

UNIVERSIDAD NACIONAL JOSÉ MARÍA ARGUEDAS

FACULTAD DE INGENIERÍA

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS



Presentado por

VICTOR ALBERTO LIMA ROMÁN

DESARROLLO DE UN MODELO DE CUBO OLAP (ON-LINE ANALYTICAL PROCESSING) PARA LA TOMA DE DECISIONES EN EL CENTRO PREUNIVERSITARIO DE LA UNIVERSIDAD NACIONAL JOSÉ MARÍA ARGUEDAS

Asesor:

Mtr. JUAN JOSÉ ORÉ CERRÓN

TESIS PARA OPTAR EL TÍTULO PROFESIONAL DE INGENIERO DE SISTEMAS

ANDAHUAYLAS – APURÍMAC - PERÚ

2019

DEDICATORIA

A mi señora madre Bernardina Román Quintana, por haber contribuido con mi educación y mi formación como profesional.

AGRADECIMIENTO

A todos los docentes de la Carrera profesional de ingeniería de Sistemas por haber contribuido en mi formación profesional y a todas aquellas personas que contribuyeron en la ejecución del presente trabajo.

ÍNDICE GENERAL

RESUMEN	9
ABSTRACT	10
CHUMASQA.....	11
CAPITULO I	12
INTRODUCCION	12
1.1 Realidad problemática.....	12
1.2 Formulación del problema	15
1.3 Objetivos	15
1.4 Objetivos Específicos	15
1.5 Justificación.....	15
1.6 Viabilidad de la investigación	16
1.6.1 Viabilidad técnica	16
1.6.2 Viabilidad económica	16
1.6.3 Viabilidad social.....	16
1.7 Limitación del estudio	16
CAPITULO II	17
MARCO TEÓRICO.....	17
2.1 Antecedentes	17
2.1.1 Antecedentes Internacionales	17
2.1.2 Antecedentes Nacionales.....	19
2.2 Inteligencia de negocios.....	21
2.2.1 Almacén de datos.....	22
2.2.2 Mercado de datos	24
2.2.3 Cubos OLAP (On-line Analytical Processing)	25
2.3. Data Warehouse	26
2.3.1 Características	26
2.3.2 Esquemas	27
2.4 Cubos de datos	30
2.5 toma de decisiones	30
2.5.1 Diversas definiciones de toma de decisiones.....	31
2.5.2 Patrones y reglas para la toma de decisiones	31

2.5.3 Tecnologías de información en la toma de decisiones	33
2.5.4 Situaciones para implementar cubos OLAP	33
CAPÍTULO III	35
MATERIALES Y MÉTODOS	35
3.1 Esquema de propuesta de solución	35
3.1.1 Modelamiento de la Base de datos del CEPRE-UNAJMA.....	35
3.1.2 Cargado de registros a la base de datos.....	36
3.1.3 Dimensiones del cubo OLAP.....	36
3.1.4 Esquema en formato XML.....	36
3.1.5 Presentación de propuesta y retroalimentación de sugerencias	36
3.2 Construcción de cubos OLAP para el Centro Preuniversitario	37
3.2.1 Reglas del negocio del Centro Preuniversitario	37
3.2.2 Modelo Físico de la base de datos del Centro Preuniversitario de la UNAJMA.....	39
CAPÍTULO III	65
RESULTADOS Y DISCUSIÓN.....	65
CONCLUSIONES.....	69
RECOMENDACIONES	70
REFERENCIAS BIBLIOGRÁFICAS	71

ÍNDICE DE TABLAS

Tabla N° 1: Registro estudiantes matriculados en el CEPRE	13
Tabla N° 2: Esquema de propuesta de solución.	35
Tabla N° 3: Pagos estudiante regular.....	38
Tabla N° 4: Pagos en partes	38
Tabla N° 5: Pago estudiantes con media beca.	39
Tabla N° 6: Tabla en formato XLS del Centro preuniversitario.....	42

ÍNDICE DE FIGURAS

Figura N° 1: Diagrama para formular problemas.....	14
Figura N° 2: Componentes de un almacén de datos.....	23
Figura N° 3: modelo multidimensional que podría crearse para representar productos, regiones, ventas reales y ventas proyectadas.	26
Figura N° 4: Esquema estrella de los comprobantes del supermercado.....	28
Figura N° 5: Organigrama del Centro Preuniversitario de la UNAJMA.....	37
Figura N° 6: Modelo físico de la base de datos del Centro preuniversitario.	40
Figura N° 7: Implementación de la base de datos en el SABD Mysql.....	41
Figura N° 8: Extracto de los registros encontrados en su base de datos.	42
Figura N° 9: Relación de registros con redundancia de datos.	43
Figura N° 10: Entorno del Software Schema Workbench.....	45
Figura N° 11: Ventana de conexión con la base de datos.....	45
Figura N° 12: Mensaje de conexión exitosa.	46
Figura N° 13: Creación de nuevo esquema.	46
Figura N° 14: Creación de un nuevos esquema.....	47
Figura N° 15: Ingresando nombre de dimensión.....	47
Figura N° 16: Agregar tabla a la dimensión.....	48
Figura N° 17: Características de la tabla.....	48
Figura N° 18: Enlaza tabla a dimensión.	49
Figura N° 19: Agregando nivel.	49
Figura N° 20: Definiendo valores de nivel.	50
Figura N° 21: Definiendo valores del nivel distrito.....	51
Figura N° 22: Creando cubos.....	51
Figura N° 23: Atributos de la dimensión de uso.	52
Figura N° 24: reporte de cubo creado con Pentaho.	53
Figura N° 25: Reporte de cubo OLAP en Pentaho V7.0.....	53
Figura N° 26: Administrando Data Sources.....	54
Figura N° 27: Creando una nueva conexión.	55
Figura N° 28: Importando análisis.	55
Figura N° 29: Mostrando dos dimensiones del cubo.....	56
Figura N° 30: Gráfico de barras del cubo OLAP.....	57
Figura N° 31: Pantalla de inicio de Pentaho Data Integration.	58

Figura N° 32: Entorno de inicio de Pentaho Data Integration	58
Figura N° 33: Tabla de hechos en el modelo estrella.	59
Figura N° 34: Creando nueva conexión.	60
Figura N° 35: Parámetros de base de datos MySql.	60
Figura N° 36: Mensaje de error en la conexión.	61
Figura N° 37: Página de descarga de archivos con extensión JAR.	62
Figura N° 38: Mensaje de conexión de BD exitoso.	63
Figura N° 39: Esquema de transformación en PDI.....	63
Figura N° 40: Consulta de Sql en tabla de entrada.	64
Figura N° 41: Configuración en tabla de salida.	64
Figura N° 42: Reporte de tabla de salida de Pentaho data Integration	64
Figura N° 43: Modelo multidimensional diseñado en Mysql.....	65
Figura N° 44: Primer reporte de Cubo Olap del Centro Preuniversitario.	66
Figura N° 45: Reporte gráfico de Cubos OLAP.	67
Figura N° 46: reporte de tres dimensiones y dos métricas.....	67
Figura N° 47: Reporte sobre la altitud sobre el nivel del mar de los colegios...	68

RESUMEN

La investigación titulada: Desarrollo de un modelo de cubo OLAP (on-line analytical processing) para la toma de decisiones en el Centro Preuniversitario de la universidad nacional José María Arguedas plantea el objetivo de Desarrollar un modelo de cubo OLAP para el soporte de la toma de decisiones estratégicas en el Centro Preuniversitario de la Universidad Nacional José María Arguedas.

Aplicando el procedimiento de desarrollo de cubos OLAP con las herramientas tecnológicas de MySQL y Pentaho se logró concluir realizar el modelamiento de la base de datos considerando las reglas de modelamiento usando el software Mysql Workbench 6.3 En base a las reglas de negocio, cargar la base de datos al Administrador de base de datos Mysql, fue una labor que requirió bastante esfuerzo ya que las tablas estaban en formato .xls y algunos registros no eran uniformes, lo que incremento el tiempo de cargado de la base de datos, se realizó la integración se usaron los softwares Pentaho, Mysql y Esquema workbench; finalmente se obtuvo el cubo OLAP obtenido indica la cantidad de postulantes por colegio, con lo que podemos tomar decisiones sobre el tema de difusión en los colegios que aportan más postulantes a nuestra universidad.

Palabras clave: Cubo OLAP, inteligencia de negocios

ABSTRACT

The research entitled: Development of an on-line analytical processing OLAP cube model for decision making at the José María Arguedas National University Pre-University Center proposes the objective of developing an OLAP cube model for the support of the intake of strategic decisions at the Pre-University Center of the José María Arguedas National University.

Applying the procedure of development of OLAP cubes with the technological tools of MySQL and Pentaho, it was possible to conclude the modeling of the database considering the modeling rules using the Mysql Workbench 6.3 software. Based on the business rules, load the database data to the Mysql Database Administrator, it was a work that required a lot of effort since the tables were in .xls format and some records were not uniform, which increased the loading time of the database, the integration was performed they used the software Pentaho, Mysql and Scheme workbench; Finally, the obtained OLAP cube was obtained indicating the number of applicants per school, with which we can make decisions on the issue of dissemination in the colleges that contribute more applicants to our university.

Keywords: OLAP cube, business intelligence

CHUMASQA

Kay sasachakuypa sutinmi:desarrollo de un modelo de Cubo OLAP (on-line analytical processing), kay sasachakuyqa kanqa allin imapas puririchinankupaq Hatun Yachaywasinchikpi Centro Pre universitario nisqampi;kayta qawarispanku allinta ñaupaqman tukuy ruwasqankuta puririchinankupaqkay Hatun Yachauwasinchikpi.

Kay sasachakuy qispirinanpaqmi ancha anchata qawarqani chay herramientas tecnológicas de MySQL Y Pentaho, chaymi ñaupaqman puririchirqa hinallataqmi tukukurqa chay modelamiento de la Base de Datos nisqan chay software Mysql Workbench 6.3.

Hinallataqmi sasachakuypi tarikurqani chay Modelamiento de Base de Datos puririchinaypaq, manan llapallanchu allin karqa, chaymi astawan qiparurqani chay cargado de la Base de Datos nisqampi.

Tukunapaqmi chapurqani los softwares Pentaho,Mysql y Esquema Workbench;,qipata vhay Cubo OLAP quwananpaq.

Chay Cubo OLAP nisqanmi qawarichiwankik haykam warmakuna sapa yachaywasimanta yaykuyta munachkan ; chayta qawarispan allinta imatapas ruwanankupaq.

Hinallataqmi yachasun qaykatan sapa yachaywasimanta warmakuna hamunku qatun Yachay Wasinchikman.

Rimanapaq : Cubo OLAP,yuyaymanaspa allinta qispichisun.

CAPITULO I

INTRODUCCION

1.1 Realidad problemática

El Centro Preuniversitario (CEPRE) brinda preparación académica a los estudiantes que deseen postular a la Universidad Nacional José María Arguedas (UNAJMA), desde el año 2007 a la fecha. Operando doce años de servicio ininterrumpido dentro del ámbito de la provincia de Andahuaylas, en la medida que han ido pasando los años y la universidad ha ido incrementando carreras profesionales. El CEPRE también incrementó la cantidad de estudiantes pasando de 300 estudiantes en promedio cuando la universidad tenía tres carreras profesionales a 600 estudiantes, ahora que se incrementó a 6 carreras profesionales. El servicio del CEPRE es la preparación para el examen de admisión a la universidad y los estudiantes que obtengan las mejores calificaciones se les otorgan vacantes directas de ingreso en número de 10 por cada una de las seis carreras que oferta la Universidad.

El Centro Preuniversitario, también es considerado como un centro de producción, aporta a la ingresos de la Universidad a la partida de los recursos directamente recaudados de la UNAJMA, es dirigido por un docente designado por la autoridad de turno su gestión implica: Elaborar un presupuesto de funcionamiento, seleccionar docentes, realizar la preparación académica y seleccionar a los mejores estudiantes para otorgar vacantes dentro de la UNAJMA por cada carrera profesional. El proceso de alimentación de datos que se da cuando inicia un ciclo académico se registra en hojas de cálculo, por lo que responder a preguntas como: ¿Cuál es la evolución de la cantidad de estudiantes por carrera?, ¿Cuánto es la demanda semestral por carrera?, ¿Cuál fue la utilidad por cada semestre de funcionamiento?, ¿Qué colegio aporta más estudiantes?, ¿De cuáles distritos provienen más estudiantes? Y ¿De cuáles distritos provienen menos estudiantes?

Estas preguntas pueden tener respuesta, pero no oportunas, ya que se tendría que analizar los diversos informes entregados por semestres y que fueron archivados físicamente producto de reportes realizadas en hojas de cálculo de Microsoft Excel.

La gestión de las organizaciones contemporánea se apoya en la toma de decisiones rápida y fiable, para reaccionar al mercado y a la competencia. Por lo tanto si no se cuenta con un sistema de inteligencia competitiva como también se le conoce a los sistemas de inteligencia de negocios, se desaprovecha oportunidades que presenta el mercado objetivo u oportunidades que surgen de la sinergia de procesos de la cadena de valor de las organizaciones, El Centro Preuniversitario no es ajeno a esta disyuntiva. Es preciso señalar que el apoyo de un sistema de cubo OLAP, mejoraría la toma de decisiones en el CEPRE de la UNAJMA. La tabla N°01 refleja el ingreso de datos al inicio de cada ciclo de preparación, los atributos de los estudiantes están incompletos sería preciso completar todos los atributos que interesaría para que un modelo de cubo OLAP pueda entrar en funcionamiento.

Tabla N° 1: Registro estudiantes matriculados en el CEPRE

CODIGO	NOMBRES	APE_PATerno	APE_MATERNO	COD_CARRERA	COD_PAGO
70100B	PATRICIA ALEXANDRA	PEÑALOZA	ALARCON	005	01
70101A	RUDY	TAIPE	FLORES	004	01
70102A	YOEL	HUARCAYA	CHOCHOCCA	001	01
70103B	KAROL INGRITH	FARFAN	BERROCAL	005	01
70104B	LIZETT ANAID	PONSECA	ROJAS	005	01
70105B	MARIA MERCEDES	PICHIHUA	VARGAS	006	01
70106B	YENIFER	YAURIS	LLACUA	005	01
70107A	SANDRA SHAILA	ALTAMIRANO	LEGUIA	003	01
70108A	ERIKA	INFANZON	HUAMANI	004	01
70109A	YHON YULISON	MAÑUICO	CUSI	004	01
70110B	THALIA EDITH	QUISPE	ALHUAY	005	01

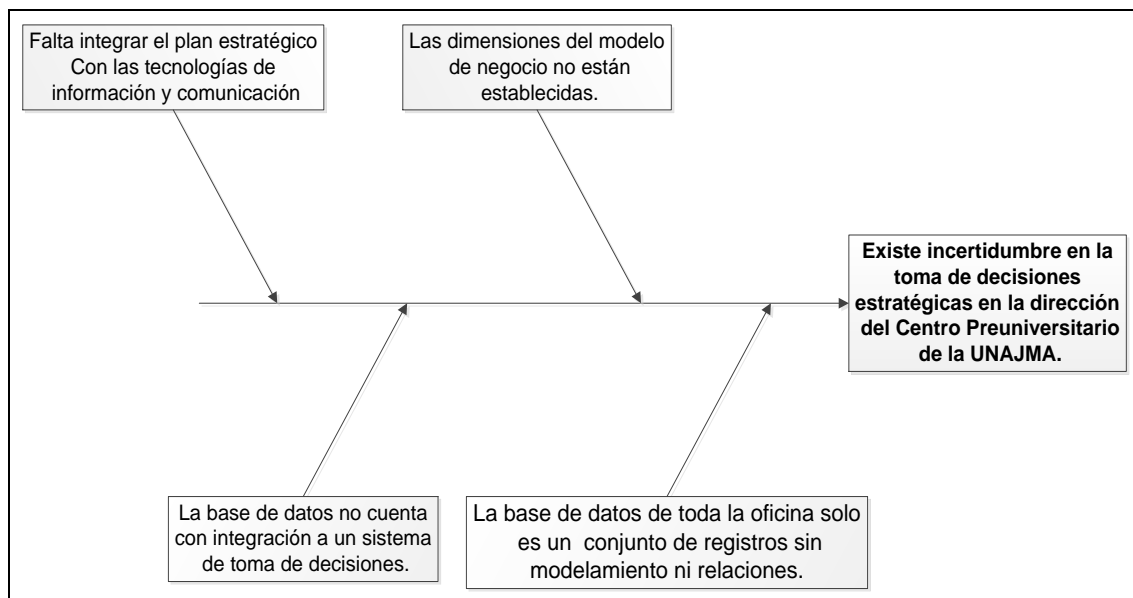
Fuente: Dirección de la oficina de Centro Preuniversitario de la UNAJMA.

No hace falta realizar un análisis más profundo, que la simple inspección, para inferir que los datos que se registran en la oficina del Centro Preuniversitario son insuficientes para que pueda servir de base en la toma de decisiones al mediano y largo plazo, por lo tanto, es preciso recabar y registrar más datos que puedan servir para producir información oportuna y veraz para la posterior toma de decisiones, de continuar con estos registros, será imposible tomar decisiones acertadas, estas estarán condenadas a la incertidumbre.

Se tienen estudios que favorecen el uso e implementación de sistemas de inteligencia competitiva, estos estudios demuestran la vital importancia de hacer uso de estos sistemas en nuestras organizaciones el Centro Preuniversitario no debe ser ajeno a esta tecnología.

Con estas premisas es posible plantear un diagrama de causa-efecto, que permita visualizar el problema del Centro Preuniversitario de la UNAJMA.

Figura N° 1: Diagrama para formular problemas



Fuente: Elaboración propia.

1.2 Formulación del problema

Incertidumbre en la toma de decisiones estratégicas en la dirección del Centro Preuniversitario de la UNAJMA

1.3 Objetivos

Objetivo General

Desarrollar un modelo de cubo OLAP para el soporte de la toma de decisiones estratégicas en el Centro Preuniversitario de la Universidad Nacional José María Arguedas.

1.4 Objetivos Específicos

- Realizar el modelamiento de Base de Datos lógico y físico de las reglas del negocio del centro preuniversitario de la UNAJMA.
- Realizar el cargado de los registros en la base de datos implementada.
- Determinar las dimensiones del modelo y crear el esquema de cubos OLAP.
- Integrar Base de datos, Esquema XML y software Pentaho para obtener el modelo de cubos OLAP.
- Presentación del modelo de cubos OLAP y retroalimentación de sugerencias.

1.5 Justificación

Los datos debidamente registrados, ofrecen un soporte para la toma de decisiones y para posibles innovaciones, la data resulta ser un bien intangible que si no se registra adecuadamente puede causar enormes gastos en el futuro. En el parágrafo de la realidad problemática se presentan estudios que propician la urgente implementación de herramientas de sistemas de inteligencia competitiva de forma que los directivos apelen más a información veraz y oportuna para tomar sus decisiones, que a su intuición y experiencia.

Los directivos siempre están sometidos a tomar decisiones y estas si acertadas van en beneficio en principio de sus propios estudiantes en segunda instancia

de sus directores y tercera instancia de la comunidad universitaria en general , ya que al ser un centro de producción las ganancias va a engrosar las cuentas de los fondos directamente recaudados.

1.6 Viabilidad de la investigación

Desde la perspectiva tecnológica la ejecución de este proyecto es totalmente viable ya que todos los programas son de uso libre, la perspectiva económica involucra más horas de investigación que se está dispuesto a invertir. La información solicitada ha sido brindada ampliamente por el actual director del Centro Preuniversitario.

1.6.1 Viabilidad técnica

Actualmente las tecnologías que dan soporte a los cubos OLAP han llegado a un pleno desarrollo, se pueden conseguir las licenciadas o las tecnologías de uso libre como Schema Workbench, Pentaho, herramientas que usaremos en este desarrollo.

1.6.2 Viabilidad económica

Como se mencionó en el párrafo anterior las herramientas libres no tienen costo y son de libre uso por los investigadores y estudiosos, todo el presupuesto esta

1.6.3 Viabilidad social

El presente estudio será para beneficio la dirección del Centro Preuniversitario, que incidirá en el mejor manejo administrativo de los cientos de estudiantes que a la fecha están matriculados en este centro de estudio.

1.7 Limitación del estudio

La data registrada fue cargada parcialmente debido a la existencia de múltiples formatos de registro usados en los diversos procesos de admisión.

CAPITULO II

MARCO TEÓRICO

2.1 Antecedentes

2.1.1 Antecedentes Internacionales

Las autoras (Arenas Lopez & Gomez Montes, 2017) en su investigación: Inteligencia de negocios aplicada a los procesos de autoevaluación de la universidad de Manizales, plantean como objetivo; Implementar una solución de Inteligencia de Negocios para administrar la información relacionada con los procesos de autoevaluación de la Universidad de Manizales; concluyen:

La Inteligencia de Negocios incorpora metodologías y aplicaciones dirigidas a la administración de información que permiten tomar las mejores decisiones a los usuarios de una organización.

Los procesos de la Inteligencia de Negocios presenta importantes beneficios enfocados a la creación de conocimiento por medio de la información suministrada de los datos. Es decir, las organizaciones optimizan la utilización de sus recursos, monitorean el cumplimiento de sus procesos y propósitos, y obtienen mejores resultados.

En los últimos años se han desarrollado tecnologías y herramientas que ayudan a las organizaciones a poder extraer y analizar la información siendo más proactivas y ágiles al momento de evaluar los resultados. Es por esto, que las soluciones de Inteligencia de Negocios suelen ser más atractivas debido al manejo visual de la información administrada.

Las habilidades de la inteligencia de negocios permiten gestionar la información de forma rápida, flexible y oportuna de los procesos internos más importantes de las organizaciones, así mismo la renovación constante de los datos. Es así, como una institución educativa puede enfrentarse a estos procesos para manejar su

información, y así involucrarse en las ventajas o inconvenientes que se presenten en la medida de sus intereses.

En el informe se puede apreciar el desarrollo de una herramienta tecnológica basada en inteligencia de negocios como producto final. Apoyado en Mysql Workbench y Pentaho

En el estudio: Research inteligencia de negocios: estudio de caso sector tecnológico colombiano los investigadores (Gutierrez camelo, Devia Llanos, & Tarazona Bermudez, 2016) pudieron concluir que el 96% de un total de 90 empresas dedicadas al rubro tecnológico conocen las herramientas de inteligencia de negocios y están en proceso de implementación nos preguntamos si en el Perú nuestra empresas estarán en este proceso y cómo será el análisis para el sector público.

Sobre inteligencia competitiva se han realizado múltiples investigaciones que propician el desarrollo de esta área, Así por ejemplo:

(Duro Novoa & Pérez Cuevas, 2016), señalan que, se hace muy necesario contar con un sistema de información que valiéndose de las tecnologías de información y las comunicaciones, gestione e integre la información adecuada, de forma eficaz y eficiente, en el momento oportuno y por la vía más rápida y factible, con vistas a lograr el desenvolvimiento de las funciones, y la toma de decisiones apropiadas (sección de resumen, párr.2).

Para el estudio de (Vélez Miranda & Martínez Martínez, 2014) se distingue una relación de la inteligencia competitiva con el proceso de innovación en las Pymes, Así el autor demuestra, que:

El subproceso de análisis de los datos tiene una relación estadísticamente significativa con la capacidad de innovación de la empresa. Se destaca la importancia que tiene esta relación debido a que si los datos no son procesados correctamente, buscando patrones, anomalías o tendencias, la empresa pierde la capacidad de innovación.

Los resultados concuerdan y apoyan a los teóricos Tanev & Bailetti (2008) que indican que las pequeñas y medianas empresas, para ser exitosas, deben tener una amplia gama de información que proviene del proceso de inteligencia competitiva. Además, los hallazgos del estudio se ajustan al estudio de Buchda (2007) quien describe que la práctica del proceso de inteligencia competitiva en las pequeñas empresas, está fundamentada en la capacidad del manejo o diseminación de la información. La habilidad de la empresa de manejar la información y el conocimiento constituye el factor esencial para el éxito del negocio (sección Hipótesis: capacidad de innovación párr. 3 y 4).

Se explica que con la implementación de sistemas de inteligencia competitiva se favorece la innovación en las organizaciones; otro estudio sobre el impacto de la inteligencia de negocios en las empresas peruanas, en principio caracteriza las variables que se debe tener en cuenta al implementar sistemas de inteligencia competitiva y luego muestra los factores a tener en cuenta para evaluar el impacto de los sistemas de inteligencia competitiva,

2.1.2 Antecedentes Nacionales

(González López , 2011), señala:

Se trata de responder a la siguiente pregunta de investigación:
¿Qué componentes tienen los principales factores que influyen en el éxito del sistema de DW y BI en el caso de un país en vías de desarrollo? La que se descompone en las siguientes preguntas:
¿cuáles son las variables más relevantes?, y ¿cuáles son los principales componentes de las mismas?, que impactan en el desempeño de la Data Warehouse e Inteligencia de Negocios en un país en vías de desarrollo.

De esta manera tenemos un procedimiento que nos da la certeza de obtener información válida y consistente en la investigación, y

con la cual determinamos que los constructos o conceptos, y sus componentes, más relevantes: Calidad de la Información, Uso del sistema de DW y BI, Calidad del Sistema, Calidad del Servicio y Satisfacción del Usuario.

(Pacco Palomino & Apaza Mamani, 2016), Implementaron la herramienta Pentaho BI en el área financiera de la Universidad Peruana Unión, desde la extracción de los datos hasta mostrar los reportes su segunda conclusión y quizás la más interesante, expresa:

Con la creación de un nuevo datamart (DATAMART_UPEU), y la elaboración de BI (Business Intelligence) se ha podido lograr la optimización para las consultas multidimensionales, facilitando el manejo dinámico de los reportes. Las operaciones de navegación son bastante flexibles, donde los usuarios finales interactúan y la herramienta muestra la información requerida para la toma de decisiones. Se logró identificar los indicadores de la unidad estratégica académica, los cuales fueron capturados de la base de datos relacional, según la necesidad de información financiera de la Gerencia de la Universidad Peruana Unión.

(Matamoros Manrique, 2015), plantea el objetivo de Determinar el efecto de la implementación de un Sistema OLAP, para la reducción de los riesgos en la toma de decisiones, respecto a la educación, en la Unidad de Estadística Educativa del Ministerio de Educación del Perú - UGEL 01 – Gestión 2014. Y concluye:

Se concluye con el Sistema OLAP se aumenta la cantidad de información que puede obtener la Unidad de Estadística Educativa del Ministerio de Educación del Perú - UGEL 01 - Gestión 2014, tal como se observa en los resultados de la investigación. De hecho con el nuevo Sistema OLAP, en promedio los reportes de interés que se pueden obtener se incrementaron de 18.17 a 132.96.

(Salazar Tataje, 2017), en su trabajo titulado: Implementación de inteligencia de negocios para el área comercial de la empresa Azaleia – basado en metodología ágil Scrum, concluye:

En la solución se ha aprovechado todas las capacidades de las herramientas Qlikview, para la explotación de la información, aumentando mayor tiempo en el porcentaje de análisis, basado en las encuestas, de la información que en el propio desarrollo de esta. Además se ha demostrado un mejor acceso a la información, al cubo OLAP, los diseño de reportes, creación de tableros de control, alerta proactivas y el acceso vía portal web a toda la información.

No obstante se buscó disponer de forma atractiva e interactiva la información, así como facilitar su elaboración por parte de los usuarios. Actualmente la empresa invierte entre 3 a 4 días en la elaboración de reportes y graficas que viene a ser el 50% del tiempo laboral que emplean. Luego de la implementación de BI, este tiempo ha sido reducido a horas (4 horas), las cuales están empleadas para la reportaría afianzando el tiempo para el análisis de los indicadores.

2.2 Inteligencia de negocios.

Es un término utilizado tanto por los distribuidores de hardware y software como por los consultores de tecnología de la información para describir la infraestructura para almacenar, integrar, crear informes y analizar los datos que provienen del entorno de negocios. La infraestructura de la base recolecta, almacena, limpia y pone la información relevante a disposición de los gerentes. Por ende, la mera esencia de la inteligencia y el análisis de negocios consiste en integrar todos los flujos de información producidos por una firma en un

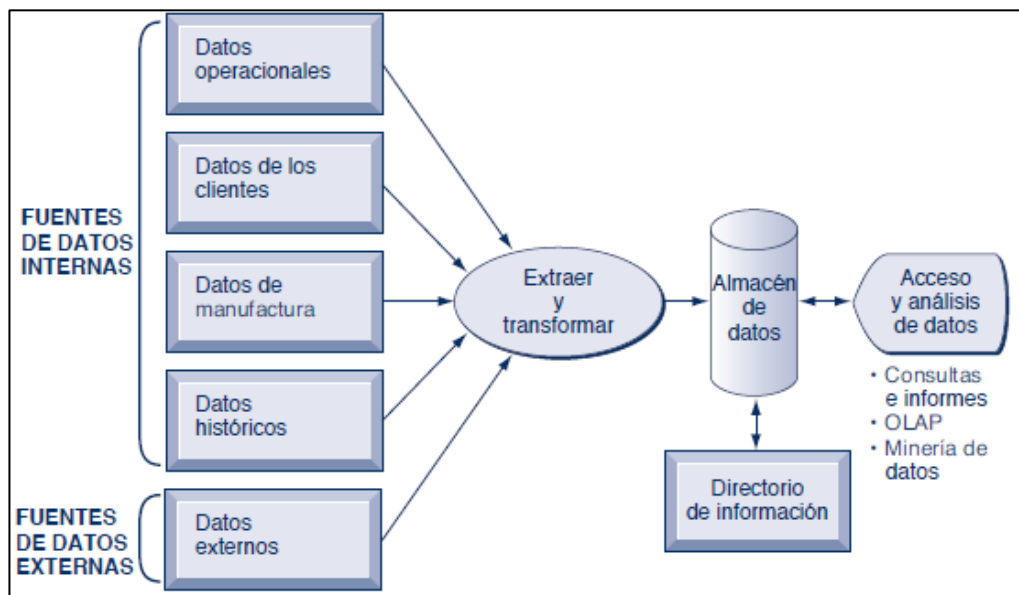
solo conjunto de datos coherente a nivel empresarial, para después, mediante el uso del modelado, la herramientas de análisis estadístico (como distribuciones normales, correlación y análisis de regresión, análisis de ji-cuadrado, pronósticos y análisis de grupos), tratar de comprender todos estos datos de modo que los gerentes puedan tomar mejores decisiones y realizar mejores planes, o por lo menos que sepan con rapidez cuando sus firmas no están cumpliendo los objetivos planeados. (Laudon & Laudon, 2012)

2.2.1 Almacén de datos

Es una base de datos que almacena la información actual e histórica de interés potencial para los encargados de tomar decisiones en la compañía. Los datos se originan en muchos sistemas de transacciones operacionales básicos, como los sistemas de ventas, las cuentas de clientes, la manufactura, y pueden incluir datos de transacciones de sitios Web. El almacén de datos consolida y estandariza la información de distintas bases de datos operacionales, de modo que se pueda utilizar en toda la empresa para el análisis gerencial y la toma de decisiones. (Laudon & Laudon, 2012)

Por lo general esta terminología es la traducción del término anglosajon muy utilizado Data Warehouse.

Figura N° 2: Componentes de un almacén de datos



Fuente: (Laudon & Laudon, 2012)

Características del almacén de datos.

Orientado a un área

Significa que cada parte del Datawarehouse está construida para resolver un problema de negocio, que ha sido definido por los tomadores de decisiones. Por ejemplo: Entender los hábitos de compra de nuestros clientes, analizar la calidad de nuestros productos, analizar la productividad de una línea de fabricación, etc. Para poder analizar un problema de negocio necesitamos información que proviene de distintos sistemas y la organizamos entorno a áreas: ventas, clientes, elementos de transporte, etc. Provee a los tomadores de decisiones de una visión completa y concisa sobre una problemática de negocio, obviando toda aquella información que no necesitan para la toma de decisiones. (Luis Cano, 2010)

Integrado

La información debe ser transformada en medidas comunes, códigos comunes y formatos comunes para que pueda ser útil. La integración permite a las organizaciones implementar la estandarización de sus definiciones, por ejemplo: La moneda en la que están expresados los importes es común. (Lluís Cano, 2010)

Indexado en el tiempo

Significa que se mantiene la información histórica y se almacena referida a determinadas unidades de tiempo, tales como horas, días, semanas, meses, trimestres o años. Ello nos permitirá analizar, por ejemplo, la evolución de las ventas en los periodos que queramos. (Lluís Cano, 2010)

No volátil

Significa que los usuarios no la mantienen, como lo harían en los entornos transaccionales. La información se almacena para la toma de decisiones. No se va actualizando continuamente, sino periódicamente, de forma preestablecida. (Lluís Cano, 2010)

2.2.2 Mercado de datos

Es un subconjunto de un almacén de datos, en el cual se coloca una porción con alto grado de enfoque en los datos de la organización en una base de datos separada para una población específica de usuarios. Por ejemplo, una compañía podría desarrollar mercados de datos sobre marketing y ventas para lidiar con la información de los clientes. Antes de implementar un almacén de datos a nivel empresarial, la librería Barnes & Noble mantenía una serie de mercados de datos: uno para los datos sobre los puntos de venta en las tiendas minoristas, otro para las ventas de las librerías universitarias y un tercero para las ventas en línea. Por lo general, un mercado de datos se enfoca en un solo tema o línea de negocios, por lo que es común que se construya con más

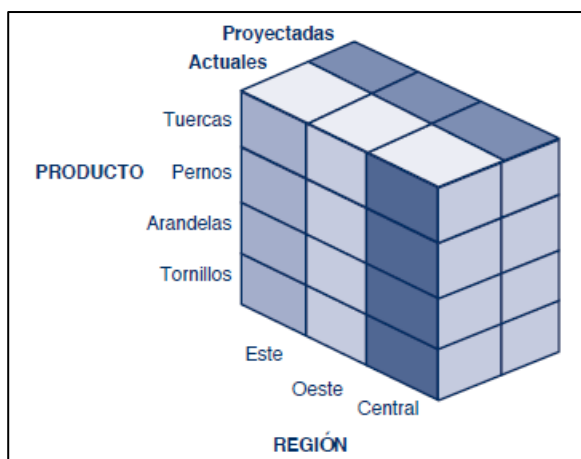
rapidez y a un menor costo que un almacén de datos a nivel empresarial. (Laudon & Laudon, 2012)

2.2.3 Cubos OLAP (On-line Analytical Processing)

Estas herramientas manejan una serie de consultas de forma interactiva sobre estructuras multidimensionales (Cubos OLAP) cargadas previamente con los datos almacenados en las bases de datos corporativas tradicionales. Permiten realizar informes y obtener grandes cantidades de información a partir de lo que resultaría ser a modo rutinario una serie de complejas consultas sobre una base de datos de forma sencilla. Con estos sistemas es posible analizar la información almacenada en un datawarehouse, pero no es estrictamente necesario, ya que la información puede provenir de diferentes bases de datos. El objetivo de estas herramientas es obtener una mejor comprensión de lo almacenado en las bases de datos. (Navarro Ruiz, 2013).

La figura N°3, muestra un modelo multidimensional que podría crearse para representar productos, regiones, ventas reales y ventas proyectadas. Una matriz de ventas actuales se puede apilar encima de una matriz de ventas proyectadas para formar un cubo con seis caras. Si gira el cubo 90 grados en un sentido, la cara que se muestre será la de producto contra ventas actuales y proyectadas, si lo gira 90 grados de nuevo, verá la cara de región contra ventas actuales y proyectadas, si lo gira 180 grados a partir de la vista original, verá las ventas proyectadas y producto contra región. Se pueden anidar cubos dentro de otros cubos para crear vistas complejas de datos. Una compañía podría utilizar una base de datos multidimensional especializada una herramienta que cree vistas multidimensionales de datos en las bases de datos relacionales. (Laudon & Laudon, 2012)

Figura N° 3: modelo multidimensional que podría crearse para representar productos, regiones, ventas reales y ventas proyectadas.



Fuente: (Laudon & Laudon, 2012)

2.3. Data Warehouse

Los conceptos aquí vertidos de Data Warehouse, sus características y esquemas fueron resumidos de los autores (Curto Díaz & Conesa, 2011) del libro Introducción al Business Intelligence.

Data Warehouse es un almacén de datos orientado a diferentes ámbitos, integrado, no volátil y variable en el tiempo que facilita tener una visión general de los datos de una organización facilitando tener la información estable, coherente, fiable e información histórica.

2.3.1 Características

Orientado a temas. La clasificación de la información se hace en base a los aspectos que son de interés para la organización y teniendo en cuenta procesos orientados a las aplicaciones; se orienta al modelamiento de datos y al diseño de la base de datos. Las principales áreas de los temas influyen en la parte más importante de la estructura como puede ser de universidad,

facultades, programas, estudiantes. La información que no será usada por el proceso de sistemas de soporte de decisiones se excluye, mientras que la información de las aplicaciones contiene datos para satisfacer de inmediato los requerimientos funcionales y de proceso, que pueden ser usados o no por el analista.

Integrado. Integra todos los datos de diferentes sistemas operacionales de la organización y fuentes externas presentando consistencia, por lo que las inconsistencias se deben eliminar. La información se estructura en niveles para adecuarse a las distintas necesidades.

Variable en el tiempo. Los datos corresponden a un periodo de tiempo y estos deben ser integrados periódicamente, los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se generen con tales variaciones.

No volátil. La información no se modifica ni se elimina, una vez almacenado un dato, este se convierte en información de solo lectura y se guarda para futuras consultas.

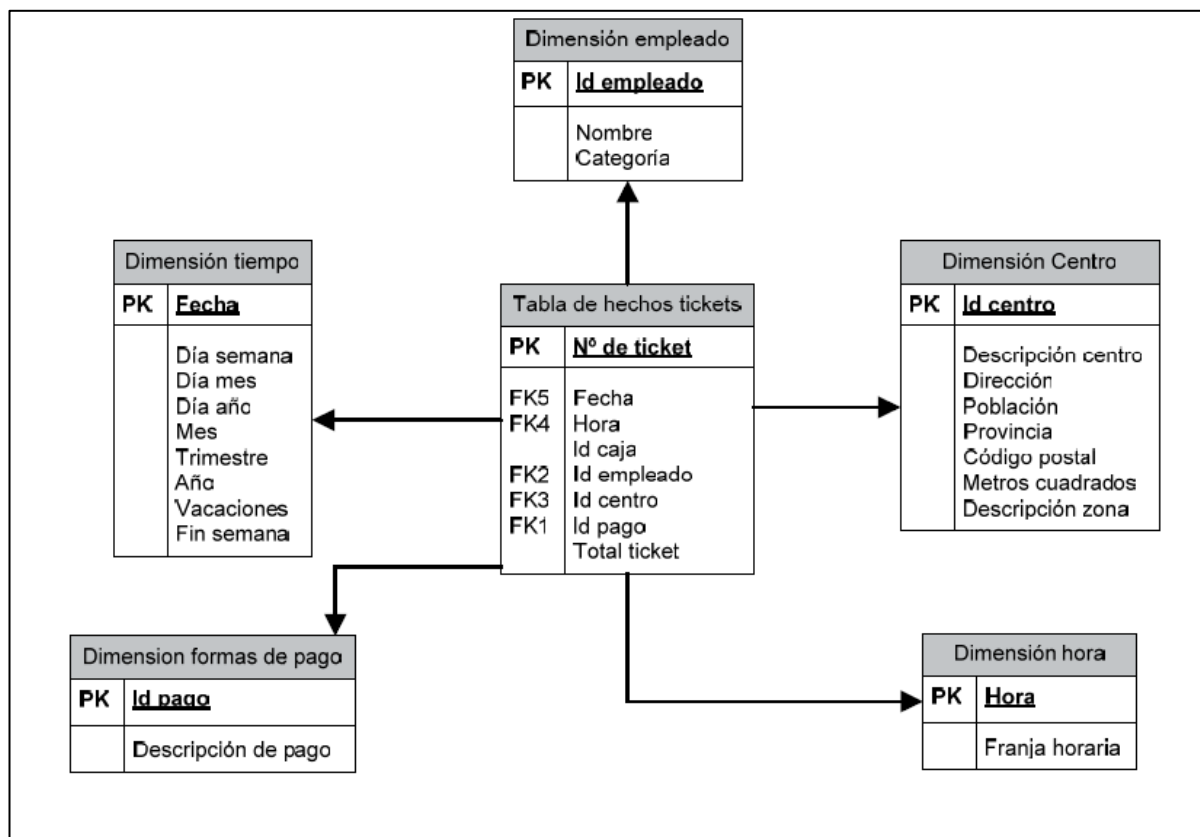
2.3.2 Esquemas

- **Estrella.** Contiene una tabla principal de hechos y algunas otras tablas de los procesos de la organización que tienen relación con la tabla de hechos; esta está ubicada en el centro y las demás tablas se ubican alrededor formando una estrella.

Para la construcción del esquema “estrella” debemos distinguir entre las tablas de hechos (aquello que queremos medir o analizar) y las tablas de dimensiones (cómo lo queremos medir), en nuestro caso, la tabla de hechos será la de los tickets y los queremos analizar por las dimensiones siguientes: tiempo, franja

horaria, centro, empleado y forma de pago. (Lluís Cano, 2010) El esquema “estrella” sería:

Figura N° 4: Esquema estrella de los comprobantes del supermercado



Fuente: (Lluís Cano, 2010)

Dimensión tiempo:

Algunos de los atributos de esta tabla no los teníamos en el modelo original de nuestra base de datos. El atributo “Fecha” lo hemos descompuesto en: Día semana, Día mes, Día año, Mes, Trimestre, Año, Vacaciones y Fin semana. Esta descomposición nos permitirá analizar si vendemos lo mismo todos los días de la semana o no, o si se vende más a principios o finales de mes, comparar entre los distintos meses, trimestres o incluso años. También hemos añadido el atributo: “Vacaciones” que nos permitirá saber si el día es festivo, y el atributo: “Fin semana” para poder diferenciar las ventas durante la semana y las de fin de semana. Esta información la deberemos añadir, ya que no está

disponible en el sistema de información de los tickets de venta.
(Lluís Cano, 2010)

Dimensión hora:

Nos ha parecido interesante analizar las ventas de las distintas franjas horarias. Hemos dividido la jornada en cuatro franjas horarias: de 9:00 a 11:59, de 12:00 a 14:59, de 15:00 a 17:59 y de 18:00 a 21:00, lo que nos permitirá saber en qué franjas tenemos más ventas. (Lluís Cano, 2010)

Dimensión formas de pago:

Esta dimensión nos permite conocer cómo nos pagan nuestros clientes: al contado, tarjeta de débito o tarjeta de crédito. (Lluís Cano, 2010)

Dimensión empleado:

Esta dimensión nos permite conocer los tickets de venta que han sido cobrados por un empleado o una categoría de empleados. (Lluís Cano, 2010)

Dimensión centro:

Con esta dimensión podremos analizar cuál es el importe de los tickets de venta de un centro, los centros de una población o de una provincia, e incluso ordenarlos por código postal. La dimensión también nos permitirá analizar el importe de los tickets de venta en función de los metros cuadrados de los centros. En la dimensión centro hemos incorporado "Descripción zona", que nos indica a qué zona está asignado el centro y nos permitirá analizar las diferencias entre las distintas zonas. (Lluís Cano, 2010)

Copo de nieve. Es una derivación del modelo estrella, la tabla de hechos deja de ser la principal y las demás tablas están relacionadas con otras tablas sin tener una relación directa con la tabla de hechos. Este esquema hace que la extracción de datos sea más difícil por lo que las tablas se vinculan con más tablas.

En el esquema “copo de nieve” aparecen relaciones entre las tablas de dimensiones, mientras que en el esquema “estrella” sólo hay relaciones entre la tabla de hechos y las de dimensiones.

En este caso, las tablas de dimensiones están totalmente normalizadas, lo que reduce el espacio que ocupan, aunque en algunos casos esta diferencia no es significativo. (Lluís Cano, 2010)

2.4 Cubos de datos

Una de las formas más comunes para analizar la información es a través de cubos OLAP (On-Line Analytical Processing), procesamiento analítico online, encargados de estructuras multidimensionales que permiten analizar bases de datos relacionales de gran volumen y variedad con gran agilidad y rapidez reduciendo el tiempo. Los cubos OLAP como bases de datos multidimensionales, las tablas, en concreto sus registros, proporcionan las medidas de los cubos, y sus dimensiones vienen determinadas por las dimensiones de los cuadros con los que cuenta cada tabla. Se debe tener en cuenta que los cubos no sustituyen a las tablas relacionales; simplemente ayudan a mejorar la eficiencia, la velocidad de respuesta y la simplicidad de las consultas. (Curto Díaz & Conesa, 2011)

2.5 toma de decisiones

Muchos gerentes de negocios operan en un banco de niebla de información, sin nunca tener realmente los datos correctos en el momento oportuno para realizar una decisión informada. En lugar de eso, los gerentes dependen de las proyecciones, los mejores planteamientos y la suerte. El resultado es una producción excesiva o baja de bienes y servicios, una mala asignación de los recursos y de los tiempos de respuesta deficientes. Estos resultados negativos elevan los costos y provocan la pérdida de clientes. En la década anterior, los sistemas y tecnologías de información hicieron posible que los gerentes usaran datos en tiempo real provenientes del mercado a la hora de tomar decisiones. (Laudon & Laudon, 2012)

2.5.1 Diversas definiciones de toma de decisiones

Es un corte entre el pasado y el futuro. (Schackle, G, 1961)

El proceso de transformación de la información en acción. (Forrester, 1992)

Identificar y resolver los problemas que se le presenta a toda organización. (Le Moigne, 1974)

Es el proceso de definición de problemas, recopilación de datos, generación de alternativas y selección de un curso de acción (Slocum, 2004)

El proceso para identificar y solucionar un curso de acción para resolver un problema específico (Wankel, 1989)

Resolución de problemas mediante los cuales tenemos que elegir la mejor decisión o alternativa. Para ello previamente debemos realizar un diagnóstico en primer lugar y una búsqueda precisa en segundo lugar (Greenwood, 1969)

2.5.2 Patrones y reglas para la toma de decisiones

Los patrones y reglas se utilizan para guiar la toma de decisiones y pronosticar el efecto de esas decisiones. (Lara Torralbo, Minería de datos, 2014)

- Las *asociaciones* son ocurrencias vinculadas a un solo evento. Por ejemplo, un estudio de los patrones de compra en supermercados podría revelar que, cuando se compran frituras de maíz, el 65 por ciento del tiempo se compra un refresco de cola, pero cuando hay una promoción, el 85 por ciento se compra un refresco de cola. Esta información ayuda a los gerentes a tomar mejores decisiones debido a que descubren la rentabilidad de una promoción.

- En las *secuencias*, los eventos se vinculan en el transcurso del tiempo. Por ejemplo, podríamos descubrir que si se compra una casa, el 65 por ciento del tiempo se compra un nuevo refrigerador dentro de las siguientes dos semanas, y el 45 por ciento se compra un horno dentro del mes posterior a la compra de la casa.

- La *clasificación* reconoce los patrones que describen el grupo al que pertenece un elemento, para lo cual se examinan los elementos existentes que hayan sido clasificados y se infiere un conjunto de reglas. Por ejemplo, las empresas como las compañías de tarjetas de crédito o las telefónicas se preocupan por la pérdida de clientes estables. La clasificación ayuda a descubrir las características de los clientes con probabilidades de dejar de serlo y puede proveer un modelo para ayudar a los gerentes a predecir quiénes son esos clientes, de modo que puedan idear campañas especiales para retenerlos.

- El *agrupamiento* funciona de una manera similar a la clasificación cuando aún no se han definido grupos. Una herramienta de minería de datos puede descubrir distintas agrupaciones dentro de los datos, como el hecho de encontrar grupos de afinidad para tarjetas bancarias o

particionar una base de datos en grupos de clientes con base en la demografía y los tipos de inversiones personales.

Aunque estas aplicaciones implican predicciones, el *pronóstico* utiliza las predicciones de una manera distinta. Se basa en una serie de valores existentes para pronosticar cuáles serán los otros valores. Por ejemplo, el pronóstico podría encontrar patrones en los datos para ayudar a los gerentes a estimar el futuro valor de variables continuas, como las cifras de ventas.

2.5.3 Tecnologías de información en la toma de decisiones

Hay una interdependencia cada vez mayor entre la habilidad de una empresa de usar la tecnología de la información y su destreza para implementar estrategias corporativas y lograr los objetivos corporativos. Lo que una empresa quiera hacer en cinco años depende a menudo de lo que sus sistemas serán capaces de realizar. Aumentar la participación en el mercado, convertirse en el productor de alta calidad o bajo costo, desarrollar nuevos productos e incrementar la productividad de los empleados son procesos que dependen cada vez más en los tipos y la calidad de los sistemas de información en la empresa. Cuanto más comprenda usted esta relación, más valioso será como gerente. En específico, las empresas de negocios invierten mucho en sistemas de información para lograr seis objetivos de negocios estratégicos: excelencia operacional; nuevos productos, servicios y modelos de negocios; intimidad con clientes y proveedores; toma de decisiones mejorada; ventaja competitiva, y sobrevivencia. (Laudon & Laudon, 2012)

2.5.4 Situaciones para implementar cubos OLAP

Existen situaciones en las que la implantación de un sistema de Cubos OLAP resulta adecuada. Destacamos, entre todas las que existen:

- La toma de decisiones se realiza de forma intuitiva en la organización.
- Identificación de problemas de calidad de información.
- Uso de Excel como repositorios de información corporativos o de usuario. Lo que se conoce como Excel caos¹.
- Necesidad de cruzar información de forma ágil entre departamentos
- Evitar silos de información.
- Las campañas de marketing no son efectivas por la información base usada.
- Existe demasiada información en la organización para ser analizada de la forma habitual. Se ha alcanzado la masa crítica de datos.
- Es necesario automatizar los procesos de extracción y distribución de información. (Curto Díaz & Conesa, 2011)

Tal y como Thomas Davenport en su libro “Competing on Analytics”, una nueva forma de estrategia competitiva está emergiendo basada en el uso de la estadística descriptiva, modelos productivos y complejas técnicas de optimización, datos de alta calidad y una toma de decisiones basada en hechos. En dicho contexto, la inteligencia de negocio es el paso previo para dicha estrategia dado que ayuda a sentar las bases para su futuro despliegue. (Curto Díaz & Conesa, 2011)

¹ Se entiende como Excel caos el problema resultante del uso intensivo de Excel como herramienta de análisis. Cada usuario trabaja con un fichero personalizado. Como resultado, la información no cuadra entre departamentos y el coste de sincronización es sumamente elevado.

CAPÍTULO III

MATERIALES Y MÉTODOS

3.1 Esquema de propuesta de solución

El esquema de solución presenta de forma resumida los objetivos que se quieren alcanzar en el presente estudio.

Tabla N° 2: Esquema de propuesta de solución.

Desarrollar un modelo de cubos OLAP que sirva de soporte a la toma de decisiones estratégicas en el Centro Preuniversitario de la Universidad Nacional José María Arguedas.				
ETAPA N°01 Realizar el modelamiento de Base de Datos lógico y físico de las reglas del negocio del centro preuniversitario de la UNAJMA.	ETAPA N°02 Realizar el cargado de los registros en la base de datos implementada.	ETAPA N°03 Determinar las dimensiones del modelo y crear el esquema de cubos OLAP.	ETAPA N°04 Integrar Base de datos, Esquema XML y software Pentaho para obtener el modelo de cubos OLAP.	ETAPA N°05 Presentación del modelo de cubos OLAP y retroalimentación de sugerencias.

Fuente: Elaboración propia.

3.1.1 Modelamiento de la Base de datos del CEPRE-UNAJMA.

El Centro preuniversitario realiza sus operaciones bajo un conjunto de reglas, que modelan sus procedimientos, en ocasiones se rigen bajo reglamentos aprobados o en todo caso por reglas convencionales, que tienen que ser enumeradas a través de entrevistas a los miembros de la dirección y personal que labora en la oficina.

La herramienta tecnológica que nos apoyará para realizar el modelamiento será el software Mysql Workbench, se considerará las reglas de normalización para tener integridad de datos y evitar

redundancia de datos. El modelo obtenido considerando entidades y atributos del Centro Preuniversitario estará relacionado bajo el esquema de entidad-relación.

3.1.2 Cargado de registros a la base de datos

De acuerdo al trabajo previo realizado en el Centro Preuniversitario, se verificó que se dispone datos registrados en periodos anteriores en formato xls, que tienen una variedad de campos, esta fue organizada según el modelo entidad relación y cargada a la base de datos implementada en la etapa previa.

3.1.3 Dimensiones del cubo OLAP

La base de datos que sirve de soporte para la construcción de cubos OLAP tiene una tabla de hechos y tres dimensiones: procedencia de estudiantes, ingresos y egresos por ciclo académico, además, demanda de carreras por ciclo, este esquema corresponde a un esquema de tipo estrella que se construyó en base al modelo entidad relación.

3.1.4 Esquema en formato XML

El esquema fue diseñado en formato XML considerando: jerarquías, dimensiones y métricas bajo el esquema XML. Una vez diseñado considerando las dimensiones y la tabla de hechos se publica en el Software que permitirá su publicación. Toda esta etapa será elaborada con ayuda del software Schema Workbench.

3.1.5 Presentación de propuesta y retroalimentación de sugerencias

Para la presentación de la propuesta de solución se debe realizar la configuración del software PENTHO, en esta etapa los usuarios podrán

realizar sugerencias de manera que se pueda considerar para una retroalimentación y la capacitación del uso de la herramienta de cubos OLAP.

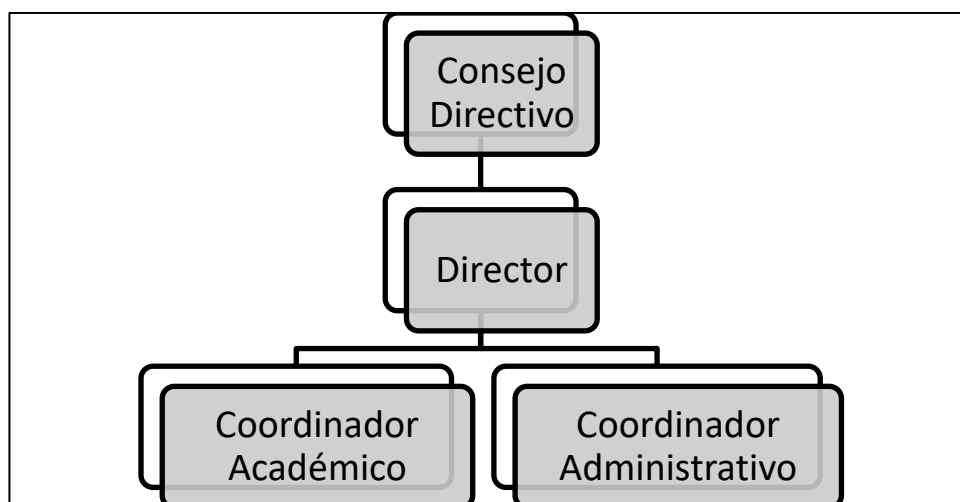
3.2 Construcción de cubos OLAP para el Centro Preuniversitario

3.2.1 Reglas del negocio del Centro Preuniversitario

Estas reglas de negocio en su mayoría están establecidas en el reglamento del Centro Preuniversitario que se aprueba con acto resolutivo cada año con una evaluación y retroalimentación de los reglamentos pasados.

La dirección del Centro Preuniversitario de la UNAJMA tiene tres ciclos de preparación durante un año denominados ciclo uno, dos y tres el primero de ellos inicia en el mes de Enero y culmina en Marzo con una duración de 8 semanas el segundo ciclo inicia el mes de Mayo y culmina el mes de Julio con una duración de diez semanas el tercer ciclo inicia el mes de Setiembre y culmina el mes de Noviembre con una duración de 10 semanas. Los responsables del Centro Preuniversitario están organizados del siguiente modo:

Figura N° 5: Organigrama del Centro Preuniversitario de la UNAJMA



Fuente: Oficina del centro Preuniversitario

Los alumnos que se presentan a través del Centro Preuniversitario (CEPRE) de la universidad también son considerados como postulantes dentro de la

modalidad de examen extraordinario. A pesar de que ellos tienen sus propias reglas para rendir su examen.

Los ingresos directamente recaudados por el Centro Preuniversitario tienen la siguiente escala:

1.- Estudiante regular:

Tabla N° 3: Pagos estudiante regular

ITEM	DESCRIPCION	MONTO S/.
1.	Matricula	20.00
2.	Carne de estudiante	10.00
3.	Pensión de enseñanza	430.00
4.	Simulacro	40.00
Total		500.00

Fuente: centro Preuniversitario

Los estudiantes realizan el abono de estos cuatro ítems por todo el periodo de tiempo que dura el ciclo.

2. Estudiante pago especial (dos partes)

Tabla N° 4: Pagos en partes

ITEM	DESCRIPCION	MONTO S/.	PAGOS
01	Matricula	20.00	380.00
02	Carné de estudiante	10.00	
03	1° Pensión de enseñanza (al matricularse)	310.00	
04	Exámenes sumativos	40.00	
05	2° Pensión de enseñanza (hasta finalizar la 5ta semana)	170.00	170.00
Total			550.00

Fuente: centro Preuniversitario

En caso de pago que se divide en dos partes los montos recaudados se incrementan debido a que los estudiantes pagan la primera parte y luego en algunos casos existe deserción estudiantil.

3. Estudiantes con media beca (Escasos recursos, violencia política, hijos de trabajadores de la UNAJMA, segundo hermano, etc.)

Tabla N° 5: Pago estudiantes con media beca.

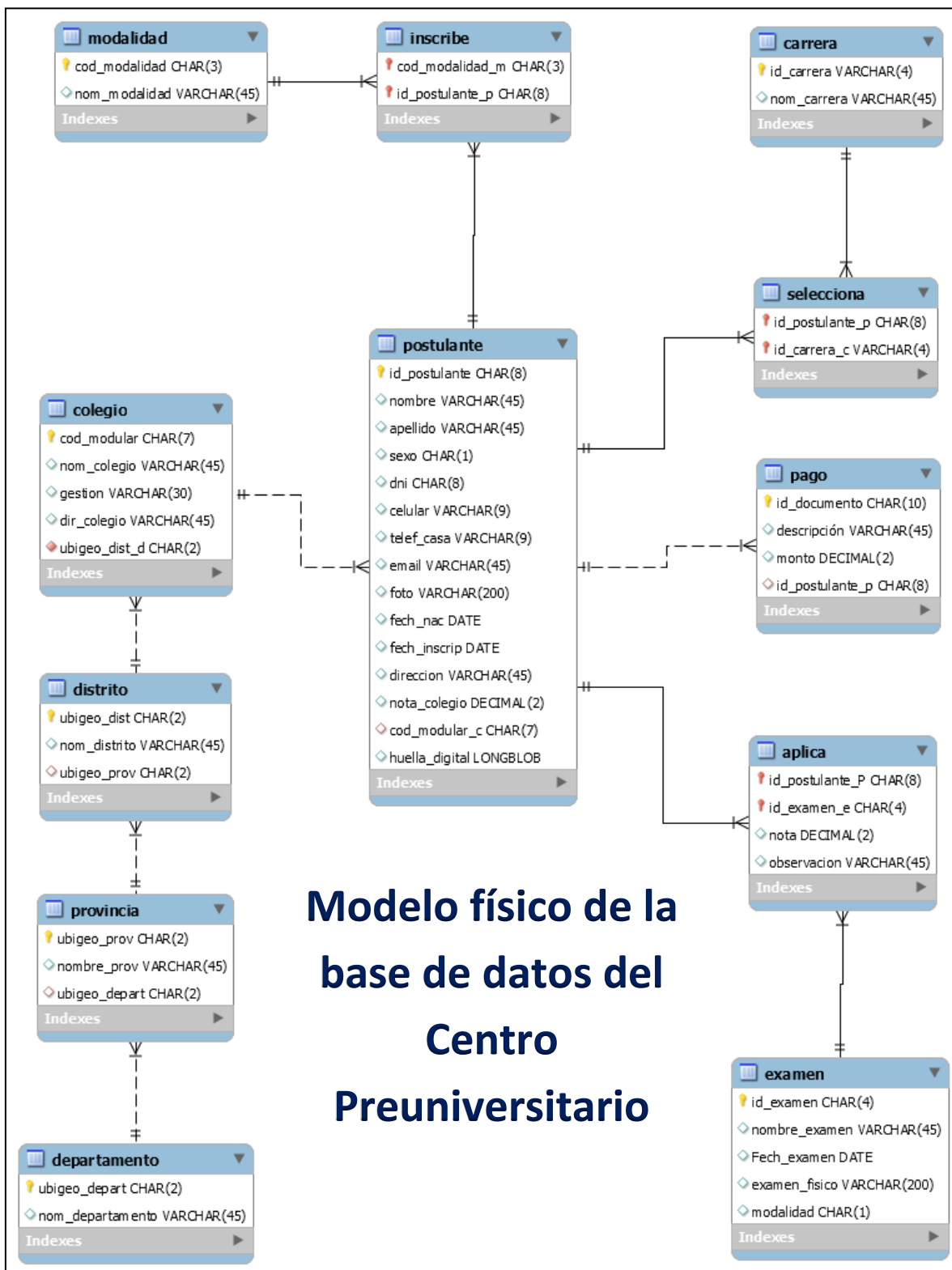
ITEM	DESCRIPCIÓN	MONTO S/.
01	Matrícula	20.00
02	Carné de estudiante	10.00
03	Simulacro	40.00
04	Pensión de enseñanza	250.00
Total		320.00

Se aplica a los estudiantes que son hijos de los trabajadores administrativos y docentes de la Universidad Nacional José María Arguedas.

Para medir la eficiencia y eficacia de la gestión financiera es preciso poder establecer todos ingresos captados por esta vía debido a que esta fuente de ingresos constituye una de las fuentes de recursos directamente recaudados (RDR) de la UNAJMA.

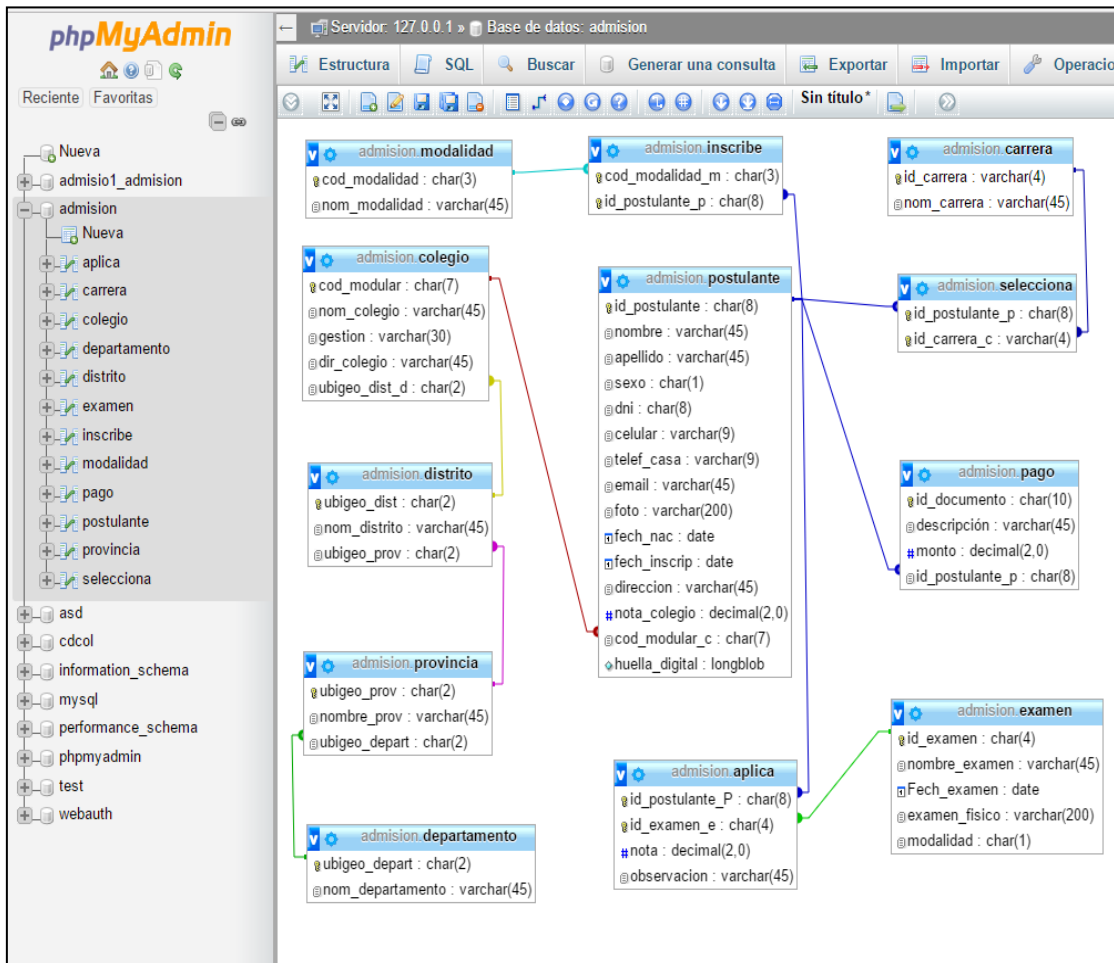
3.2.2 Modelo Físico de la base de datos del Centro Preuniversitario de la UNAJMA

Figura N° 6: Modelo físico de la base de datos del Centro preuniversitario.



Este modelo diseñado en el software Mysql Workbench se exporta a un archivo de extensión (.log) que a su vez será importado por el software Mysql, obteniéndose el siguiente diagrama relacional de base de datos:

Figura N° 7: Implementación de la base de datos en el SABD Mysql



Fuente: Elaboración propia

Manual de referencia del Mysql tomado de: (Oracle and/or its affiliates, 2014)

Insertar registros en una tabla procedentes de otra tabla.

Nota previa: la base de datos original del sistema de admisión cuenta con una tabla independiente o sin relacionar con la siguiente estructura y con demasiada redundancia además de muchos registros con información nula podemos apreciar un extracto en la siguiente figura:

Tabla N° 6: Tabla en formato XLS del Centro preuniversitario.

CODIGO	NOMBRES	APE_PATerno	APE_MATERNO	COD_CARRERA	COD_PAGO
70100B	PATRICIA ALEXANDRA	PEÑALOZA	ALARCON	005	01
70101A	RUDY	TAIPE	FLORES	004	01
70102A	YOEL	HUARCAYA	CHOCHOCCA	001	01
70103B	KAROL INGRITH	FARFAN	BERROCAL	005	01
70104B	LIZETT ANAID	PONSECA	ROJAS	005	01
70105B	MARIA MERCEDES	PICHIHUA	VARGAS	006	01
70106B	YENIFER	YAURIS	LLACUA	005	01
70107A	SANDRA SHAILA	ALTAMIRANO	LEGUIA	003	01
70108A	ERIKA	INFANZON	HUAMANI	004	01
70109A	YHON YULISON	MAÑUICO	CUSI	004	01
70110B	THALIA EDITH	QUISPE	ALHUAY	005	01

Fuente: Secretaría del Centro Preuniversitario

Se puede apreciar que la tabla postulante no está relacionada indicador que no cumple las condiciones de integridad que debe cumplir, en esta captura de pantalla tampoco se ve la totalidad de campos de esta tabla que en número son cincuenta campos en las siguientes figuras apreciaremos su contenido que están cortadas en dos capturas de pantalla.

Figura N° 8: Extracto de los registros encontrados en su base de datos.

anyo_cole	Nota	mail_post	fono_post	sexo_post	fech_naci	FECH_INSC	nrod_rec1	impo_rec1	fcom_rec1	nrod_rec2	impo_rec2	fcom_rec2	tota_
2015	12.30	FRAYANTHONY@GMAIL.COM	0000000000	M	25/08/1997	08/03/2016	NULL	NULL	NULL	NULL	NULL	NULL	140.6
2009	12.86	ARC-24@OUTLOOK.ES	995396746	M	24/06/1994	09/03/2016	NULL	NULL	NULL	NULL	NULL	NULL	180.6
2015	12.16	MAURO@HOTMAIL.COM	999999999	M	15/10/1999	09/03/2016	NULL	NULL	NULL	NULL	NULL	NULL	140.6
2015	13.16	AARON@HOTMAIL.COM	953520388	M	06/08/1999	14/03/2016	NULL	NULL	NULL	NULL	NULL	NULL	180.6
2015	14.31	C@GMAIL.COM	983950144	F	07/12/1999	16/03/2016	00002255	180.60	16/03/2016	NULL	NULL	NULL	180.6
2014	13.95	BRAYANSAMUEL@GMAIL.COM	973225029	M	28/05/1998	16/03/2016	00002276	180.60	21/03/2016	NULL	NULL	NULL	180.6
2015	13.69	GINO@HOTMAIL.COM	969058445	M	29/04/1999	17/03/2016	00001969	180.60	17/03/2016	NULL	NULL	NULL	180.6
2015	13.47	NELSON@HOTMAIL.COM	951144149	M	02/09/1999	18/03/2016	00002349	180.60	18/03/2016	NULL	NULL	NULL	180.6
2013	12.96	EDITH@GMAIL.COM	947694548	F	31/03/1997	18/03/2016	00002353	160.00	18/03/2016	NULL	NULL	NULL	180.6
2016	13.42	JOSUE@GMAIL.COM	953451554	M	01/02/1998	18/03/2016	00002305	180.60	18/03/2016	NULL	NULL	NULL	180.6
2015	12.81	JHONATAN@GMAIL.COM	980284579	M	02/11/1998	21/03/2016	2453	180.60	21/03/2016	NULL	NULL	NULL	180.6
2014	12.51	PPIMILAGROS@HOTMAIL.COM	983620744	M	15/03/1999	22/03/2016	2479	180.60	22/03/2016	NULL	NULL	NULL	180.6

Fuente: Elaboración propia.

Se aprecia la cantidad de valores nulos que posee la tabla, ahora en lo que resta de la tabla se puede apreciar la redundancia de datos existente en esta tabla con la dirección de los postulantes no guarda la debida relación sino por el contrario repiten los datos en los diferentes registros:

Figura N° 9: Relación de registros con redundancia de datos.

nomb_papa	nomb_mama	dire2_post	depa_proc	depa_naci	inst_supe	nive_inst	tipo_inst	egre_inst	estu_inst	codi_cana	nrod_aula	ingr_post	nota_ingr
JUAN VELASQUE RINCON	NELY RUTH ROJAS JAUREGUI	PSJE MIRADOR S/N	APURIMAC-ANDAHUAYLAS-ANDAHUAYLAS	APURIMAC-ANDAHUAYLAS-ANDAHUAYLAS						A	01		NULL
GUILLERMO CAÑARI GUZMAN	KATIA ACUÑA ALCARRAZ	PSJ TAHUANTINSUYO N° 231	APURIMAC-ANDAHUAYLAS-ANDARAPA	APURIMAC-ANDAHUAYLAS-ANDARAPA						B	01		NULL
RUBEN DARIO VALLEJOS SAENZ	RUTH LERZUNDI GALVAN	JR.JOSÉ C. MAREATEGUI N°317	APURIMAC-ANDAHUAYLAS-ANDAHUAYLAS	APURIMAC-ANDAHUAYLAS-ANDAHUAYLAS						A	01		NULL
ALFREDO ALLCCA HUAMAN OSCCO	REYNALDA HILDA JUNCO HERHUAY	JR ASTUHUARACA N° 309	APURIMAC-ANDAHUAYLAS-SAN JERONIMO	APURIMAC-ANDAHUAYLAS-SAN JERONIMO						A	01		NULL
CARLOS MOLINA HURTADO	ELENA OALOMINO HUAMAN	VISTA ALEGRE S/N	APURIMAC-ANDAHUAYLAS-ANDAHUAYLAS	APURIMAC-ANDAHUAYLAS-ANDAHUAYLAS						A	02		NULL
JUAN GARFIAS CCAHUANA	CONSTANTINA TAPASCO VELASQUE	HUERTOS DE MANCHAY ZONA F MZ. F-6 LT 09	LIMA-LIMA-PACHACAMAC	LIMA-LIMA-PACHACAMAC						B	02		NULL
CLINARIA QUISPE URBINA	CLINARIA QUISPE URBINA	PSJ. LOS CELAJES S/N	APURIMAC-ANDAHUAYLAS-ANDAHUAYLAS	APURIMAC-ANDAHUAYLAS-ANDAHUAYLAS						A	03		NULL
ABDON ZENON PALOMINO HURTADO	VICYORIA HUARCAYA ROJAS	JR. LAS PALMERAS S/N	APURIMAC-ANDAHUAYLAS-ANDAHUAYLAS	APURIMAC-ANDAHUAYLAS-ANDAHUAYLAS						A	03		NULL
WHASINTON PARIONA QUINTANA	MARIA CCAHUANA CCENTE	AV. CELAJES 184	APURIMAC-ANDAHUAYLAS-ANDAHUAYLAS	APURIMAC-ANDAHUAYLAS-ANDAHUAYLAS						A	03		NULL
JUAN CCORISONCCO GARFIAS	MARGARITA HUAMAN BAUTISTA	COM. ROSASPATA	APURIMAC-ANDAHUAYLAS-PACUCHA	APURIMAC-ANDAHUAYLAS-PACUCHA						A	01		NULL
MANUEL OSCAR YUTO QUISPE	ANGELICA UMERES PAUCAR	JR. TUPAC AMARU S/N	APURIMAC-ANDAHUAYLAS-ANDAHUAYLAS	APURIMAC-ANDAHUAYLAS-ANDARAPA						A	04		NULL
JOSE MARTIN SOTO ANO	YANETH ALARCON GALVAN	AV. CONFRATERNIDAD S/N	APURIMAC-ANDAHUAYLAS-ANDAHUAYLAS	APURIMAC-ANDAHUAYLAS-ANDAHUAYLAS						A	04		NULL

Fuente: Elaboración propia.

Es preciso señalar por ejemplo que el nombre de los padres de los postulantes es innecesario almacenarlos esta base de datos muy probablemente fue adaptada de otra universidad esto genera los abundantes despropósitos de esta base de datos, nótese la gran cantidad de datos repetidos en depa_proc y depa_nac.

Con estos antecedentes debo informar que para realizar las pruebas posteriores de calificación de exámenes de admisión en caso de empate se desempata con las notas de colegio. Por lo que se requiere copiar la información del campo nota de la tabla postulante de la base de datos primigenia a la base de datos construida para este informe de nombre admisión a la tabla postulante y campo nota_colegio. El problemita es que la copia no es inmediata. Si no que se realizó ciertas tareas:

1.- Se exporto la tabla postulante a un archivo de formato Excel.

2.- En la tabla de Excel se eliminó todas las columnas innecesarias y nos quedamos con las columnas: id_postulante y nota, esta tabla se cargó la data dentro de una tabla auxiliar con el código:

```
LOAD DATA INFILE "D:/NOTITA.TXT" INTO TABLE NOTAS;
```

Notita.txt es el archivo de texto generado por el Excel y notas es el nombre de la tabla auxiliar.

3.- Para trasladar datos de un campo a otro campo vacío pero con ciertos registros es necesario aplicar las siguientes sentencias:

```
UPDATE postulante
INNER JOIN notas
ON notas.Id = postulante.Id_postulante
SET postulante.nota_colegio = notas.nota;
```

De lo anterior:

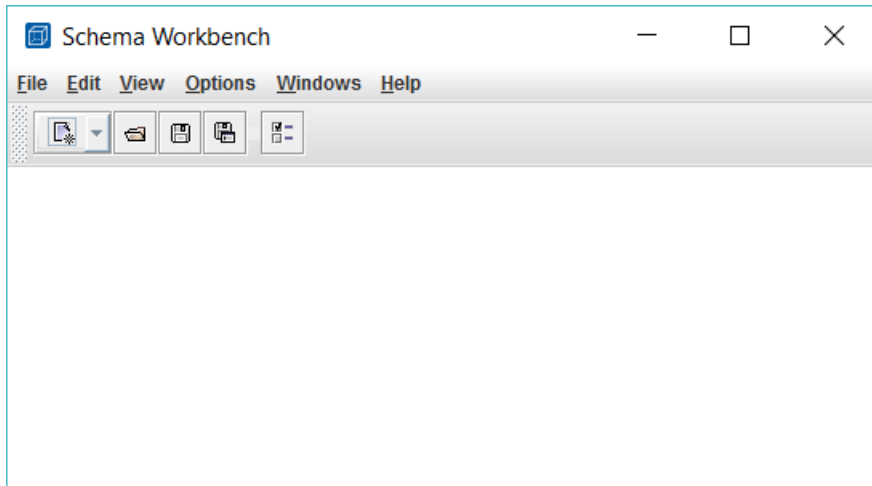
- postulante = tabla destino.
- notas = tabla origen
- nota_colegio nota= las columnas que se desea copiar

Elaborando cubos OLAP en Schema Workbench

Primer cubo: discriminara la cantidad de postulantes por colegio además a qué distrito pertenece cada colegio de manera que la dimensión sea la tabla distrito y la tabla colegio y la dimensión sea la tabla postulante.

1.- Ingresar al software SCHEMA WORKBENCH Y CONECTAR LA BASE DE DATOS DE ADMISIÓN.

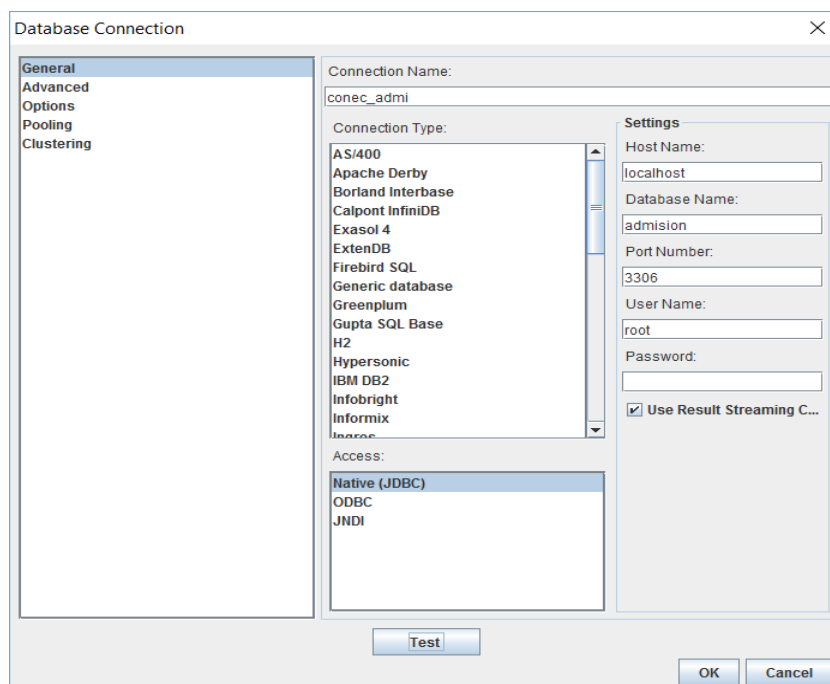
Figura N° 10: Entorno del Software Schema Workbench.



Fuente: Elaboración propia.

Clic en el menú OPTIONS/CONNECTION para que pueda aparecer el siguiente formulario:

Figura N° 11: Ventana de conexión con la base de datos.



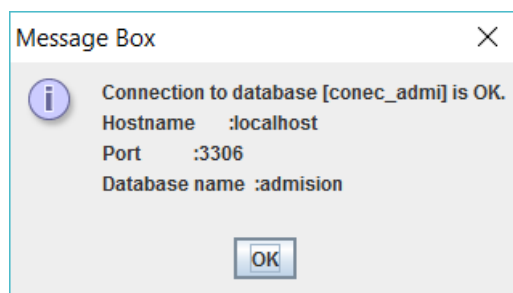
Fuente: Elaboración propia.

Este formulario solicita un nombre de conexión que le asigne “admisioncita”, Host Name para nuestro caso es “localhost” si tuvieran la base de datos en otro

servidor también podría requerir la dirección URL, el nombre de la base datos “admisión” Port Number que por defecto es: 3306 de la misma manera el User Name que es root y como no asigne password a la base de datos se queda en blanco.

Al hacer un clic en test debe mostrar:

Figura N° 12: Mensaje de conexión exitosa.

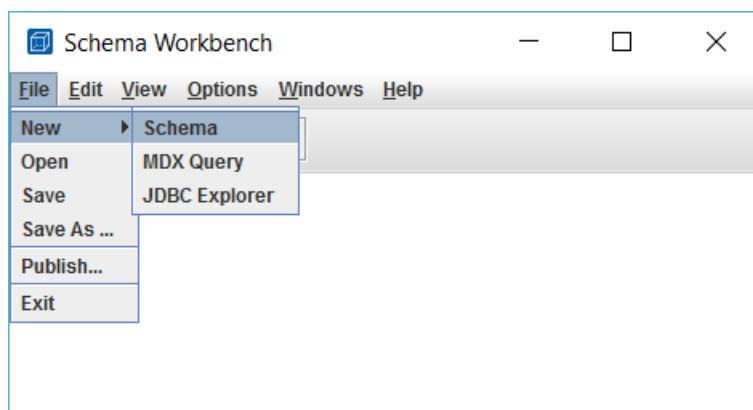


Fuente: Elaboración propia.

De lo contrario “mysql-connector-java-5.1.17.jar” de la página oficial de java y guardarlo en la carpeta drivers de schema workbench.

2.- Diseñar cubo: Del menú File/New/Schema

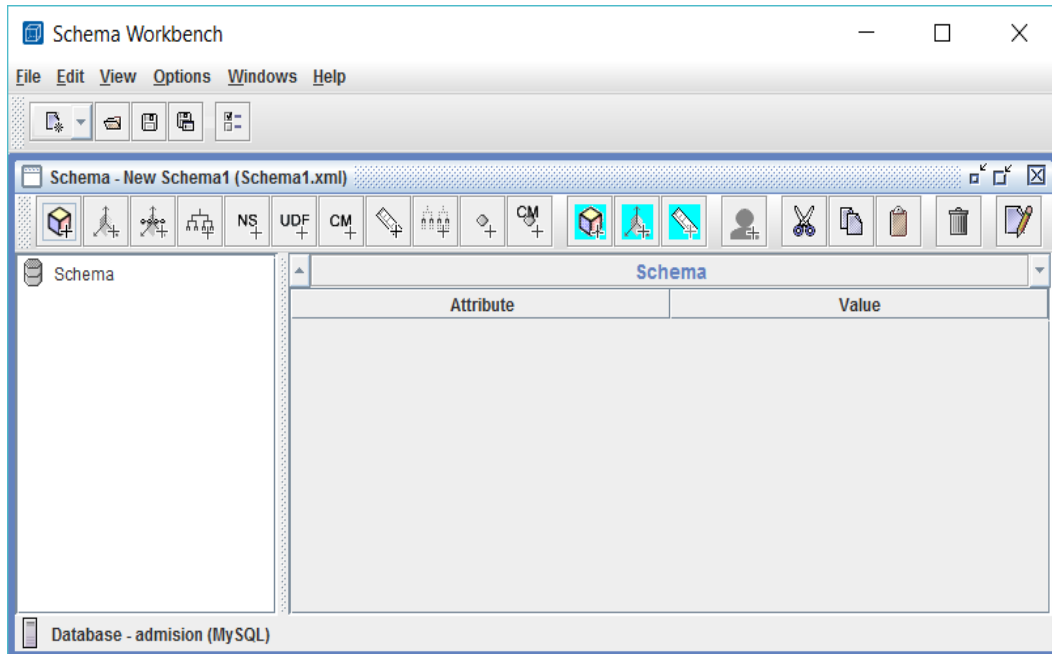
Figura N° 13: Creación de nuevo esquema.



Fuente: Elaboración propia.

Debe mostrar la siguiente ventana:

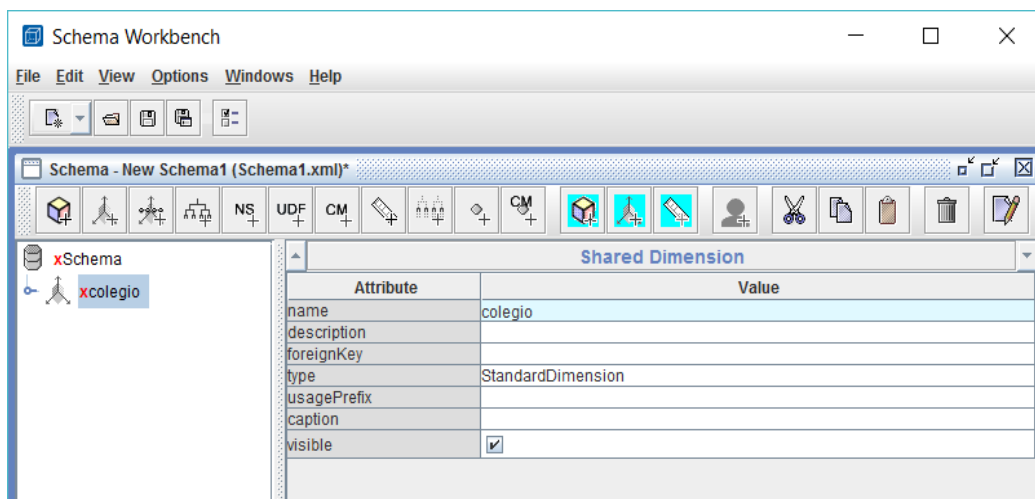
Figura N° 14: Creación de un nuevos esquema.



Fuente: Elaboración propia.

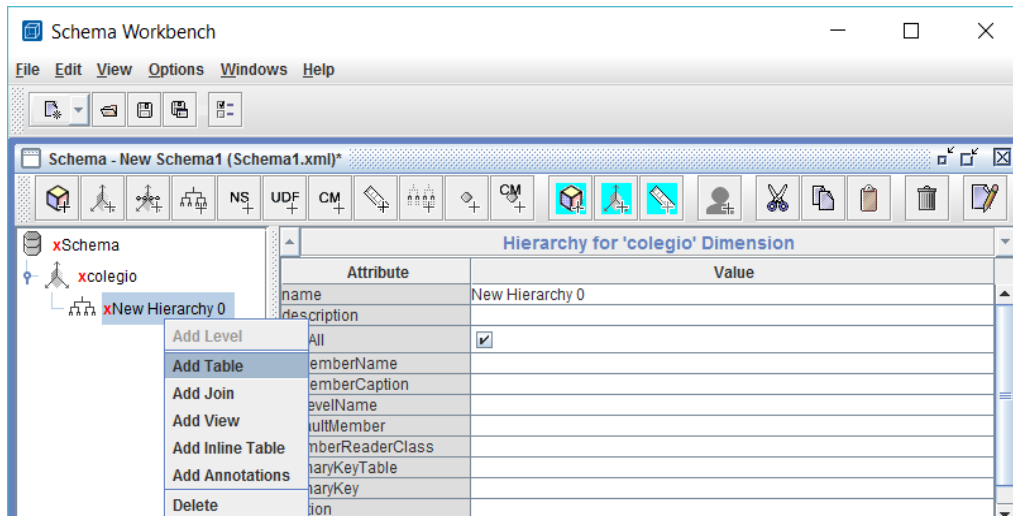
En principio agregamos dimensiones; clic derecho sobre schema y agregamos dimensiones recuerden que las dimensiones son la tabla colegio y la tabla distrito. La primera dimensión la llamaremos colegio que es el mismo nombre de la tabla también tenemos que agregar una tabla, al crear la dimensión automáticamente se crea una jerarquía en esta jerarquía agregamos la tabla “colegio”

Figura N° 15: Ingresando nombre de dimensión.



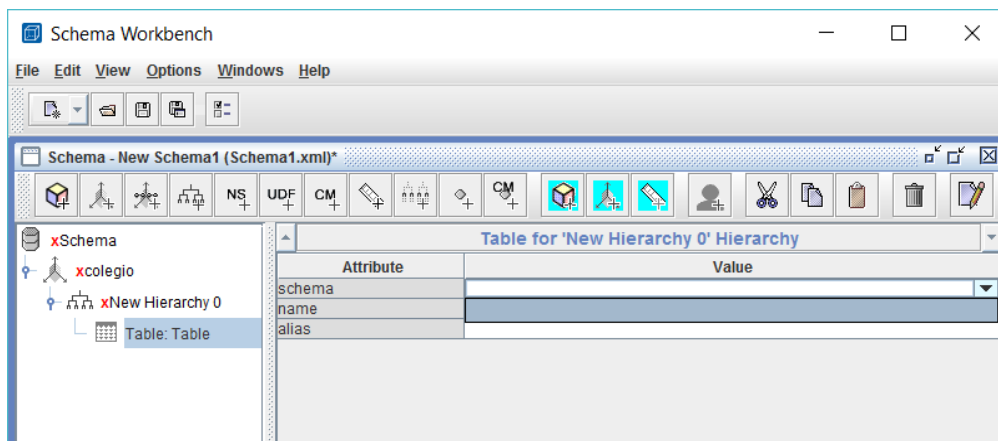
Fuente: Elaboración propia.

Figura N° 16: Agregar tabla a la dimensión.



Fuente: Elaboración propia.

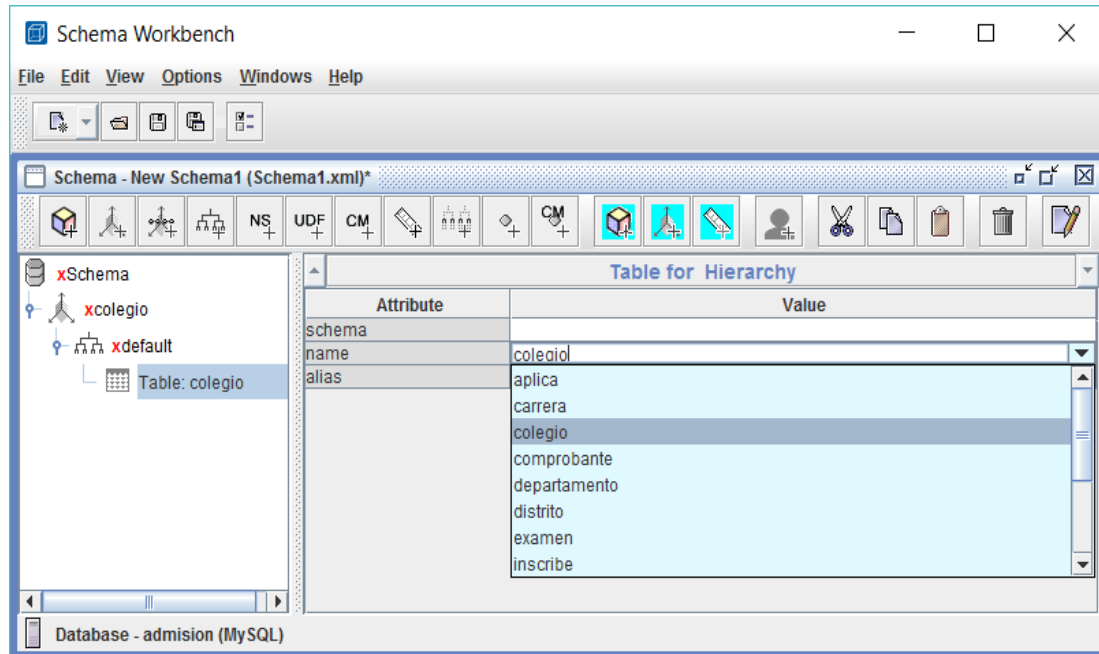
Figura N° 17: Características de la tabla.



Fuente: Elaboración propia.

Esta dimensión de nombre colegio estará enlazada a la tabla del mismo nombre, así que seleccionar la tabla, como está bien conectada además de la tabla colegio se mostrara todas las tablas de la base de datos:

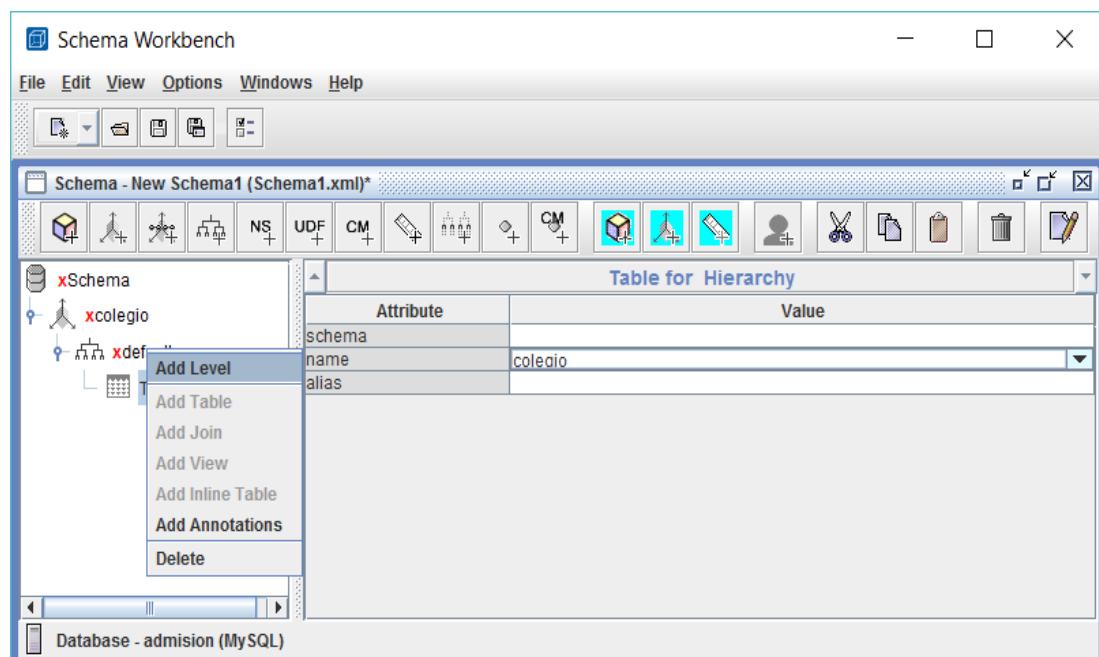
Figura N° 18: Enlaza tabla a dimensión.



Fuente: Elaboración propia.

De la jerarquía que aparece con el nombre default porque se borró el nombre newherarquy agregamos con un clic derecho un nivel. Con la opción add level:

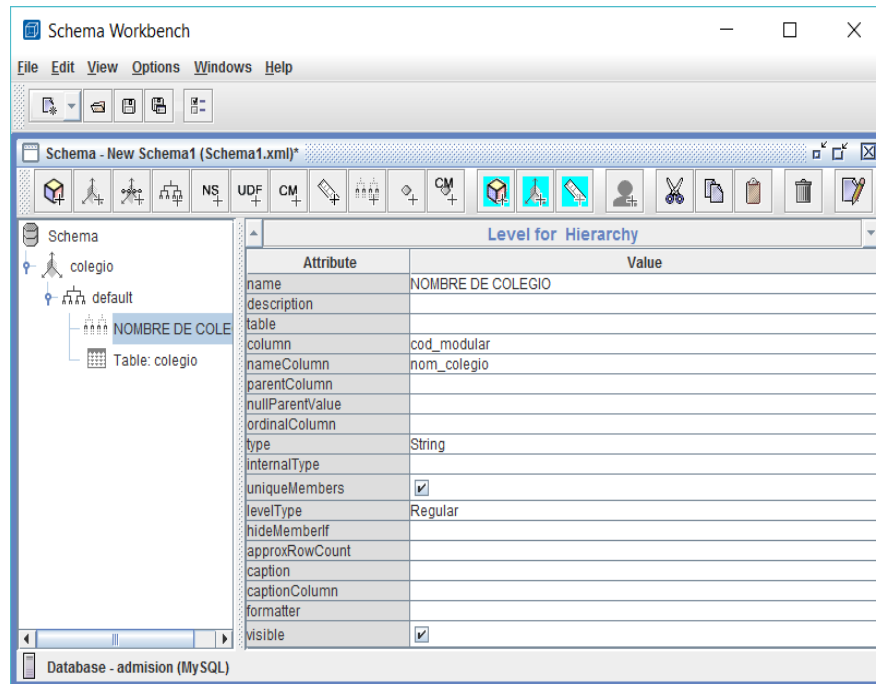
Figura N° 19: Agregando nivel.



Fuente: Elaboración propia.

Le ingresamos en el nivel “Nombre de colegio” ya que será una etiqueta que servirá en la navegación del cubo: Además del nombre del campo que sirve como relación entre la tabla colegio y la tabla postulante que es justamente la clave primaria de nombre “cod_modular” y el nombre de la columna que será el que se mostrara en nuestro cubo: “nom_colegio” y los parámetros que se muestran:

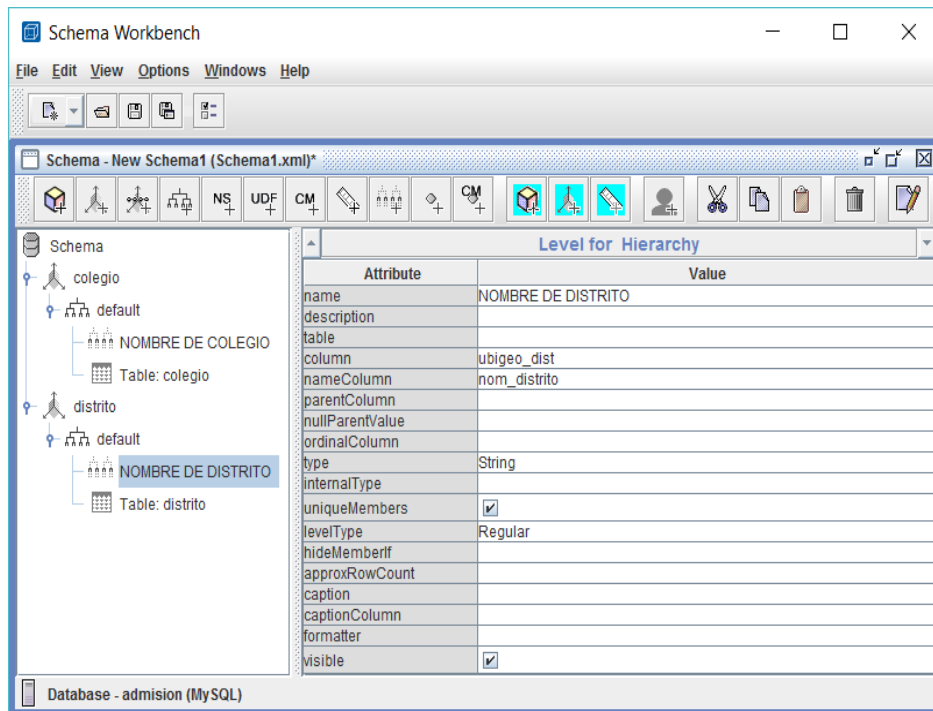
Figura N° 20: Definiendo valores de nivel.



Fuente: Elaboración propia.

De modo similar creamos la otra dimensión que es el distrito análogamente a la primera dimensión creada con lo que tenemos lo siguiente:

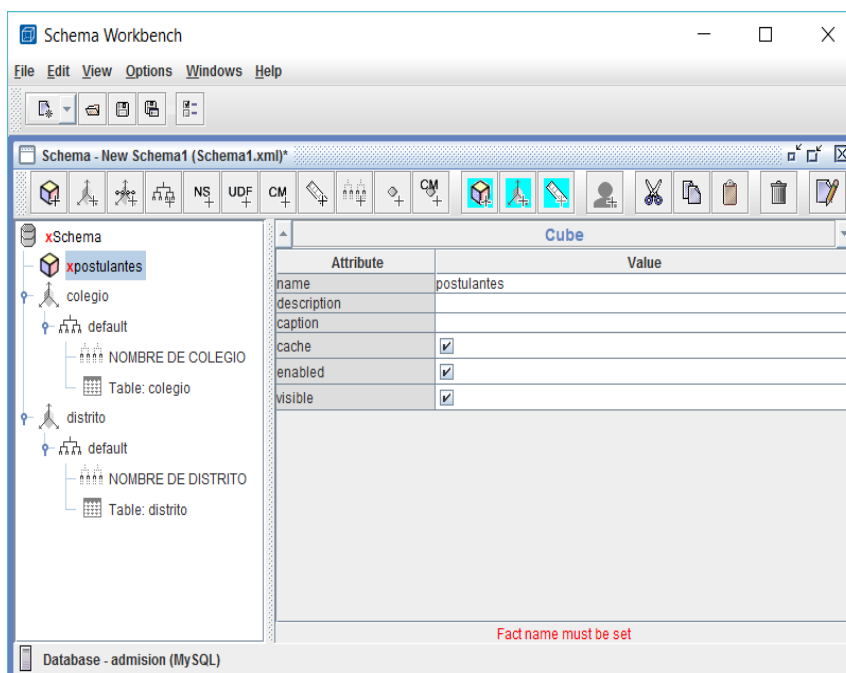
Figura N° 21: Definiendo valores del nivel distrito.



Fuente: Elaboración propia.

Por último le damos el nombre al cubo con un clic derecho sobre Schema: le dimos el nombre de postulantes.

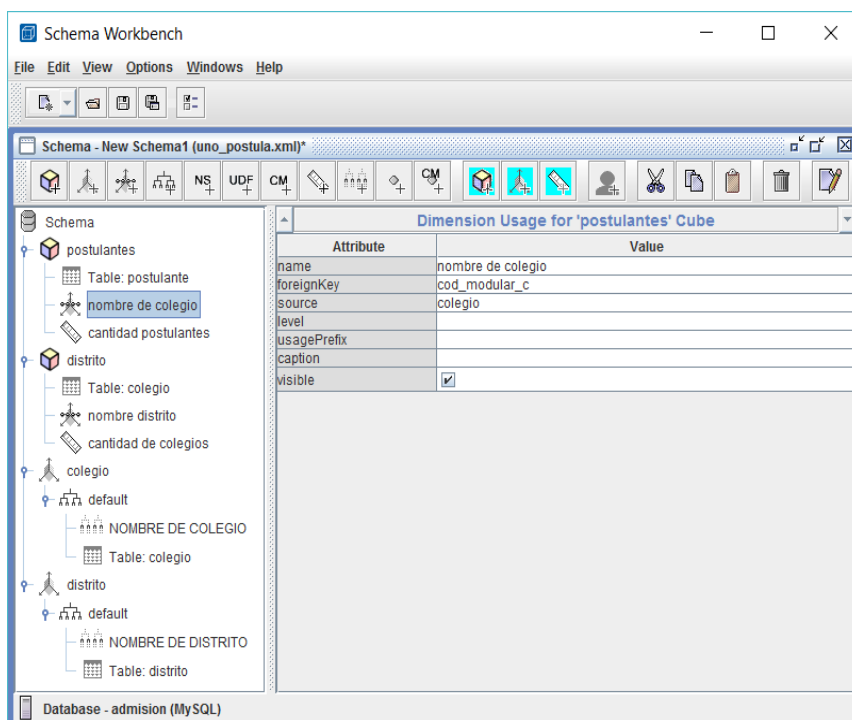
Figura N° 22: Creando cubos.



Fuente: Elaboración propia.

Luego adicionamos dimensiones de uso con el clic derecho sobre el cubo postulantes y seleccionar add dimensión usage dándole el nombre de “nombre de colegio” también se le agrega una dimensión que servirá para contar la cantidad de postulantes por cada colegio y la tabla de echo es postulante, quedara del siguiente modo e incrementaremos un cubo adicional para determinar la cantidad de colegios por distrito.

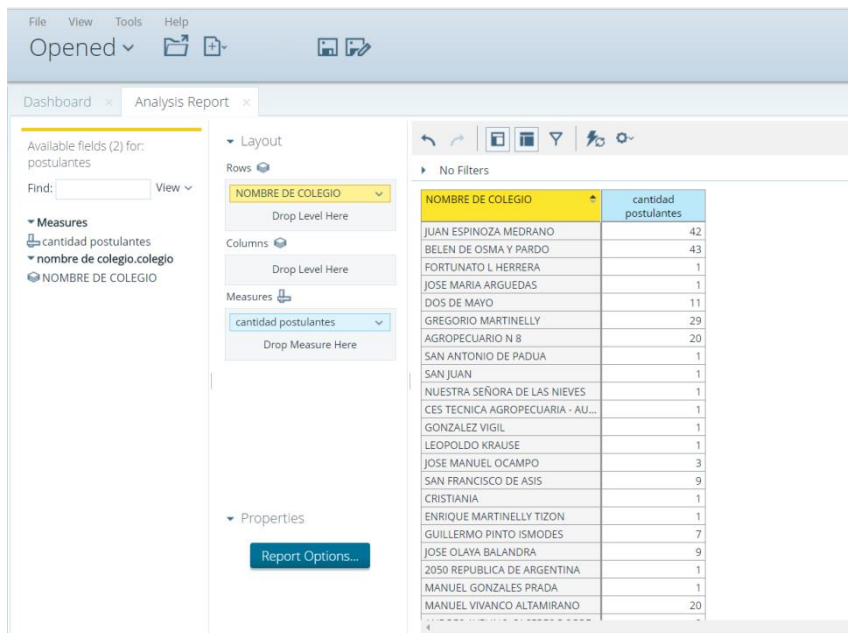
Figura N° 23: Atributos de la dimensión de uso.



Fuente: Elaboración propia.

Una vez terminado este diseño se precisa publicarlo en el software PENTAHO 7.0 en la opción de ANALYSIS REPORT se puede visualizar los siguientes reportes: Es preciso recordar que el PENTAHO lo que hace es publicar los diseños realizados en el software SCHEMA WORKBENCH, prácticamente estos son parte de los resultados propuestos

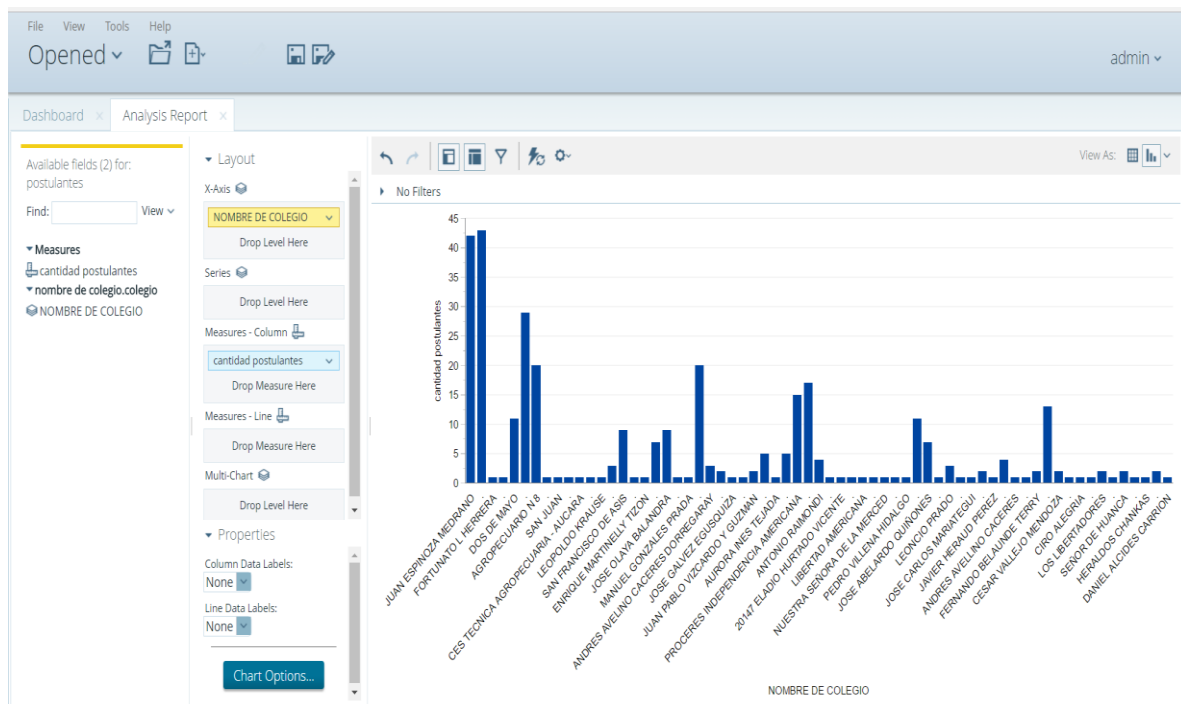
Figura N° 24: reporte de cubo creado con Pentaho.



Fuente: Elaboración propia.

Otro reporte gráfico se da a continuación:

Figura N° 25: Reporte de cubo OLAP en Pentaho V7.0

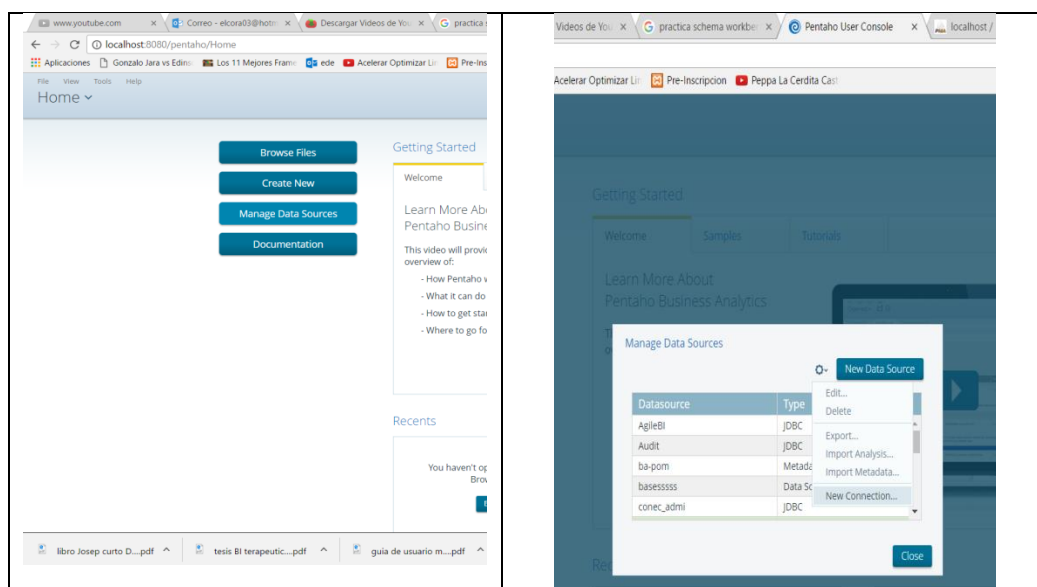


Fuente: Elaboración propia.

2.- El cubo anterior involucra una dimensión y una “measure” medición en el reporte siguiente se obtendrá un cubo con más dimensiones y mediciones en la tabla de hechos, situación que se abordará en las siguientes pantallas.

2.1 En principio es necesario crear una nueva conexión en Pentaho esta nueva conexión luego será usado para importar el cubo diseñado en SCHEMA WORKBENCH. En el menú contextual de la izquierda de la consola de usuario de Pentaho tenemos la opción MANAGE DATA SOURCES al hacer clic obtenemos la ventana de la derecha en esta hacemos clic en la parte superior existe un icono en forma de rueda dentada y seleccionamos nueva conexión (new connection). Es posible configurar la nueva conexión que para el caso particular es una conexión con la base de datos MYSQL

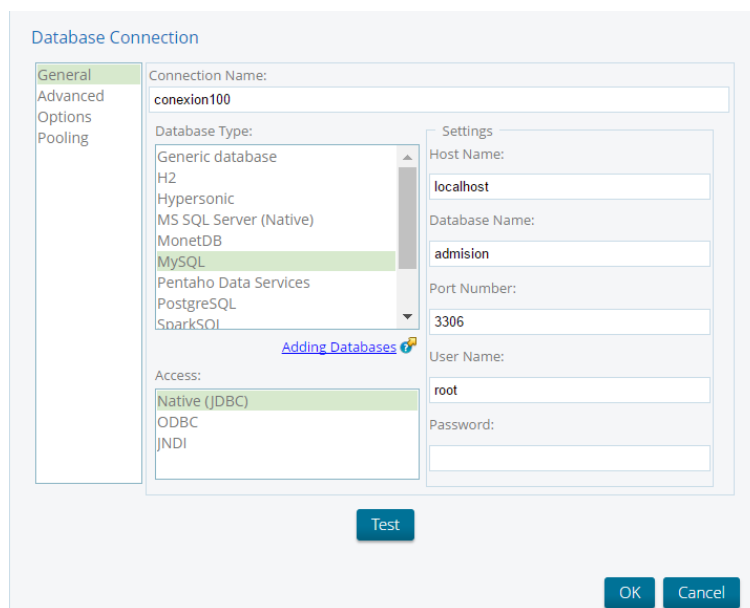
Figura N° 26: Administrando Data Sources.



Fuente: Elaboración propia.

Al hacer clic en New Connection se muestra la siguiente ventana donde podemos configurar la siguiente ventana con todos los datos que se requiere para una conexión con nuestra base de datos de MYSQL:

Figura N° 27: Creando una nueva conexión.

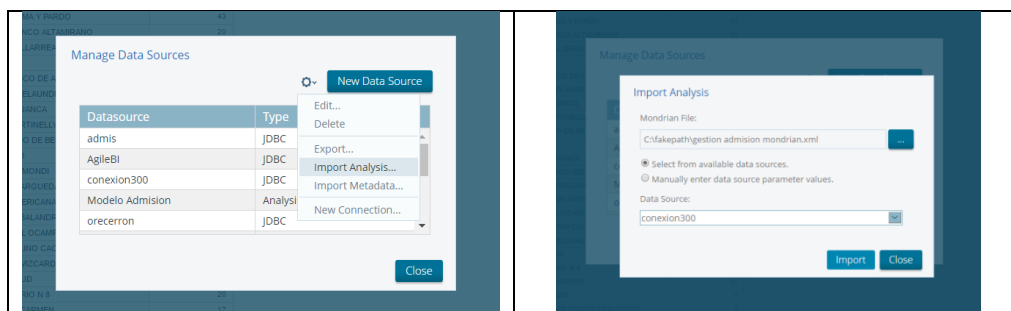


Fuente: Elaboración propia.

Para el trabajo desarrollado se eligió servidor local será; Localhost en el espacio del Host Name en caso de que la base de datos se encuentre alojada en un hosting se tendría que cambiar con la URL correspondiente. El nombre de la base de datos es admisión, el puerto por defecto utilizado por el MYSQL es el 3306, el user name es root y como no le puse password a la base de datos ese casillero queda en blanco.

Creada esta conexión se importará el modelo realizado en el schema workbench considerando la ruta donde se guardó el modelo en formato XML, y luego seleccionar el nombre de la conexión, nosotros estamos trabajando con conexión300 así que importaremos este esquema a Pentaho.

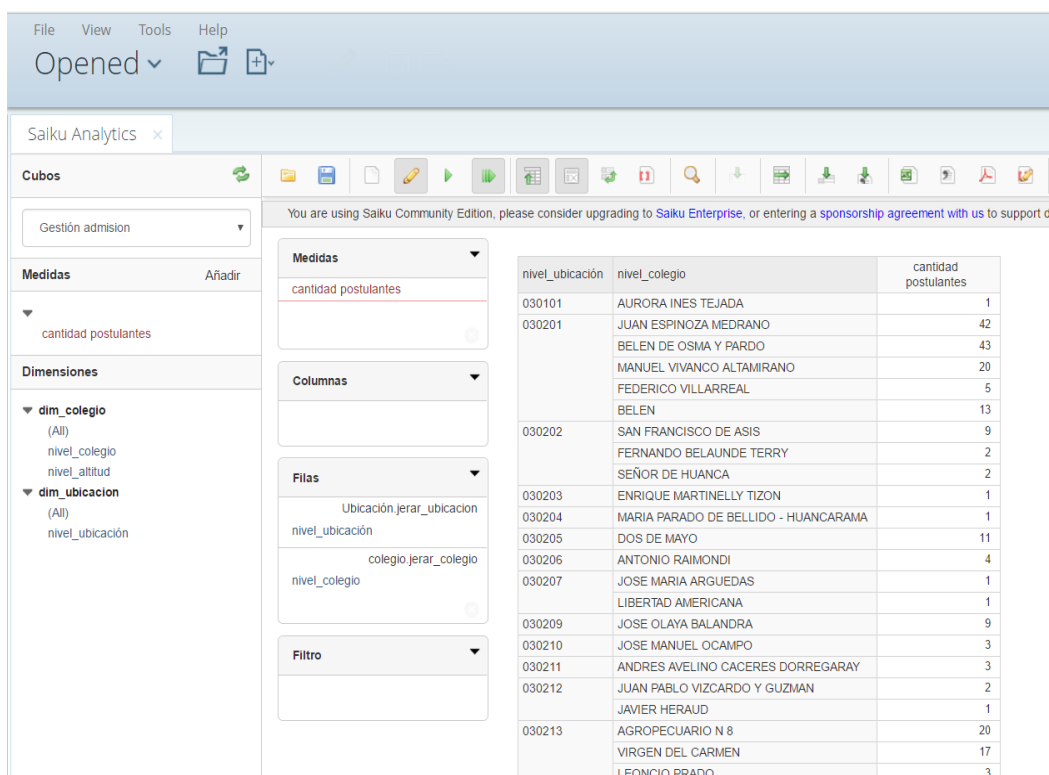
Figura N° 28: Importando análisis.



Fuente: Elaboración propia.

Ahora podemos visualizar el nuevo Schema con el visor de SAIKU:

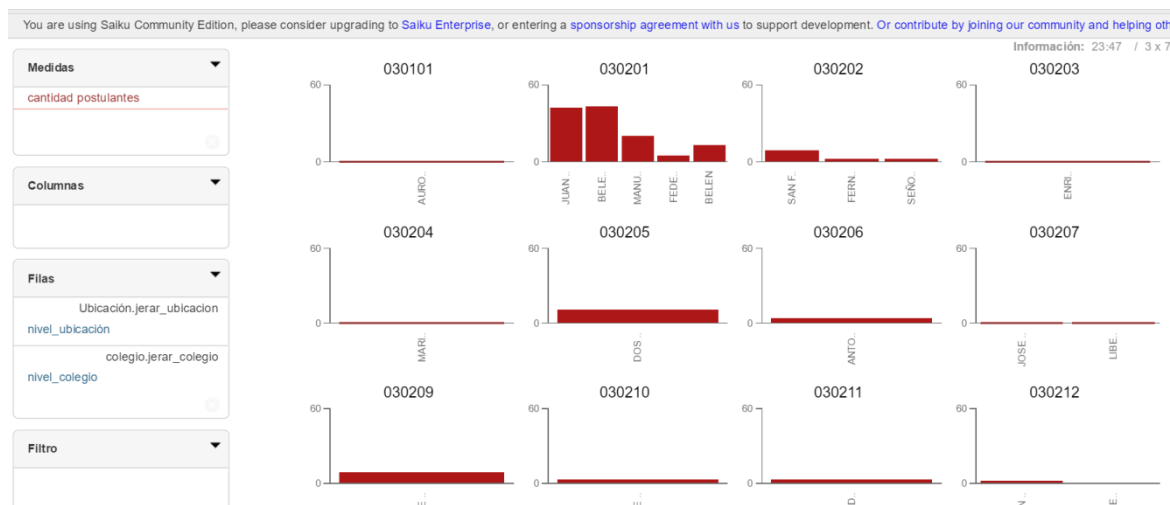
Figura N° 29: Mostrando dos dimensiones del cubo.



Fuente: Elaboración propia.

La Figura N°29, muestra la cantidad de estudiantes por colegio y por ubicación lo cual constituye un cubo de presentación de información, esta pestaña incide en el nivel_ubicación y al nivel colegio por la simple razón de que con este reporte se puede apreciar cuantos postulantes provee cada colegio pero además el nivel_ubicación nos reporta el código de ubigeo de distrito así el cubo por simple inspección muestra el código 030201 que pertenece al distrito de Andahuaylas se consigue inscribir a 123 postulantes y en el sector inferior se aprecia que entre los tres colegios del distrito de San Jerónimo se consigue inscribir a 40 postulantes. Lo ideal sería que en vez de códigos de ubigeo nos proporcione directamente el nombre del distrito en lugar del código de ubigeo. Lo interesante del Pentaho es que además de darte los reportes en forma de tabla también te entrega reportes en forma gráfica como los doce gráficos en dos dimensiones que nos indican según el código de ubigeo la cantidad de postulantes por colegio, existen distritos que no aportan significativamente con postulantes a nuestra universidad.

Figura N° 30: Gráfico de barras del cubo OLAP.



Fuente: Elaboración propia.

3.- Diseño del DATA WAREHOUSE

La teoría indica que el almacén de datos o Data Warehouse se obtiene de una o varias bases de datos incluso cuando solo se tiene una base de datos de donde obtendremos nuestros reportes para la inteligencia de negocios es necesario realizar el proceso de Extracción, Transformación y carga de la data (proceso ETL) ya que el modelo de inteligencia de negocios precisa de un modelo multidimensional (modelo estrella) y no un modelo entidad relación como normalmente se tiene modelado en las bases de datos convencionales. Esta operación se consigue con el Software Pentaho Data Integration que lo descargamos de la página: <http://community.pentaho.com/> el archivo tiene un tamaño de 882 Mb su nombre es pdi-ce-7.1.0.0-12.zip y es la última versión disponible a la fecha, este archivo se descomprime y se ejecuta con el archivo de extensión .bat de nombre Spon.bat para iniciar con la siguiente pantalla y realizar nuestra operación de proceso ETL. Este aplicativo hace posible extraer datos de cualquier fuente de datos transformar estos datos eliminando columnas o registros o adicionando nuevas columnas a una tabla especial denominada por el marco teórico como tabla de hechos y las demás tablas que estarán a su alrededor serán conocidas como las tablas dimensionales y como varias tablas estarán relacionadas con la tabla de hechos adoptan el nombre de modelo multidimensional, estos modelos están conformados en principio por dimensiones, luego por jerarquías y niveles además de métricas que se extraen

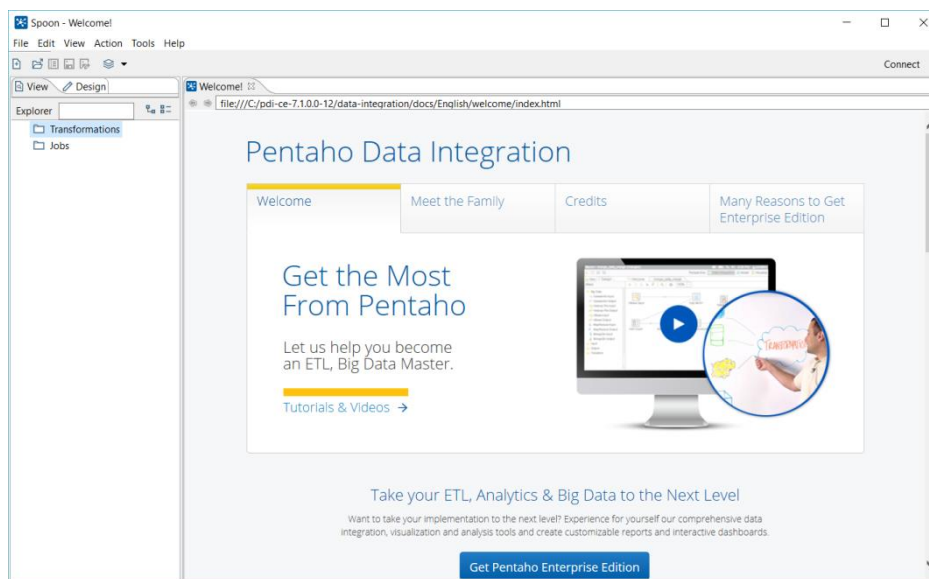
de la tabla de hechos tanto las dimensiones como la tabla de hechos están relacionadas con sus respectivas tablas.

Figura N° 31: Pantalla de inicio de Pentaho Data Integration.



Fuente: Elaboración propia.

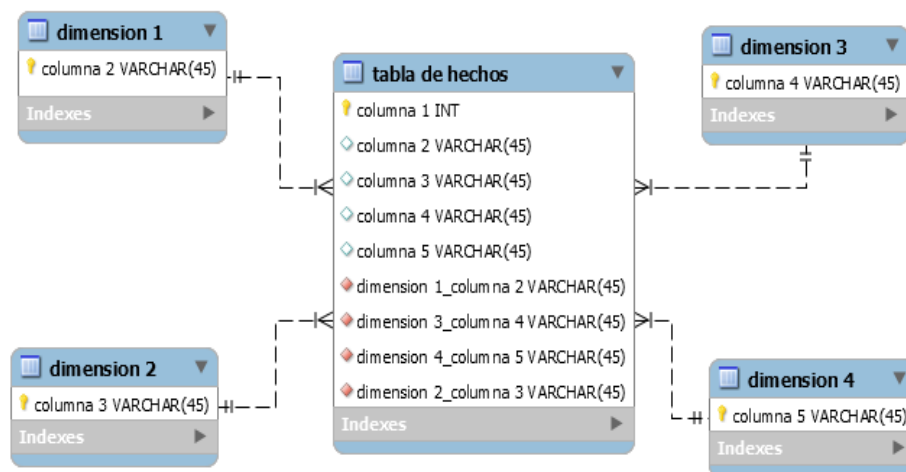
Figura N° 32: Entorno de inicio de Pentaho Data Integration



Fuente: Elaboración propia.

El modelo multidimensional tendrá la siguiente apariencia:

Figura N° 33: Tabla de hechos en el modelo estrella.

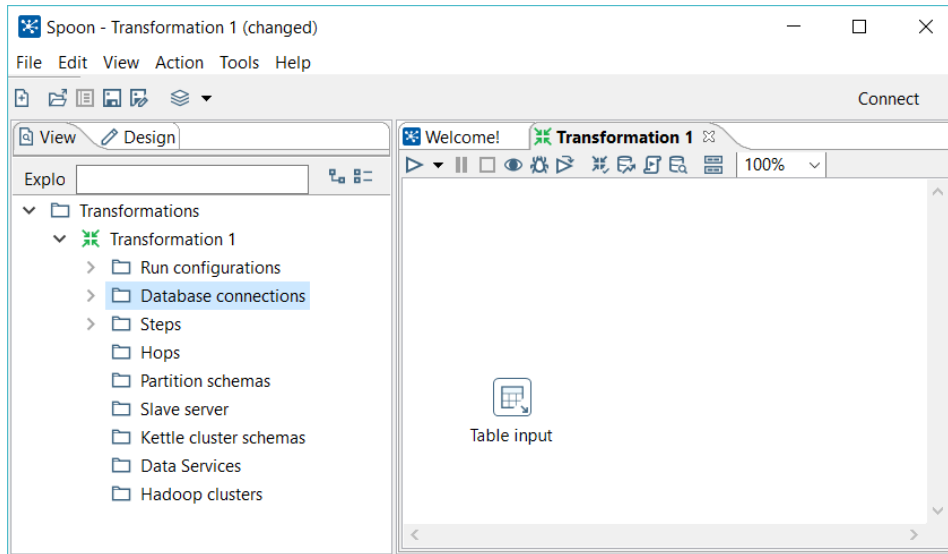


Fuente: Elaboración propia.

La figura N°31 representa un modelo estrella de base de datos, para ser utilizado en cubos OLAP es indispensable que la base de datos tenga este modelo para nuestro modelo la tabla de hechos será la tabla estudiante que almacena los principales datos del Centro Preuniversitario la dimensión 1 es la tabla colegio la dimensión 2 es la tabla de ingresos y la dimensión 3 es la tabla de egresos. La importancia de mostrar esta figura radica en que los modelos entidad relación convencionales no necesariamente tienen esta forma mientras que en esta modelo se aprecia claramente que la tabla de hechos es la tabla más importante y las dimensiones serán los cubos posteriores.

Lo primero que se realizó con el software Pentaho Data Integration es conectar con la base de datos que usaremos como fuente para construir nuestro modelo multidimensional; luego en la pestaña "View" del sector izquierdo al abrir transformaciones se creó la conexión en database connections con un clic derecho sobre este botón eligiendo la opción NEW:

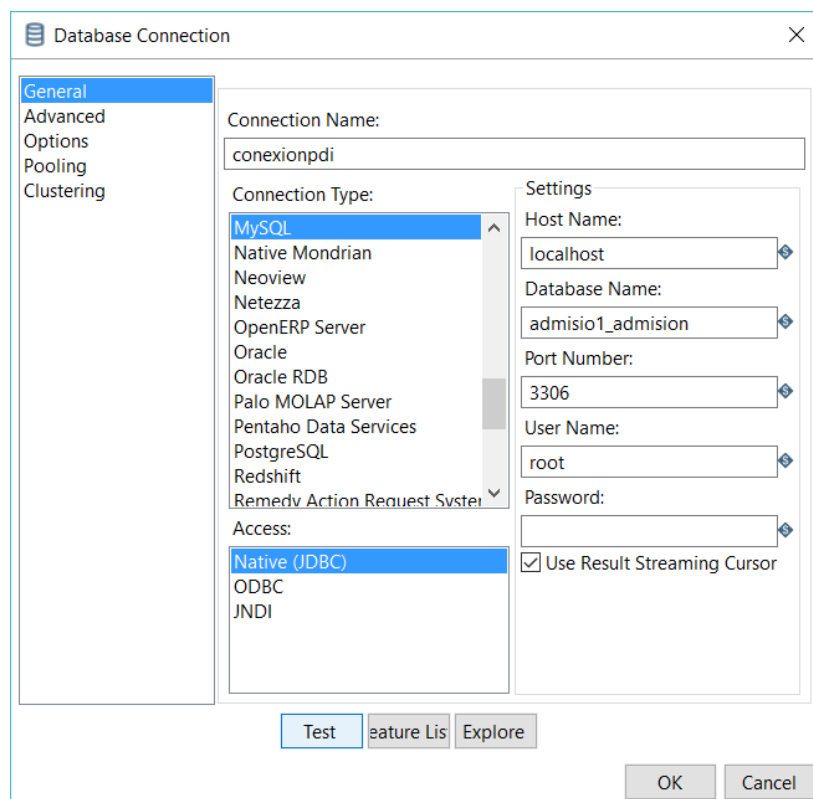
Figura N° 34: Creando nueva conexión.



Fuente: Elaboración propia.

Al hacer doble clic sobre el botón NEW se mostrara la siguiente ventana en donde podremos configurar la conexión con la base de datos MYSQL

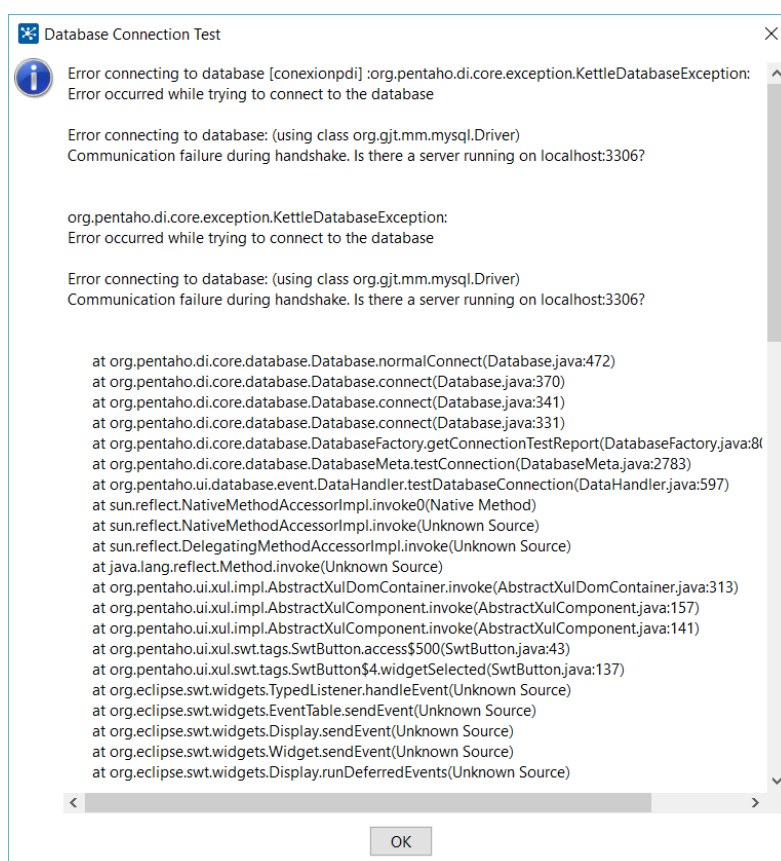
Figura N° 35: Parámetros de base de datos MySql.



Fuente: Elaboración propia.

Se le asigna un nombre y todos los parámetros de los ajustes (SETTINGS) ver figura N°35, antes de hacer un clic en OK es preciso realizar un TEST así que debemos realizar un clic sobre el botón TEST, normalmente se visualiza la siguiente ventana mostrando un error fue solucionado ya que en Software libre es la principal dificultad con que debemos lidiar, en este sector el error típico es el siguiente.

Figura N° 36: Mensaje de error en la conexión.



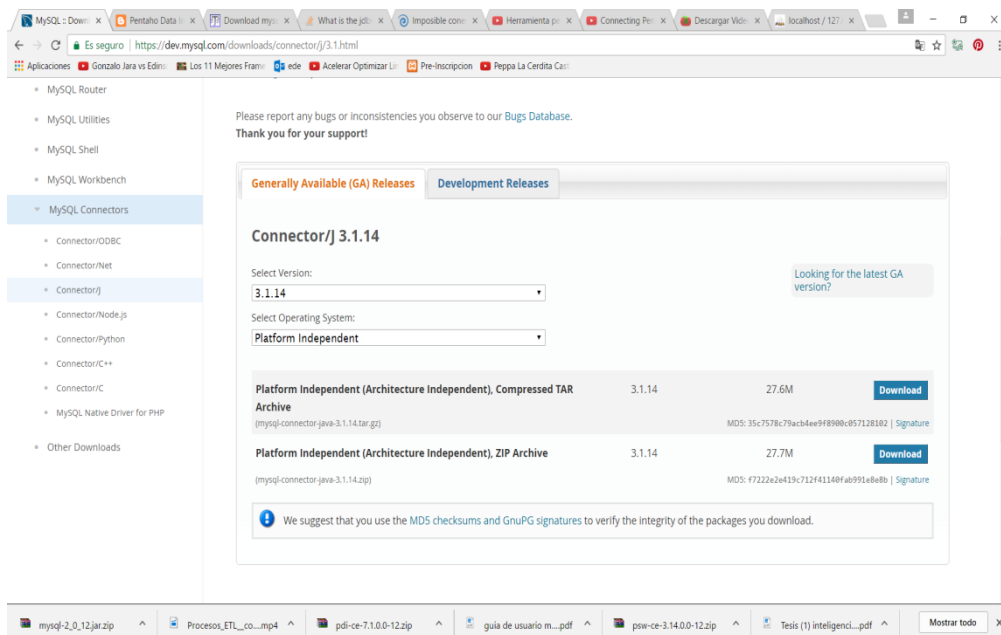
Fuente: Elaboración propia.

Pues bien es el momento de buscar un archivo de extensión .jar (archivo de java) que nos permita realizar esta conexión. Así que lo tenemos que descargar de la siguiente página: <https://dev.mysql.com/downloads/connector/j/3.1.html>; de internet, debajo le mostramos incluso la página.

Casi todos los programas que tienen que ver con Pentaho tienen este inconveniente sucede con el propio Pentaho, con el software Schema Workbench y ahora con Data Integration por lo tanto debemos ubicar el archivo

requerido para cada caso y copiarlo a la ruta especificada para este caso la ruta donde debemos copiarlo es: C:\pdi-ce-7.1.0.0-12\data-integration\lib; también dependerá donde copiaron el archivo del software pentaho data integration

Figura N° 37: Página de descarga de archivos con extensión JAR.

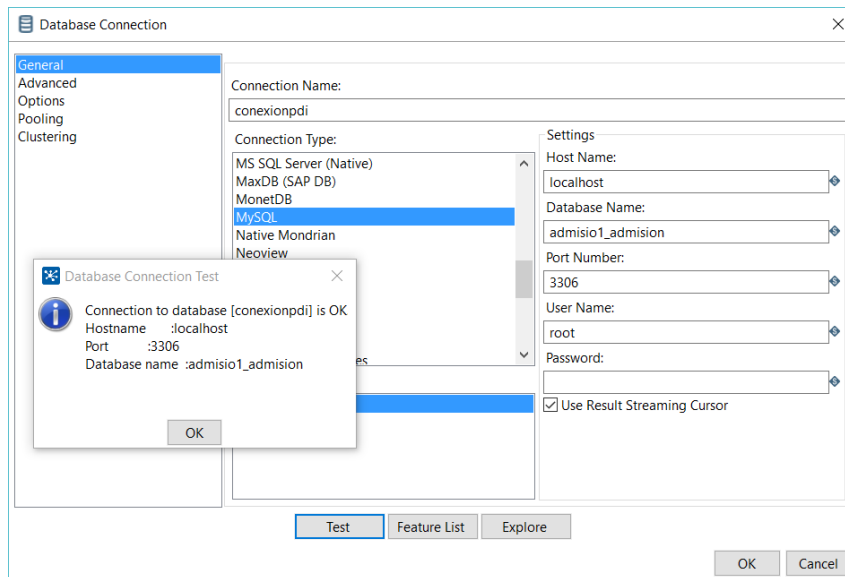


Fuente: Elaboración propia.

El archivo se denomina: mysql-connector-java-3.1.14-bin.jar una vez descomprimido se almacena en la ruta C:\pdi-ce-7.1.0.0-12\data-integration\lib; se cierra el software data integration para volverlo a iniciar y ya podemos realizar nuestra primera operación ETL, ojo esta operación es repetitiva para cada software del pentaho. En caso de requerir información adicional recurrir a la documentación que proporciona la página oficial de Pentaho <http://wiki.pentaho.com/display/EALes/.03+Conexiones+a+Bases+de+Datos> (Bernabeu, 2009).

El siguiente será el mensaje que se obtiene cuando las cosas salen bien:

Figura N° 38: Mensaje de conexión de BD exitoso.

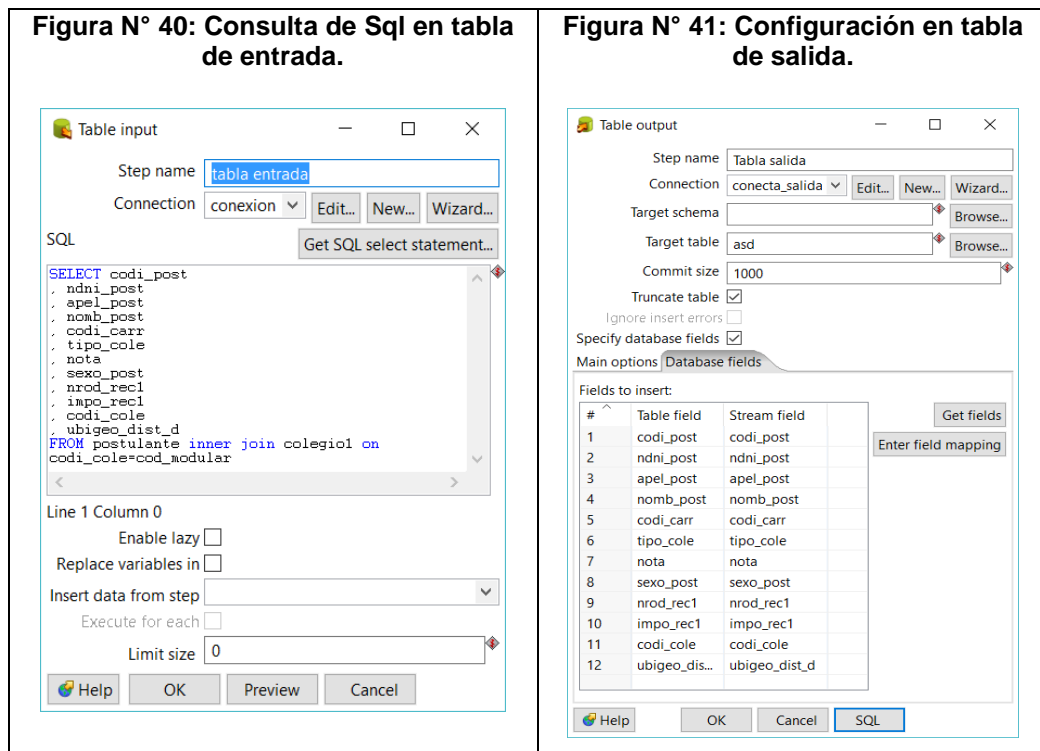


Fuente: Elaboración propia.

Ya con esta conexión es posible realizar la extracción, transformar y cargar en una nueva base de datos la tabla que se le asignara de nombre postulante pero que será la tabla de hechos luego copiaremos las tablas de colegio y distrito para poder armar dos de sus dimensiones, se agregó las tablas de carrera y también de ingresos. Todo ello se hizo en la siguiente secuencia de imágenes:

<p>El gráfico de la derecha posee una transformación de nombre <code>transforma_uno</code>, esta transformación tiene una tabla de entrada asociada a una conexión de nombre "conexión" y una tabla de salida asociada a una conexión de nombre "conecta_salida", entonces la tabla entrada hace referencia a la tabla inicial que tenemos y la tabla salida se asocia a la otra base de datos</p>	<p>Figura N° 39: Esquema de transformación en PDI</p>
--	--

La configuración de la tabla de entrada y de la tabla de salida podemos visualizarlo en las siguientes imágenes:



Y al ejecutar en el botón de corrido de transforma_uno se obtuvo la preciada tabla con todos los registros habiendo creado previamente una base de datos y asignándole el nombre de la tabla para nuestro caso “asd” que se pone la configuración de tabla salida.

Figura N° 42: Reporte de tabla de salida de Pentaho data Integration

codi_post	ndni_post	apel_post	nomb_post	codi_carr	tipo_cole	nota	sexo_post	nrod_rec1	impo_rec1	codi_cole	ubigeo_dist_d
010001	72882457	VALLEJOS LERZUNDI	MAURO RUBEN	03	E	12.16	M	00001584	180.6	0236190	030201
010055	73440376	CCORAHUA CASTRO	DAVID	01	E	12.53	M	00001583	180.6	1205996	030609
010056	75688976	FALCON CHUNGUILLO	NOEMI YENIFER	01	E	13.72	F	00001610	180.6	0362681	050701
010059	77575565	VELASQUE SERNA	AYDEE	01	E	13.80	F	00001582	180.6	1311216	030201
010060	71526418	YANAHUAMAN GALLEGOS	WILDER REYNAN	01	E	13.71	M	00001609	180.6	0535716	030202
010061	72909029	URRUTIA GONZALES	NILFA YULISA	01	E	13.16	F	00001608	180.6	0236570	030216
010062	71855775	GUTIERREZ HUARACA	ROCIO	01	E	12.77	M	00001623	180.6	1331537	030602
010063	70676971	MAUCAYLLE COBRISONCCO	YONY	01	E	13.58	F	00001581	180.6	0671248	030213
010064	73516237	PAHUARA PICHIRUA	MARISELA	01	E	14.29	F	00001607	180.6	0575480	030209
010065	72706163	CORONADO ROJAS	RONALD	01	E	12.62	M	00001606	180.6	0742023	030206
010066	71876344	VILLAGARAY ACEVEDO	FLORMIRA	01	E	13.93	F	00001605	180.6	1331438	030607
010067	71855986	CRUZ QUISPE	YANETH	01	E	12.86	F	00001604	180.6	1437375	150117
010069	70397529	CUSI ALTAMIRANO	MIGUEL ANGEL	01	E	13.27	M	00001622	180.6	1205236	030201
010070	70396345	RODRIGO ALTAMIRANO	JOSE LUIS	01	E	13.33	M	00001621	180.6	0236323	030201
010071	70148219	VELASQUE ROJAS	FRAY ANTHONY	01	E	12.30	M	00001620	180.6	0236190	030201
Consola	70675512	MALLCCO	DANIFI	01	P	14.75	M	00001580	180.6	1330685	030201

Fuente: Elaboración propia.

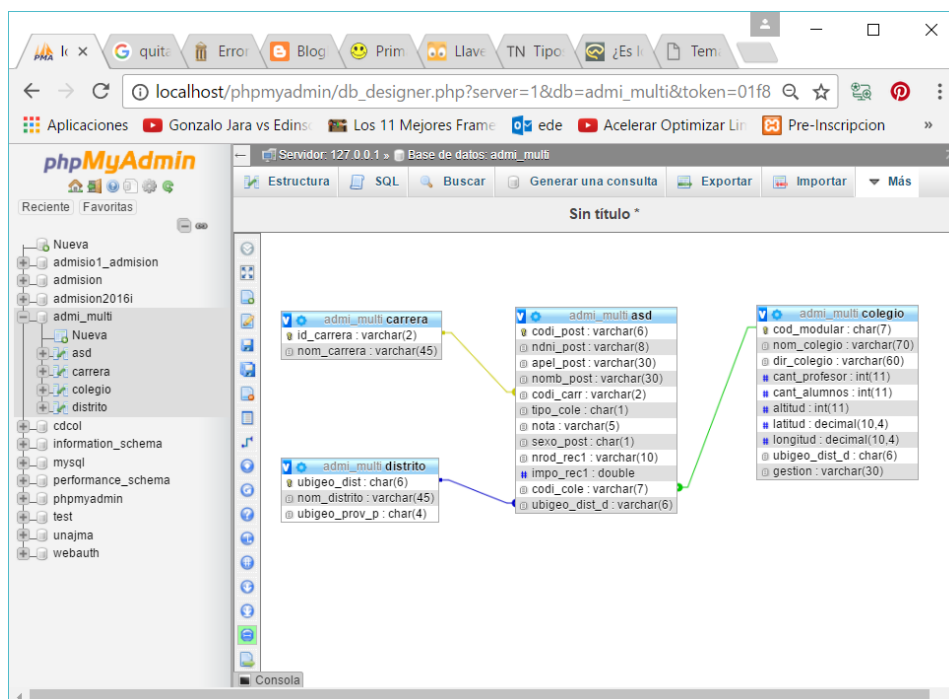
En el gráfico de la figura N°42, visualizamos dentro de la base de datos admulti la tabla asd con los registros insertados esta tabla también fue creada desde el software Data Integration de Pentaho.

CAPÍTULO III

RESULTADOS Y DISCUSIÓN

A continuación el modelo multidimensional con una tabla de hechos y tres dimensiones podría existir una dimensión adicional de nombre tiempo procesos de admisión pero faltaría data para poder insertar esta nueva dimensión: Siempre se está repitiendo que el Data Warehouse es una base de datos pero no es la misma base de datos que usa un sistema de información normal ya que este está diseñado como un modelo entidad relación y el modelo multidimensional está diseñado bajo el esquema de estrella

Figura N° 43: Modelo multidimensional diseñado en Mysql.




Fuente: Elaboración propia.

Este modelo multidimensional que pertenece al modelo estrella permitirá realizar un cubo OLAP con mayor información que el primero que presentamos ya que estamos agregando dos dimensiones más.

Lo primero es visualizar la cantidad de postulantes por colegio además del lugar a donde pertenece en un reporte anterior solo lo podíamos tener a nivel de código de ubigeo (código de ubicación de distrito)

Figura N° 44: Primer reporte de Cubo Olap del Centro Preuniversitario.

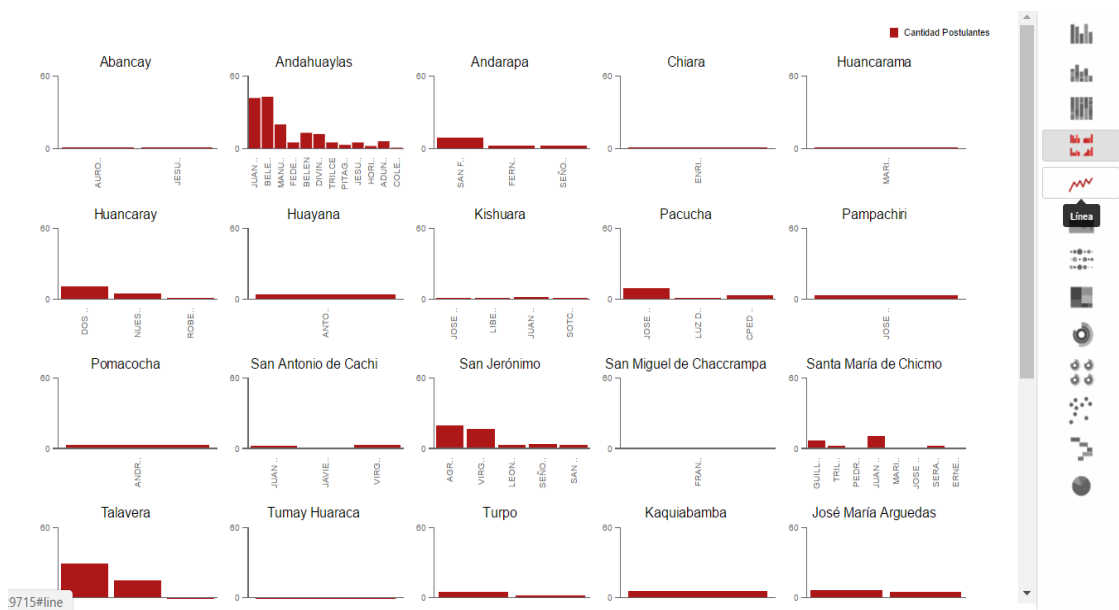


Nombre distrito	Nombre de Colegio	Cantidad Postulantes
Abancay	AURORA INES TEJADA	1
	JESUS DE LAS AMERICAS	1
Andahuaylas	JUAN ESPINOZA MEDRANO	42
	BELEN DE OSMA Y PARDO	43
	MANUEL VIVANCO ALTAMIRANO	20
	FEDERICO VILLARREAL	5
	BELEN	13
	DIVINO MAESTRO	12
	TRILCE	5
	PITAGORAS NIVEL A	3
	JESUS NAZARENO	5
	HORIZONTE DEL SABER	2
	ADUNI MAS SAC	6
	COLEGIO MILITAR ANCCOHUAYLLO	1
	Andarapa	SAN FRANCISCO DE ASIS
FERNANDO BELAUDE TERRY		2
SEÑOR DE HUANCA		2
Chiara	ENRIQUE MARTINELLY TIZON	1
Huancarama	MARIA PARADO DE BELLIDO - HUANCARAMA	1
Huancaray	DOS DE MAYO	11
	NUUESTRA SEÑORA DE ASUNCION	5
	ROBERTO ZEVALLOS CORDOVA	1
Huayana	ANTONIO RAIMONDI	4
Kishuara	JOSE MARIA ARGUEDAS	1
	LIBERTAD AMERICANA	1
	JUAN VELASCO ALVARADO	2
	SOTCCOMAYO	1

Fuente: Elaboración propia.

La riqueza de los cubos OLAP, es su entorno gráfico para presentar la información, La figura N°45 presenta la cantidad de postulantes por cada distrito de la provincia de Andahuaylas, además de la cantidad de postulantes que cada colegio dentro de un distrito proporciona a la universidad, indudablemente que esta información es muy requerida para que la dirección del centro preuniversitario pueda tomar decisiones.

Figura N° 45: Reporte gráfico de Cubos OLAP.



Fuente: Elaboración propia.

Y este es un gráfico más rico: con tres dimensiones y dos métricas cantidad de postulantes e ingresos por cada postulante:

Figura N° 46: reporte de tres dimensiones y dos métricas.

iku Community Edition, please consider upgrading to [Saiku Enterprise](#), or entering a [sponsorship agreement](#) with us to support development. Or [contribute by joining our commu](#)

Nombre distrito	Nombre de Colegio	Nombre de Carrera	Cantidad Postulantes	Ingresos
Abancay	AURORA INES TEJADA	Ingeniería de Sistemas	1	160
	JESUS DE LAS AMERICAS	Ingeniería de Sistemas	1	180,6
Andahuaylas	JUAN ESPINOZA MEDRANO	Administración de Empresas	29	5.157,4
		Ingeniería Agroindustrial	2	340,6
		Ingeniería de Sistemas	11	1.986,6
	BELEN DE OSMAY PARDO	Administración de Empresas	30	5.418
		Ingeniería Agroindustrial	4	722,4
	MANUEL VIVANCO ALTAMIRANO	Ingeniería de Sistemas	9	1.625,4
		Administración de Empresas	13	2.347,8
		Ingeniería Agroindustrial	3	541,8
	FEDERICO VILLARREAL	Ingeniería de Sistemas	4	722,4
		Administración de Empresas	1	180,6
	BELEN	Administración de Empresas	9	1.625,4
		Ingeniería Agroindustrial	2	361,2
		Ingeniería de Sistemas	2	361,2
	DIVINO MAESTRO	Administración de Empresas	8	1.444,8
		Ingeniería Agroindustrial	1	180,6
		Ingeniería de Sistemas	3	541,8
TRILCE	Administración de Empresas	2	461,2	
	Ingeniería de Sistemas	3	691,8	
PITAGORAS NIVEL A	Administración de Empresas	2	461,2	
	Ingeniería de Sistemas	1	230,6	
	Administración de Empresas	3	621,2	
JESUS NAZARENO	Ingeniería Agroindustrial	1	230,6	
	Ingeniería de Sistemas	1	230,6	
	Administración de Empresas	1	150,6	
HORIZONTE DEL SABER	Administración de Empresas	1	150,6	

Fuente: Elaboración propia.

Otro reporte:

Figura N° 47: Reporte sobre la altitud sobre el nivel del mar de los colegios.

Nombre de Carrera	Nombre de Colegio	Altitud	Cantidad Postulantes	Ingresos
Administración de Empresas	JUAN ESPINOZA MEDRANO	2903	29	5.157,4
	BELEN DE OSMA Y PARDO	2903	30	5.418
	DOS DE MAYO	2906	8	1.444,8
	GREGORIO MARTINELLY	2815	20	3.572
	AGROPECUARIO N 8	2954	13	2.307,8
	SAN ANTONIO DE PADUA	115	1	180,6
	SAN JUAN	117	1	180,6
	NUUESTRA SEÑORA DE LAS NIEVES	3178	1	180,6
	CES TECNICA AGROPECUARIA - AUCARA	3220	1	180,6
	GONZALEZ VIGIL	2633	1	180,6
	LEOPOLDO KRAUSE	1464	1	180,6
	JOSE MANUEL OCAMPO	3378	1	180,6
	SAN FRANCISCO DE ASIS	2977	6	1.083,6
	CRISTIANA	45	1	230,6
	GUILLERMO PINTO ISMODES	3277	4	722,4
	JOSE OLAYA BALANDRA	3137	9	1.625,4
	MANUEL GONZALES PRADA	3462	1	140,6
	MANUEL VIVANCO ALTAMIRANO	2868	13	2.347,8
	ANDRES AVELINO CACERES DORREGARAY	3662	1	180,6
	TRILCE	3193	1	180,6
	JUAN PABLO VIZCARDO Y GUZMAN	3250	1	180,6
	JUAN ANTONIO LIGARDA PINEDA	3310	2	361,2
	FEDERICO VILLARREAL	2903	1	180,6
	PROCERES INDEPENDENCIA AMERICANA	2815	10	1.806
	VIRGEN DEL CARMEN	2954	15	2.709
	JOSE MARIA ARGUEDAS	3117	3	541,8
	ANTONIO RAIMONDI	3181	3	541,8

Fuente: Elaboración propia.

CONCLUSIONES

- Se realizó el modelamiento de la base de datos considerando las reglas de modelamiento usando el software Mysql Workbench 6.3 En base a las reglas de negocio.
- El cargado de la base de datos al Administrador de base de datos Mysql, fue una labor que requirió bastante esfuerzo ya que las tablas estaban en formato .xls y algunos registros no eran uniformes, lo que incremento el tiempo de cargado de la base de datos.
- Luego de la carga de datos las dimensiones que se implementaron son: colegio, inscribe, pagos.
- Para realizar la integración se usaron los softwares Pentaho, Mysql y Esquema workbench.
- El cubo OLAP obtenido indica la cantidad de postulantes por colegio, con lo que podemos tomar decisiones sobre el tema de difusión en los colegios que aportan más postulantes a nuestra universidad.

RECOMENDACIONES

- El modelamiento se realizó para visualizar los colegios que aportan más postulantes lo que conlleva a calcular los mayores ingresos en el centro preuniversitario, sin embargo sería preciso que en posteriores investigaciones se pueda modelar una base de datos que sea el soporte de todo un sistema integral del Centro preuniversitario.
- Se recomienda que para ingresar los datos a las tablas esta debe ser una labor que involucre integridad de la información porque cuando se tiene errores en los registros no se puede obtener resultados para los cubos OLAP.
- Para obtener información es necesario crear las dimensiones, éstas pueden ser incrementadas para tener más datos para la toma de decisiones.
- Sería preciso proponer otros trabajos en creación de cubos OLAP pero con otras herramientas más integradas es decir que eviten usar esquemas Workbench.
- Un trabajo adicional a esta investigación es conseguir incrementar sobre lo desarrollado un cuadro de mando integral para el centro preuniversitario de la UNAJMA.

REFERENCIAS BIBLIOGRÁFICAS

- Arenas Lopez, M., & Gomez Montes, A. (2017). *Inteligencia de negocios aplicada a los procesos de autoevaluación de la universidad de manizales*. Manizales: Universidad de Manizales.
- Baron Lopez, F., & Tellez Montiel, F. (2013). *Apuntes y videos de Bioestadística*. Recuperado el 02 de Junio de 2017, de Sitio web Apuntes y videos de Bioestadística: <https://www.bioestadistica.uma.es/baron/apuntes/ficheros/cap08.pdf>
- Bernabeu, D. (06 de Noviembre de 2018). *Metodología para la construcción de un Datawarehouse*. Obtenido de <http://www.bussinessintiligencie.info>
- Bouckaert, R., Frank, E., Hall, M., Kirkby, R., Reutemann, P., Seewald, A., & Scuse, D. (2012). *WEKA Manual for version 3-6-7*. Hamilton: the University of Waikato.
- Cabezas Gutierrez, M., & Peña, B. F. (21 de Octubre de 2012). *Scielo*. Recuperado el 01 de Junio de 2017, de Sitio web de Scielo.org: <http://www.scielo.org.co/pdf/rudca/v15n2/v15n2a15.pdf>
- Cañon Bolaño, P., Osorio Ospina, L., & Pinzón Chica, C. (2014). *Balanced Scorecard como modelo articulador entre la estrategia organizacional y el sistema de*. Medellin: Universidad de Medellin.
- Casado Fernandez, C. (2013). *Manual básico de Matlab*. Servicios informáticos UCM.
- Curto Díaz, J., & Conesa, J. (2011). *Introducción al Business Intelligence*. Barcelona: Editorial UOC.
- Duro Novoa, V., & Pérez Cuevas, C. (2016). Inteligencia de negocios y sistema de soporte a las decisiones de la gestión económica en la Universidad de la Habana. *3C TIC*, 38-54.
- Eppen, G., Gould, F., Schmindt, C., More, J., & Weatherford, L. (2000). *Investigación de operaciones en la Ciencia Admistrativa* (Quinta ed.). Mexico: PEARSON.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (31 de July de 1997). *Computer Science*. Recuperado el 01 de Junio de 2017, de sitio web de Computer Science: <https://www.csd.uwo.ca/faculty/ling/cs435/fayyad.pdf>
- Fernandez Castillo, E., Molerio Pérez, O., Grau, R., & Cruz Peña, A. (19 de Julio de 2016). *ELSEVIER*. Recuperado el 03 de Junio de 2017, de sitio web de ELSEVIER: www.elsevier.es/es-revista-ansiedad-estres-242-pdf-S1134793716300094-S300
- Forrester, J. (1992). *Dinámica Industrial*. Buenos Aires: El Ateneo.
- Fuentes Tapia, L., & Valdivia Pinto, R. (2010). Incorporación de ionteligencia de negocios en el proceso de admsión y matrícula de una universidad chilena. *Ingeniare*, 383-394.

- González López, R. (2011). *Impacto de la Datawarehouse e inteligencia de negocios en el desempeño de las empresas: Investigación empírica en Perú como país en vías de desarrollo*. Lima: ESAN.
- Greenwood, W. (1969). *Decision Theory and Information System – An Introduction To Management Decision*. Cincinnati, Ohio: South Western Publishing Co.
- Gutiérrez Camelo, A., Devia Llanos, M., & Tarazona Bermudez, G. (2016). Research inteligencia de negocios: estudio de caso sector tecnológico colombiano. *Redes de Ingeniería*, 156-169.
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2006). *Metodología de la investigación* (Cuarta ed.). México: McGraw-Hill/Interamericana Editores.
- Iannacone, J., & Alvarino, L. (30 de Mayo de 2011). *sisbib.unmsm.edu.pe*. Recuperado el 01 de Junio de 2017, de Sistema de bibliotecas y biblioteca central: <http://sisbib.unmsm.edu.pe/BVRevistas/neohel/v5n1/pdf/a08v5n1.pdf>
- Lara Torralbo, J. (2014). *Minería de datos*. Madrid - España: Centro de estudios financieros.
- Lara Torralbo, J. (24 de Julio de 2014). *UDIMA*. Recuperado el 02 de Junio de 2017, de Sitio web de la UDIMA: <http://www.udima.es/en/integracionymineria.html>
- Laudon, K., & Laudon, J. (2012). *Sistemas de Información gerencial*. Mexico: Pearson Educación.
- Le Moigne, J. (1974). *Les systèmes de décision dans les organisations*. París: PUF.
- Lluis Cano, J. (2010). (ESADE, Editor, & ESADE) Recuperado el 08 de Setiembre de 2019, de <http://itemsweb.esade.edu>: http://itemsweb.esade.edu/biblioteca/archivo/Business_Intelligence_competir_con_informacion.pdf
- Matamoros Manrique, J. (2015). *“IMPLEMENTACIÓN DE UN SISTEMA OLAP PARA REDUCIR LOS RIESGOS EN LA TOMA DE DECISIONES RESPECTO A LA EDUCACION POR PARTE DE LAS ORGANIZACIONES DEL ESTADO EN LA UNIDAD DE ESTADISTICA EDUCATIVA DELESTADO EN LA UNIDAD DE ESTADISTICA EDUCATIVA DEL MINISTERIO DE. Huancavelica-Perú : Universidad Nacional de Huancavelica.*
- Maya Pérez, P., Aguilar Cisneros, J., Zamora Rios, R., & Barrón Adame, J. (13 de Junio de 2016). *ECORFAN*. Recuperado el 03 de Junio de 2017, de sitio web de EORFAN: http://www.ecorfan.org/spain/researchjournals/Aplicacion_Cientifica_y_Tecnica/vol2num4/Revista_de_Aplicacion_Cientifica_y_Tecnica_V2_N4_5.pdf
- Mitchell, T. (1997). *Machine Learning*. USA: Mc Graw Hill.
- Navarro Ruiz, F. (18 de Septiembre de 2013). Recuperado el 01 de Agosto de 2019, de Universitat Oberta de Catalunya: <http://openaccess.uoc.edu/webapps/o2/bitstream/10609/26761/6/fnavarroruTFC0913memoria.pdf>

- OCDE. (2003). *Manual de Frascati 2002*. Paris.
- ORACLE. (2016). *¿Qué es inteligencia de negocios?* Recuperado el 03 de Junio de 2017, de Sitio web de Oracle:
http://www.oracle.com/ocom/groups/public/@otn/documents/webcontent/317529_esa.pdf
- Pacco Palomino, R., & Apaza Mamani, G. (2016). Sistema de gestión financiera basado en sistemas de información ejecutiva para vicerrectorado académico de una universidad privada de Lima este. *Revista de investigación universitaria*, 61-67.
- Rodriguez Rodriguez, J. (2010). *Fundamentos de Minería de datos*. Bogotá: Universidad distrital Francisco José de Caldas.
- Rodriguez Ulloa, R. (1994). *La sistémica, los sistemas blandos y los sistemas de información*. Lima: Universidad del Pacífico.
- Salazar Tataje, J. (2017). *IMPLEMENTACIÓN DE INTELIGENCIA DE NEGOCIOS PARA EL ÁREA COMERCIAL DE LA EMPRESA AZALEIA - BASADO EN METODOLOGÍA ÁGIL SCRUM*. Lima-Perú: Universidad San Ignacio de Loyola.
- Schackle, G. (1961). *Decision, Order and Time in Human Affairse*. Oxford: Oxford University Press.
- Slocum, H. &. (2004). *Comportamiento organizacional (10ª ed.)*. Mexico: Ediciones Paraninfo.
- Vélez Miranda, S., & Martínez Martínez, J. (2014). EL PROCESO DE INTELIGENCIA COMPETITIVAY EL INDICADOR CLAVE DE DESEMPEÑO, CAPACIDAD DE INNOVACIÓN, EN LA PEQUEÑA Y MEDIANA EMPRESA (PYMES) DE MANUFACTURA EN LA REGIÓN OESTE DE PUERTO RICO . *Universidad de Puerto Rico*, 1-23.
- Vizcaino Garzón, P. (2008). *Aplicación de técnicas de inducción de árboles de decisión a problemas de clasificación mediante el uso de weka (waikato environment for knowledge analysis)*. Bogota, Colombia: Fundación universitaria Konrad Lorenz.
- Wankel, S. &. (1989). *Administración (3ª ed.)*. Méjico: Prentice Hall.
- Winston, P. (1994). *Inteligencia Artificial*. Estados Unidos: Adison Wesley.
- Zarria Torres, C., Arce Ramos, C., & Lam Moraga, J. (30 de Mayo de 2016). *Dialnet*. Recuperado el 03 de Junio de 2017, de Dialnet:
<https://dialnet.unirioja.es/download/articulo/5608574.pdf>