNER TO postgres;
CREATE INDEX idx_adm_procedencia_lookup
ON public.adm_procedencia
USING btree
(id_procedencia, distrito, provincia, departamento);

-----
```

```
CREATE TABLE public.adm_religion
(
    id_religion text NOT NULL,
    religion text,
    CONSTRAINT id_religion PRIMARY KEY (id_religion)
)
WITH (
    OIDS=FALSE
);
ALTER TABLE public.adm_religion OWNER TO postgres;
CREATE INDEX idx_adm_religion_lookup
    ON public.adm_religion
    USING btree
    (id_religion, religion);
```

```
CREATE TABLE public.adm_sectoradm
(
    id_sectoradm text NOT NULL,
    sector_adm text,
    CONSTRAINT id_sectoradm PRIMARY KEY (id_sectoradm)
)
WITH (
```

```
    OIDS=FALSE
);

ALTER TABLE public.adm_sectoradm OWNER TO postgres;

CREATE INDEX idx_adm_sectoradm_lookup
    ON public.adm_sectoradm
    USING btree
    (id_sectoradm, sector_adm);

-----

CREATE TABLE public.adm_tiempo
(
    fecha_campana date,
    id_tiempo text NOT NULL,
    CONSTRAINT id_tiempo PRIMARY KEY (id_tiempo)
)

WITH (
    OIDS=FALSE
);

ALTER TABLE public.adm_tiempo OWNER TO postgres;

CREATE INDEX idx_adm_tiempo_lookup
    ON public.adm_tiempo
    USING btree
    (id_tiempo, fecha_campana);
```

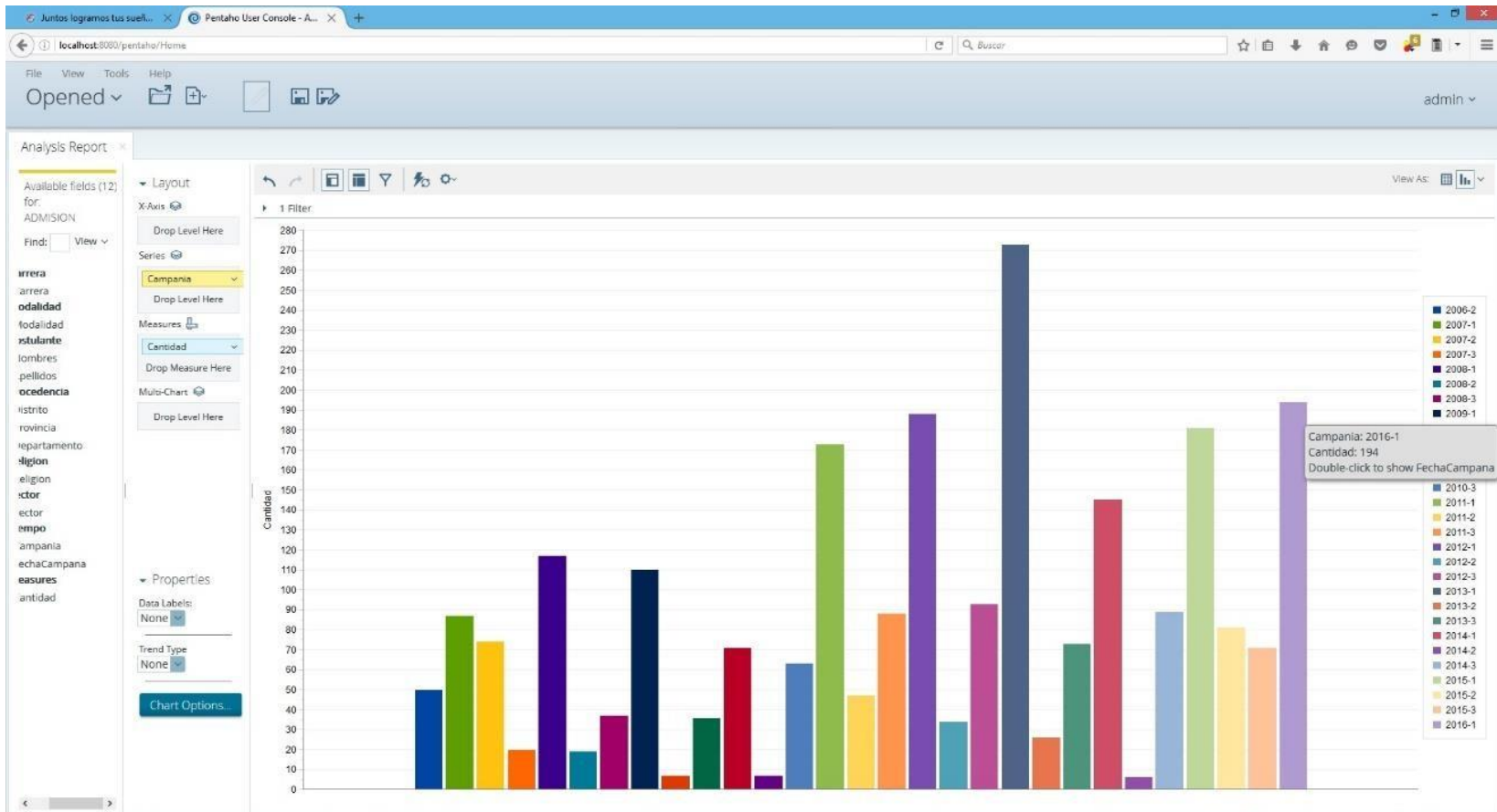
```
CREATE TABLE public.hecho_admision
(
  id_postulante text NOT NULL,
  id_procedencia text NOT NULL,
  id_religion text NOT NULL,
  id_sectoradm text NOT NULL,
  id_modalidad text NOT NULL,
  id_tiempo text NOT NULL,
  id_carrera text NOT NULL,
  edad_postulacion double precision,
  CONSTRAINT pk_hechoadmision PRIMARY KEY (id_postulante, id_procedencia,
id_religion, id_sectoradm, id_modalidad, id_tiempo, id_carrera),
  CONSTRAINT id_carrera FOREIGN KEY (id_carrera)
  REFERENCES public.adm_carrera (id_carrera) MATCH Unknown
  ON UPDATE NO ACTION ON DELETE NO ACTION,
  CONSTRAINT id_modalidad FOREIGN KEY (id_modalidad)
  REFERENCES public.adm_modalidad (id_modalidad) MATCH Unknown
  ON UPDATE NO ACTION ON DELETE NO ACTION,
  CONSTRAINT id_postulante FOREIGN KEY (id_postulante)
  REFERENCES public.adm_postulante (id_postulante) MATCH Unknown
  ON UPDATE NO ACTION ON DELETE NO ACTION,
```

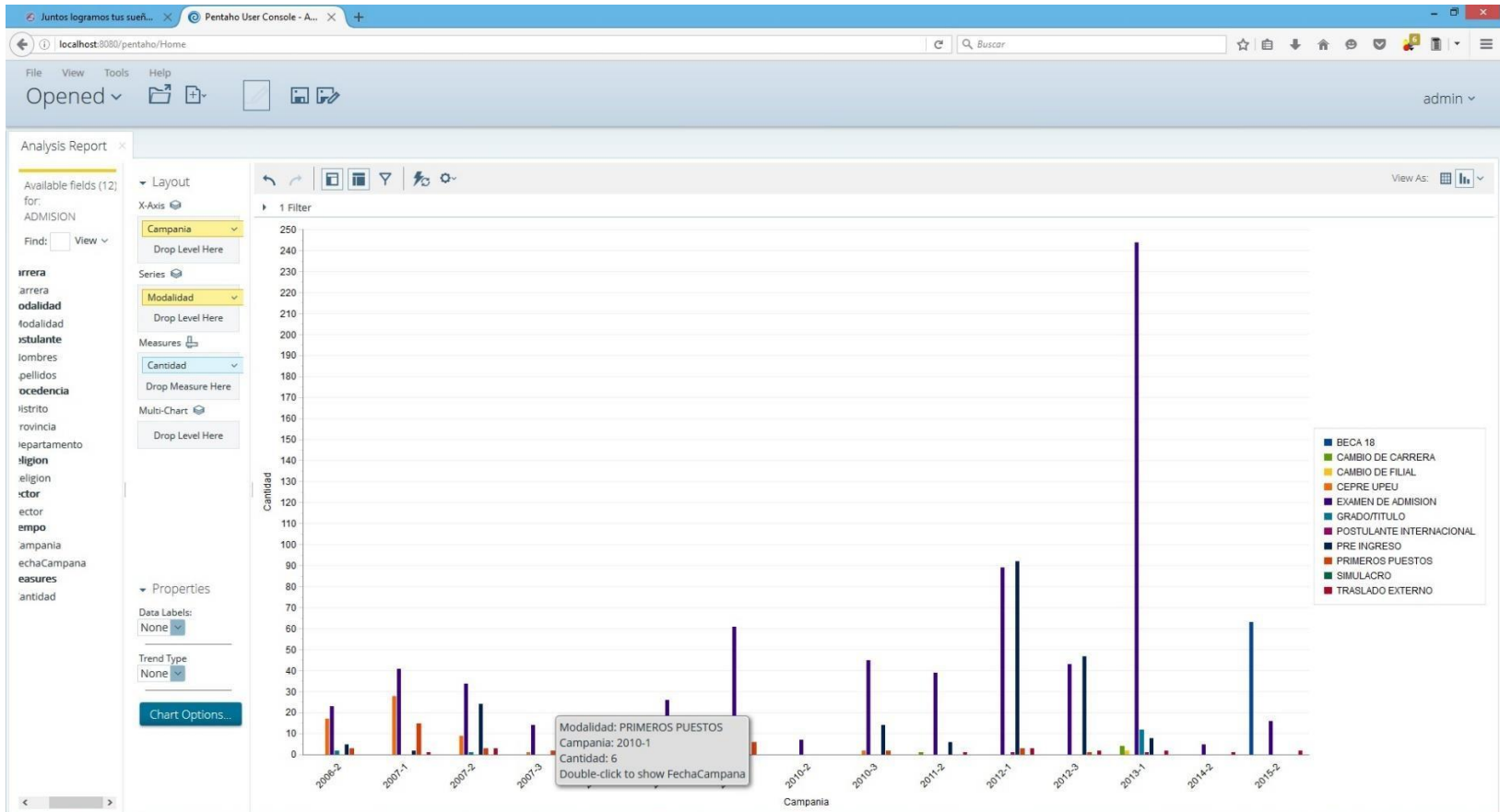
```

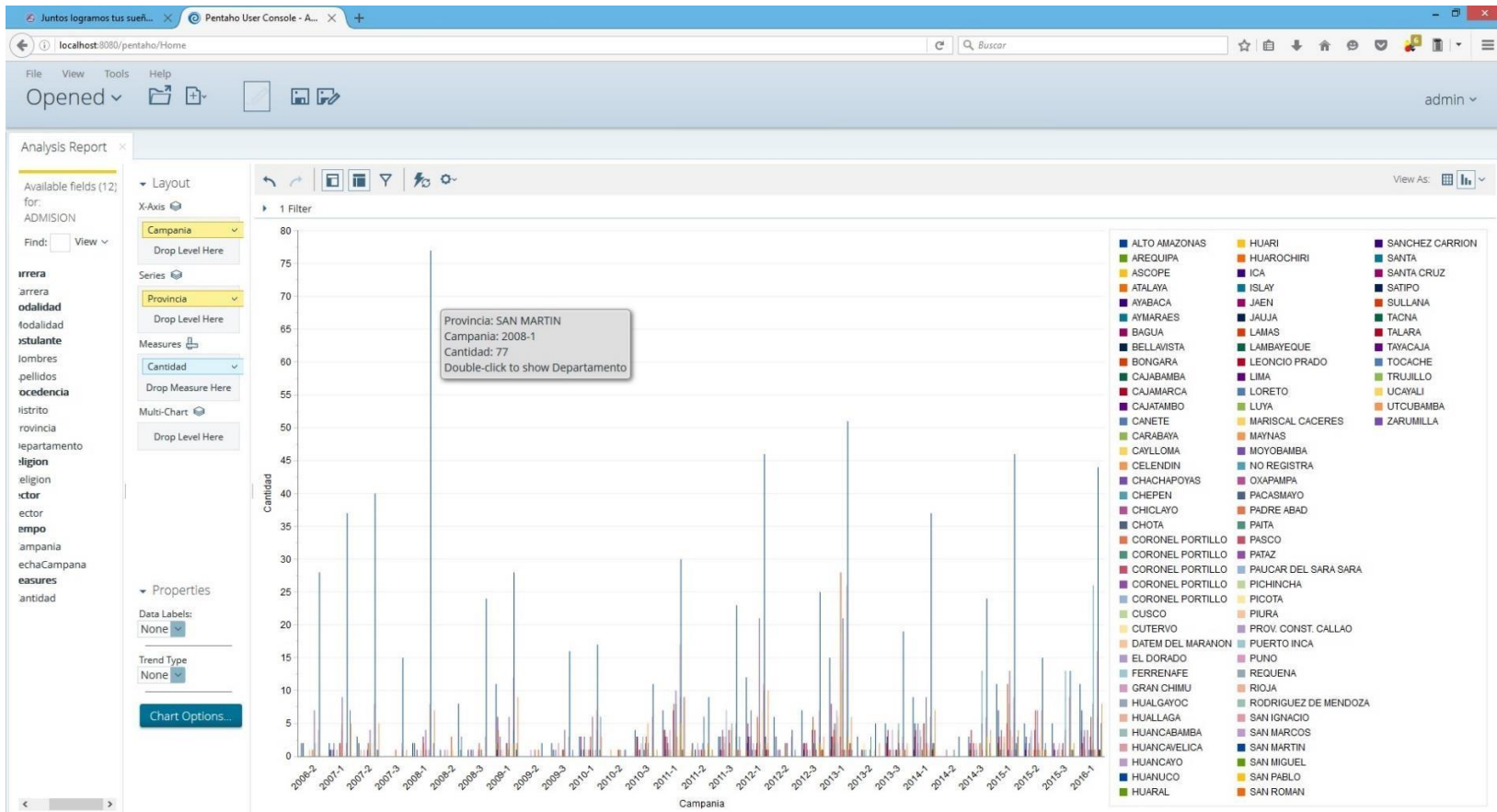
CONSTRAINT id_procedencia FOREIGN KEY (id_procedencia)
REFERENCES public.adm_procedencia (id_procedencia) MATCH Unknown
ON UPDATE NO ACTION ON DELETE NO ACTION,
CONSTRAINT id_religion FOREIGN KEY (id_religion)
REFERENCES public.adm_religion (id_religion) MATCH Unknown
ON UPDATE NO ACTION ON DELETE NO ACTION,
CONSTRAINT id_sectoradm FOREIGN KEY (id_sectoradm)
REFERENCES public.adm_sectoradm (id_sectoradm) MATCH Unknown
ON UPDATE NO ACTION ON DELETE NO ACTION,
CONSTRAINT id_tiempo FOREIGN KEY (id_tiempo)
REFERENCES public.adm_tiempo (id_tiempo) MATCH Unknown
ON UPDATE NO ACTION ON DELETE NO ACTION
)
WITH (
OIDS=FALSE
);
ALTER TABLE public.hecho_admision OWNER TO postgres;
CREATE INDEX idx_hecho_admision_lookup
ON public.hecho_admision
USING btree
(id_postulante, id_procedencia, id_religion, id_sectoradm, id_modalidad, id_tiempo,
id_carrera, edad_postulacion);

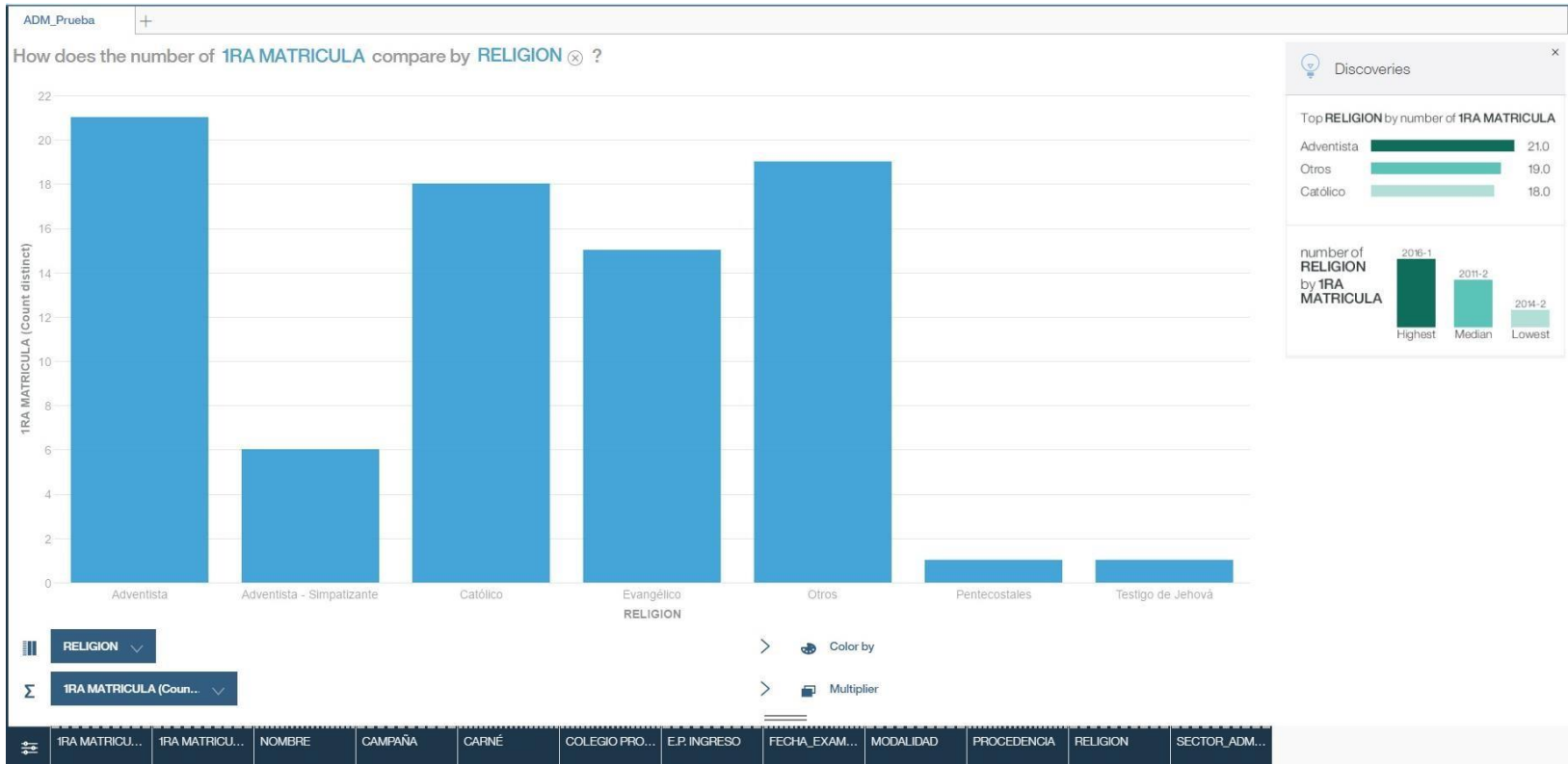
```

Anexos 3. Otros reportes.









Anexos 4. Constancia de conformidad del Departamento de Admisión.



UNIVERSIDAD PERUANA UNIÓN
EDUCATIO·HVIC·VITAE·ET·AETERNITATI

Una Institución Alceñista

UNIVERSIDAD PERUANA UNIÓN

Juntos logramos tus sueños

**EL COORDINADOR DEL DEPARTAMENTO DE ADMISIÓN
DE LA UNIVERSIDAD PERUANA UNIÓN, FILIAL
TARAPOTO:**

ADMISIÓN

- Administración en Gestión Empresarial
- Contabilidad y Gestión Tributaria
- Ingeniería Ambiental
- Ingeniería de Sistemas
- Marketing y Negocios Internacionales
- Psicología
- Arquitectura

PROESAD

- Administración y Negocios Internacionales
- Contabilidad y Gestión Tributaria

HACE CONSTAR:

Que, el Señor, Ivan Guadaña Quiroz, identificado con DNI N° 46598091, prestó servicio AD HONOREM como Bachiller en Ingeniería de Sistemas, diseñando un modelo de DataMart, de ayuda para la toma de decisiones del Departamento de Admisión, de la Universidad Peruana Unión, Filial Tarapoto, habiendo concluido satisfactoriamente el desarrollo de su investigación y presentando reportes importantes para el departamento.

Se expide la presente constancia a solicitud del interesado.

Morales, 05 de agosto del 2016



JOSÉ TARRILLO PAREDES
COORDINADOR ADMISIÓN

www.upeu.edu.pe - Jr. Los Martires 218 - Santa Lucia - Morales / Telf.: (042)520024 RPC 968701943 / 968701966
E-mail: admission.tpp@upeu.edu.pe

Anexos 5. Constancia del Centro de Idiomas.



 UNIVERSIDAD PERUANA UNIÓN

 EDUCAT HUIIC VITAE ET AETERINITATI

Una Institución Adventista

“Año de la Consolidación del Mar de Grau”

Tarapoto, viernes 05 de agosto de 2016

CI-FT-02-2016

Ingeniero
Joel Pérez Suárez
 Coordinador de la Escuela de Ingeniería de Sistemas
 Tarapoto.

Estimado ingeniero Pérez:

Tengo el agrado de saludarles, deseándole éxitos en la labor que desempeña.

Asimismo, me dirijo a usted con el propósito de informarle que el Centro de Idiomas UPeU Tarapoto realizó la traducción de español a inglés del resumen de tesis que corresponde al siguiente alumno:

1.- Guadaña Quiroz, Iván. COD: 201121396

Le expreso mi gratitud por su gentil atención.

Con afecto y cordialidad,




Lic. Freddy Chavez Moleros
 Director de Centro de Idiomas

UNIVERSIDAD PERUANA UNIÓN
 FILIAL TARAPOTO
 FACULTAD DE INGENIERIA Y
 ARQUITECTURA
 05 AGO 2016
RECIBIDO
 Expediente:

Jr. Los Mártires N° 218 - Urb. Santa Lucía - Morales (Carretera Fernando Belaunde Terry Km 1.5) Tels (042) 52-1414, 52-5537, 52-6276
 Fax: (042) 52-9450 - Telulink: 968701994 - E-mail: admision.tpp@upeu.edu.pe - www.upeu.edu.pe

Vita

Ivan Guadaña Quiroz, nació el 24 de julio de 1990, en la Provincia de Celendín, Provincia de Celendín, Departamento de Cajamarca. Sus padres son: el Sr. Félix Guadaña Chávez y la Sra. Victoria Quiroz Cabrera. Sus primeros años los vivió en dicha ciudad, a sus 3 años ingresó a la Institución Educativa Pública Inicial N° 72. Su primaria lo realizó en el Centro Educativo de Gestión No Estatal “Cristo el Redentor”. Su secundaria lo realizó en la Institución Pública Estatal “Coronel Cortegana”.

Desde el año 2007 estudió la Carrera Técnica de Computación e Informática en el Universal Instituto Basado en el Avance de la Ciencia y la Tecnología (UNIBACT).

Trabajó en la Red III de Salud Celendín de la Dirección Regional de Salud Cajamarca en el área de Estadística y Epidemiología, los años 2009 y 2010.

El año 2011 ingresó a la Universidad Peruana Unión, Filial Tarapoto para estudiar la carrera de Ingeniería de Sistemas.

Los años 2014 y 2015, se desempeñó como Asistente en Procesos de Bibliotecología, en la UPeU-FT.

El primer semestre del 2016, se desempeñó como Asistente de Acreditación de la UPeU-FT.

Actualmente es Bachiller en Ingeniería de Sistemas y labora en el Área de Desarrollo de la Dirección General en Tecnologías de Información, de la universidad.