



Universidad Nacional Mayor de San Marcos
Universidad del Perú. Decana de América
Facultad de Ingeniería de Sistemas e Informática
Escuela Académico Profesional de Ingeniería de Sistemas

**Implementación de un sistema de consultas analíticas
para el soporte de las decisiones en instituciones
educativas públicas basado en un Datamart, aplicando
la metodología HEFESTO. Caso de estudio: I.E. N°
170 Santa Rosa de Sauce**

TESIS

Para optar el Título Profesional de Ingeniero de Sistemas

AUTOR

Jorge Alberto VILCA YPANAQUÉ

ASESOR

César Augusto ANGULO CALDERÓN

Lima, Perú

2016



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Vilca, J. (2016). *Implementación de un sistema de consultas analíticas para el soporte de las decisiones en instituciones educativas públicas basado en un Datamart, aplicando la metodología HEFESTO. Caso de estudio: I.E. N° 170 Santa Rosa de Sauce*. [Tesis de pregrado, Universidad Nacional Mayor de San Marcos, Facultad de Ingeniería de Sistemas e Informática, Escuela Académico Profesional de Ingeniería de Sistemas]. Repositorio institucional Cybertesis UNMSM.

1318

2 (o.m.)

Xi
94



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
FACULTAD DE INGENIERIA DE SISTEMAS E INFORMATICA
Escuela Académico Profesional de Ingeniería de Sistemas

Acta de Sustentación de Tesis

Siendo las ^{6:00 PM}.....horas del día ²³...de Setiembre del año 2016, se reunieron los docentes designados como miembros de Jurado de la Tesis, presidido por el Lic. Luis Alberto Alarcón Loayza (Presidente), el Lic. Jorge Luis Chávez Soto (Miembro) y el Lic. Cesar Augusto Angulo Calderón (Miembro Asesor) para la sustentación de la Tesis Intitulada: **"IMPLEMENTACIÓN DE UN SISTEMA DE CONSULTAS ANALÍTICAS PARA EL SOPORTE DE LAS DECISIONES EN INSTITUCIONES EDUCATIVAS PÚBLICAS BASADO EN UN DATAMART, APLICANDO LA METODOLOGÍA HEFESTO. CASO DE ESTUDIO: I.E. N° 170 SANTA ROSA DE SAUCE"**. Por el Bachiller: VILCA YPANAQUÉ, JORGE ALBERTO; para optar el Título Profesional de Ingeniero de Sistemas.

Acto seguido de la exposición de la Tesis, el presidente invitó al graduado a dar las respuestas a las preguntas establecidas por los Miembros del Jurado.

EL graduado en el curso de sus intervenciones demostró pleno dominio del tema, al responder con acierto y fluidez a las observaciones y preguntas formuladas por los señores miembros del Jurado.

Finalmente habiéndose efectuado la calificación correspondiente por los miembros del Jurado, el graduado obtuvo la nota de.....^{A.F.}..... (En letras).....^{Diecisiete}.....

A continuación el Presidente de Jurados el Lic. Luis Alberto Alarcón Loayza, declara al graduado **Ingeniero de Sistemas**.

Siendo las ^{6:50 PM}..... Horas, se levantó la sesión.

.....
Lic. Luis Alberto Alarcón Loayza

.....
Miembro
Lic. Jorge Luis Chávez Soto

.....
Miembro Asesor
Lic. Cesar Augusto Angulo Calderón

FICHA CATALOGRÁFICA

VILCA YPANAQUÉ, Jorge Alberto

“Implementación de un sistema de consultas analíticas para el soporte de las decisiones en instituciones educativas públicas basado en un datamart, aplicando la metodología HEFESTO. Caso de estudio: I.E. N° 170 Santa Rosa de Sauce”

Inteligencia de Negocios
(Lima, Perú 2016)

Tesis, Facultad de Ingeniería de Sistemas e Informática, Pregrado, Universidad Nacional Mayor de San Marcos.

DEDICATORIA

Dedico esta tesis a Dios, por haberme dado esta grandiosa oportunidad y por las fuerzas que me brinda para seguir adelante. A mi familia por apoyarme incondicionalmente en el descubrimiento de mi vida y formación como persona, por su esfuerzo, dedicación, paciencia y sabiduría que han compartido conmigo.

AGRADECIMIENTOS

Agradezco a Dios por la fuerza y coraje con la que logro mis metas, a mis padres, hermana, tia y abuela por confiar en mí, y haberme apoyado en las decisiones que he tomado. A todos los docentes que han dictado mis clases tanto en el colegio como en la universidad. Finalmente, agradezco a la UNMSM, por ser una institución que mantiene sus valores y calidad.

EPIGRAFE

“Vive como si fueras a morir mañana;
aprende como si el mundo fuera a durar
para siempre.”

Mahatma Gandhi

RESUMEN

Título	:	Implementación de un sistema de consultas analíticas para el soporte de las decisiones en instituciones educativas públicas basado en un datamart, aplicando la metodología HEFESTO. Caso de estudio: I.E. N° 170 Santa Rosa de Sauce
Autor	:	Vilca Ypanaqué, Jorge Alberto
Asesor de Tesis	:	Angulo Calderón, César
Fecha	:	Septiembre, 2016

En las instituciones educativas, los docentes utilizan diferentes metodologías de enseñanza, por lo que sus estudiantes logran obtener diferentes resultados. En este sentido, los resultados de los estudiantes, pueden definir si un docente usa sus metodologías adecuadamente o no. Por otro lado, en las instituciones educativas se realiza un proceso periódico de evaluación docente (conocido como monitoreo docente) que permite verificar la utilización de lineamientos proporcionados por el ministerio de educación (plan curricular y recomendaciones de metodologías). Estos datos se generan en grandes cantidades en las instituciones educativas públicas, y los tiempos dentro de las mismas son tan cortos que muchas veces no es posible realizar el análisis correspondiente de la información, desde este punto de vista, este proyecto busca brindar una herramienta que, gracias a la utilización de un sistema de consultas analíticas, facilite el acceso a esta información de manera centralizada y automática.

Cuando se califica a un estudiante, se utilizan diferentes criterios que deben cumplirse para que sea merecedor de dicha calificación, esta indica que los estudiantes logran satisfacer determinado criterio gracias a su esfuerzo personal, pero el factor externo de la enseñanza que reciben tiene su parte en estos resultados. Es por esta razón que se considera, en este trabajo, la medición del resultado de los estudiantes como el resultado de las metodologías utilizadas por el docente.

Por otro lado, en las instituciones educativas se realiza el monitoreo docente, esta actividad busca evaluar la metodología de enseñanza que se utiliza en las aulas de clases de manera objetiva desde el punto de vista de otros docentes y basándose en las lineaciones brindadas por el ministerio de educación (plan curricular y recomendaciones de metodologías). Este criterio permite tener una visión objetiva de la metodología que el docente usa.

Palabras claves

Monitoreo docente, Datamart, Instituciones educativas.

ABSTRACT

Title	:	Implementation of an analytics system to support decision taking in public schools based on a datamart, by applying the HEFESTO methodology. Case study: I.E. N° 170 Santa Rosa de Sauce
Author	:	Jorge Alberto Vilca Ypanaqué
Thesis advisor	:	César Angulo Calderón
Date	:	September, 2016

In schools, teachers use different teaching methodologies, this is why students get different results, in this way, student's results can define if a teacher uses it's methodologies in a good manner or doesn't. Besides in schools there is a process of teacher's evaluation, this evaluation is driven periodically and is intended to verify that teachers use the guidelines provided by the education authorities. This data is massively generated in public schools and time available to process this data is so short that often it is not analyzed, from this point of view, this project aims to provide easy access to information in a centralized and automatic manner via analytics.

When a student is graded, teachers use different criteria that students must accomplish so that they can get their grade, students deserve these grades thanks to their personal effort, but the teaching they get has an important impact in their results. That's why in this work the student's results are considered as result of the methodologies used by teachers.

On the other hand, in public schools there is a process of teacher's evaluation, this activity targets to objectively evaluate the teaching methodology they use in their class rooms, having the opinion of other teachers and taking into consideration the guidelines provided by the Peruvian education authorities.

Key words

Teacher Evaluation, Datamart, Public Schools

ÍNDICE

DEDICATORIA	I
AGRADECIMIENTOS	II
EPIGRAFE	III
RESUMEN	IV
ABSTRACT	V
CAPÍTULO I. INTRODUCCIÓN	1
1.1 Título de la Tesis	1
1.2 Resumen	1
1.3 Planteamiento del problema	2
1.3.1 Antecedentes	2
1.3.2 Formulación del problema	4
1.3.3 Justificación	5
1.4 Alcances	5
1.5 Objetivos	6
1.5.1 Objetivo General	6
1.5.2 Objetivos Específicos	6
CAPÍTULO II. MARCO TEÓRICO	7
2.1 Definiciones teóricas	7
2.1.1 Data Warehouse	7
2.1.2 Datamart	8
2.1.3 OLAP	8
2.1.4 Proceso ETL (Extracción, Transformación y Carga)	10
2.2 Marco normativo	12
2.2.1 Diseño curricular nacional de la educación básica regular	12
2.2.2 Manual de construcción de un Data Warehouse	13
2.3 Metodologías de desarrollo de Data warehouse	13
2.3.1 Diseño del data warehouse por Bill Inmon	13
2.3.2 Ciclo de vida de Ralph Kimball	15
2.3.3 Metodología para la construcción de un Data Warehouse (HEFESTO)	18
2.3.4 Metodología DWEP (Data Warehouse Engineering Process)	22

CAPÍTULO III. ESTADO DEL ARTE	24
3.1 Herramientas para la construcción de datamart	24
3.1.1 Oracle Business Analytics	24
3.1.2 Pentaho	24
3.1.3 Microsoft SQL Integration Services y Analysis Services	25
3.1.4 MicroStrategy	26
3.1.5 Jaspersoft	27
3.1.6 QlikView	28
3.2 Trabajos realizados	29
3.2.1 Construcción de una herramienta de desarrollo de soluciones para inteligencia de negocios – explotación (Falcón Cisneros, Palma Stanciuc, & Celi Peña, 2007)	29
3.2.2 Solución de inteligencia de negocios en el área de finanzas en la municipalidad de metropolitana de Lima	30
3.2.3 Escuelas públicas de Tacoma	30
3.2.4 Soluciones de Inteligencia de negocios en la práctica: apoyo a la toma de decisiones en proyectos educativos para población infantil vulnerable en el caribe Colombiano	31
CAPÍTULO IV. APORTE TEÓRICO	33
4.1 Metodología de desarrollo	33
4.2 Herramienta de procesamiento analítico	35
4.3 Arquitectura de la solución	37
4.4 Herramientas tecnológicas	38
4.5 Especificaciones técnicas	38
CAPÍTULO V. APORTE PRÁCTICO	40
5.1 Evaluación de la organización	40
5.1.1 Contexto del Negocio	40
5.1.2 Visión	40
5.1.3 Misión	40
5.1.4 Factores Externos	40
5.1.5 Factores Internos	41
5.2 Análisis del negocio	43
5.2.1 Casos de uso del negocio	43
5.2.2 Diagrama de casos de uso del negocio	44
5.2.3 Modelo de análisis del negocio	45
5.2.4 Diagramas de actividades	45
5.2.5 Requisitos del proyecto	48
5.3 Diseño del datamart - Aplicación de la Metodología HEFESTO	49
5.3.1 Análisis de requerimientos	49

5.3.2	Análisis de los OLTP	55
5.3.3	Modelo Lógico del datamart	58
5.3.4	Definición del proceso ETL	60
5.3.5	Creación de cubos multidimensionales	68
5.4	Implementación de consultas con SAIKU BI	70
CAPÍTULO VI. RESULTADOS		75
6.1	Resultados	75
CAPÍTULO VII. CONCLUSIONES Y TRABAJOS FUTUROS		77
7.1	Conclusiones	77
7.2	Recomendaciones y trabajos futuros	78
REFERENCIAS BIBLIOGRÁFICAS		79
ANEXOS		82
ANEXO 01: EVALUACIÓN DE LA ORGANIZACIÓN		83
ANEXO 02: FICHA DE MONITOREO 2014		91

Índice de figuras

Ilustración 1 SIAGIE, Múltiples Vistas.	3
Ilustración 2 Aplicaciones transaccionales y de toma de decisiones.	7
Ilustración 3 Ejemplo de Modelo estrella.	9
Ilustración 4 Ejemplo de modelo de copo de nieve.	10
Ilustración 5 Ejemplo de modelo de Constelación.....	10
Ilustración 6 División entre gestión de datos y capa de presentación de un data warehouse.....	11
Ilustración 7 Pasos del proceso de staging.	12
Ilustración 8 Ejemplo de niveles de esquematización que podría encontrarse en un data warehouse.	13
Ilustración 9 Contraste entre el ciclo de vida del desarrollo de sistemas (SDLC) y la propuesta de Inmon (CLDS).	14
Ilustración 10 Elementos básicos del data warehouse.	15
Ilustración 11 Diagrama del ciclo de vida del negocio dimensional.	16
Ilustración 12 Cuadrante de priorización.	17
Ilustración 13 Metodología HEFESTO, pasos.....	19
Ilustración 14 Ejemplo de modelo conceptual.....	20
Ilustración 15 Arquitectura utilizada por una solución en MSSQL Server.	26
Ilustración 16 Arquitectura planteada por MicroStrategy.....	27
Ilustración 17 Arquitectura propuesta por jaspersoft.....	28
Ilustración 18 Arquitectura QlikView.....	29
Ilustración 19 Arquitectura planteada para la implementación de la solución.	37
Ilustración 20 Topología de la propuesta.....	38
Ilustración 21 Monitoreo Educativo.....	42
Ilustración 22 Desarrollo de año escolar.....	43
Ilustración 23 Diagrama de casos de uso del negocio.....	44
Ilustración 24 Modelo de análisis del negocio.....	45
Ilustración 25 Diagrama de actividades de monitoreo docente.	46
Ilustración 26 Diagrama de actividades de la evaluación a los estudiantes.....	47
Ilustración 27 Diagrama de actividades de procesamiento de resultados.....	47
Ilustración 28 Diagrama de casos de uso.....	49
Ilustración 29 Información solicitada por el ministerio de educación en la matriz de monitoreo - consolidado. Fuente: Ministerio de Educación.....	50
Ilustración 30 Detalle de resultados de monitoreo solicitado por el ministerio de educación.....	50
Ilustración 31 Datos solicitados por el ministerio de educación relacionada a las calificaciones de los estudiantes.....	51
Ilustración 32 Modelo Conceptual Calificación Estudiante.....	54
Ilustración 33 Modelo Conceptual Calificación Monitoreo.....	54
Ilustración 34 Modelo OLTP.....	56
Ilustración 35 Diagrama estrella para los hechos de calificaciones.....	59

Ilustración 36 Diagrama en estrella para los hechos de monitoreo.....	60
Ilustración 37 Formato de archivos excel procesados para la carga inicial.	61
Ilustración 38 Formato PDF de calificaciones exportado desde el SIAGIE.....	61
Ilustración 39 Proceso ETL con Pentaho Data Integration.....	62
Ilustración 40 Transformación Carga Periodo	62
Ilustración 41 Consulta de carga de datos para ETL en Pentaho Data Integration.	63
Ilustración 42 Procedimiento definido para la carga de datos en la tabla de hechos de calificaciones en Pentaho Data Integration.	63
Ilustración 43 Consulta para carga de hechos de calificaciones.	64
Ilustración 44 Consulta para carga de hechos monitoreo.....	65
Ilustración 45 Job de migración de calificaciones modificado con validación de datos.	67
Ilustración 46 Validación de calificaciones.	67
Ilustración 47 Validación de datos en migración de dimensiones.	68
Ilustración 48 Schema de definición de cubos en Schema Workbench de Pentaho.	69
Ilustración 49 Definición de jerarquías para el periodo de calificación.....	70
Ilustración 50 Vista de calificaciones promedio por periodo, grado, sección y curso en saiku.....	71
Ilustración 51 Vista comparativa de calificaciones promedio en estudiantes por año en secciones diferentes. Fuente: (Elaboración propia).....	72
Ilustración 52 Cuadro comparativo que evidencia la calificación promedio en los estudiantes en tres periodos académicos.	72
Ilustración 53 Gráfico de las calificaciones de los estudiantes en diferentes periodos.	73
Ilustración 54 Resultados obtenidos por rangos en tres periodos académicos por curso, grado y sección. Fuente: (Elaboración propia).....	73
Ilustración 55 Puntajes promedios de los monitoreos por periodo y aspecto observado en SAIKU.	74
Ilustración 56 Número de docentes que cumplen un indicador por periodo y grado en SAIKU.....	74

Índice de Tablas

Tabla 1 Porcentajes de Nivel de Desaprobados y Repitentes del 2014.	4
Tabla 2 Motores de bases de datos soportados por la solución.....	30
Tabla 3 Cuadro resumen Objetivos, Dificultades y Beneficios del proyecto.	32
Tabla 4 Metodologías para BI y sus componentes.....	32
Tabla 5 Comparación de metodologías.....	33
Tabla 6 Procesos para la implementación del datamart basado en la metodología HEFESTO.	34
Tabla 7 Comparación de herramientas.....	36
Tabla 8 Casos de uso del negocio identificados.....	43
Tabla 9 Requisitos funcionales.	48
Tabla 10 Requisitos no funcionales.	48
Tabla 11 Requerimientos de información.	52
Tabla 12 Indicadores y perspectivas.	53
Tabla 13 Conformación de indicadores.	55
Tabla 14 Correspondencias de datos OLTP – ETL.	56
Tabla 15 Descripción de Tablas de dimensiones-.....	58
Tabla 16 Descripción de Tabla de hechos.....	59
Tabla 17 Criterios para limpieza de datos.....	66
Tabla 18 Comparativa proceso actual vs proceso propuesto.	75

CAPÍTULO I. INTRODUCCIÓN

Este capítulo presenta la formulación del problema, presentando sus antecedentes, justificación y alcance. Además, se especifica el objetivo de la investigación tanto el objetivo general como los objetivos específicos.

1.1 Título de la Tesis

“Implementación de un sistema de consultas analíticas para el soporte de las decisiones en instituciones educativas públicas basado en un datamart, aplicando la metodología HEFESTO. Caso de estudio: I.E. N° 170 Santa Rosa de Sauce”

1.2 Resumen

Últimamente, las instituciones educativas han sido parte de un proyecto ambicioso del Ministerio de Educación en el cual se provee de una plataforma uniforme de gestión de instituciones educativas. Esta solución se llama SIAGIE (Sistema de Información de Apoyo para la Gestión de Instituciones Educativas) y tiene como principal función dar soporte a las funciones operacionales de las instituciones educativas a lo largo del territorio peruano.

Otra de las características de este sistema es que brinda información vital para los altos niveles de gestión como son el Ministerio de Educación, las direcciones regionales y las UGEL; sin embargo, las Instituciones Educativas no son consideradas en esta cadena de toma de decisiones y son vistas como proveedores de información. Siendo una institución educativa el punto de acción más cercano a los usuarios (los estudiantes y padres de familia), la información que necesita para gestionar los centros educativos es necesaria para poder guiar el personal a la mejora de la calidad educativa en las aulas.

Tomando en cuenta el poder de influencia del ambiente sobre los estudiantes, la institución educativa junto con los padres de familia, pueden hacer la diferencia en el resultado a fin de año de los estudiantes. Es por esta razón que el siguiente trabajo busca proveer a las instituciones educativas con información más allá de las notas; que les permita realizar un análisis de metas alcanzadas de los estudiantes y proporcionar un análisis de brechas del desarrollo de los mismos a nivel individual y de aula.

Esta información proporcionará un análisis del estilo de desarrollo de temas de los docentes permitiéndoles saber el nivel de desarrollo de sus estudiantes, las tendencias respecto a los temas que se realizan en clase y les permitirá poder realizar cambios en sus metodologías de enseñanza adecuándolas al grupo de estudios que dirige.

1.3 Planteamiento del problema

En la siguiente sección, se detallarán los puntos básicos del proyecto de tesis con el fin de dar a conocer las características del mismo.

1.3.1 Antecedentes

En el Perú, la educación pública fue considerada asunto de estado, desde el año 1837; año en el cual el Capitán Don Andrés de Santa Cruz decide crear un Ministerio con el título de Ministerio de Instrucción Pública, Beneficencia y Negocios Eclesiásticos. Actualmente, dicho ministerio se conoce como Ministerio de Educación; el Ministerio de Educación tiene como misión la *promoción del desarrollo de la persona humana, a través de un nuevo sistema educativo en el cual las capacidades individuales se vean fortalecidas, gracias a una formación integral y permanente*. Para cumplir con su misión, el Ministerio de Educación cuenta aproximadamente con 94941 instituciones educativas en el territorio nacional de las cuales 72171 (76 %) son instituciones públicas.

El sistema educativo actual se organiza en tres niveles organizacionales a nivel nacional. La institución principal es el Ministerio de Educación, que se encarga del aspecto normativo y de gestión a nivel macro de la educación del Perú, la Dirección Regional de Educación (DRE), la cual es una instancia intermedia de gestión educativa que garantiza el servicio educativo y promueve el desarrollo del educando comprometido con el desarrollo regional. En un nivel más modular, encontramos las Unidades de Gestión Educativa Local (UGEL), la cual es un órgano descentralizado del Ministerio de Educación con autonomía y responsable de la administración y desarrollo de la educación de las instituciones educativas públicas y privadas que se encuentren dentro de la localidad asignada. En este marco, cada Institución Educativa es asignada, de acuerdo a su localización, a una determinada UGEL, la cual pertenece a una región gestionada por una DRE y que es normada por el Ministerio de Educación.

En el marco de la tecnología, el Ministerio de Educación ha implementado un Sistema de Información de Apoyo a la Gestión de la Institución Educativa (SIAGIE); sin embargo, no ha proporcionado las utilidades informáticas para que se pueda realizar una gestión adecuada dentro de la Institución Educativa.



• Herramienta gratuita • Acceso 24/7 vía Internet • Soporte multiplataforma

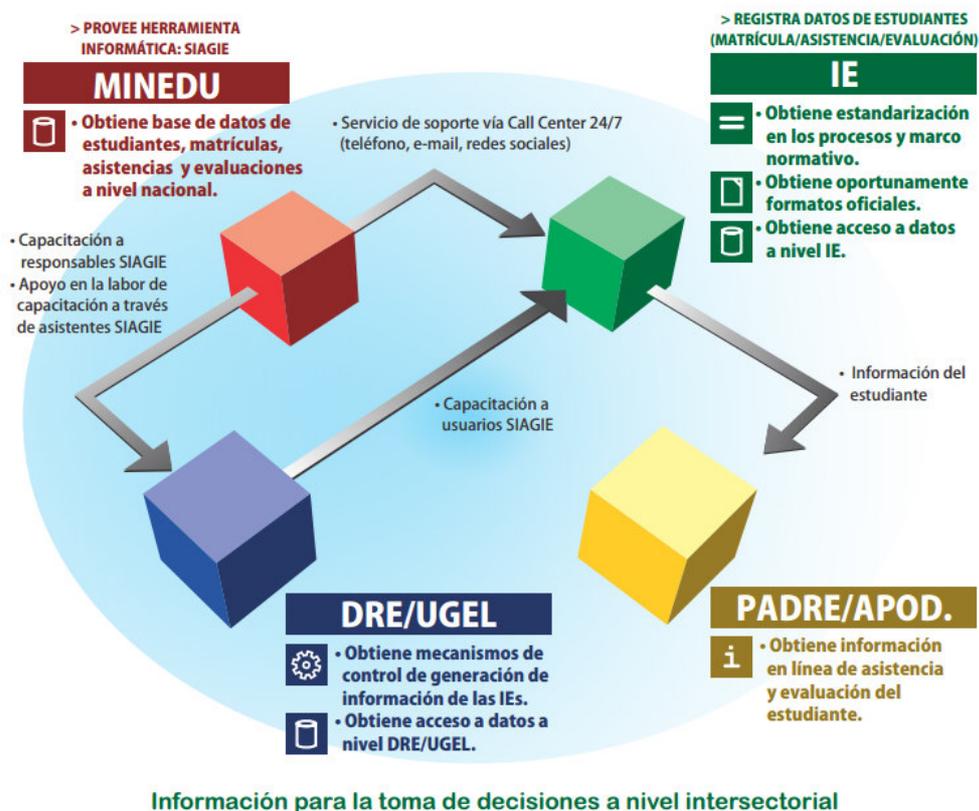


Ilustración 1 SIAGIE, Múltiples Vistas.

Fuente: (MINEDU, 2016)

Como se observa en la Ilustración 1, vemos que el SIAGIE, tiene como objetivo, en las instituciones educativas, la normalización de procesos y consolidar la información en un solo repositorio. De acuerdo con el Plan Estratégico de Tecnologías de la Información (PETI) del Ministerio de Educación, se tiene como objetivo cumplir con los siguientes proyectos, (MINEDU, 2016):

- Sistema de Información de la Gestión de Instituciones Educativas
- Sistema de Información de Instancias de Gestión Intermedia

- Sistema de Información de la Acreditación y Certificación Educativa

Los sistemas mencionados anteriormente apoyan a la gestión interna de los datos y procedimientos internos de las Instituciones Educativas.

De acuerdo con ESCALE, Estadística de la Calidad Educativa, en Lima metropolitana, los porcentajes de desaprobados en los niveles primario y secundario se muestran en la Tabla 1:

Tabla 1 Porcentajes de Nivel de Desaprobados y Repitentes del 2014.

Fuente (MINEDU, 2016)

PRIMARIA							
%	Total	1er Grado	2do Grado	3er Grado	4to Grado	5to Grado	6to Grado
Desaprobados	2.7	0.1	4.7	3.6	3.0	2.9	2.0
Repitentes	1.5	0.3	2.8	1.9	1.5	1.4	0.8
SECUNDARIA							
	Total	1er Grado	2do Grado	3er Grado	4to Grado	5to Grado	
Desaprobados	6.6	8.2	8.0	7.2	5.6	3.9	
Repitentes	3.0	4.1	4.0	3.2	2.2	1.2	

Como podemos observar en la Tabla 1, los datos estadísticos proporcionados por ESCALE, en el 2014, se ha registrado un porcentaje de desaprobados y repitentes bajo pero considerable en nuestra educación; sin embargo, con la información correcta, el momento indicado, estos valores podrían ser disminuidos a través de políticas institucionales que corrijan los errores y de esta manera mejorar el nivel educativo peruano.

1.3.2 Formulación del problema

En las instituciones educativas públicas el proceso de retroalimentación y mejora continua se realiza de manera manual y tradicional, lo que genera que se deba realizar un trabajo adicional para poder conocer cuáles son los resultados de la institución educativa, sus estudiantes y de las metodologías utilizadas por los docentes. Este trabajo adicional necesita un tiempo que es empleado para la

recopilación y explotación de los resultados de los estudiantes y de los monitoreos docentes.

En este sentido, las reuniones son pospuestas o llevadas con datos incompletos, además, al requerir de un tiempo adicional para plantear los resultados, se debe hacer uso del tiempo de otra actividad que muchas veces puede ser perjudicial.

Además, con el pasar de los años, la cantidad de datos de estudiantes y docentes que se maneja dentro de la institución educativa crece, lo que nos genera datos históricos, estos, no son usados o son usados vagamente en la realización de las reuniones de retroalimentación por lo que los docentes están limitados a usar información actual sin tener una visión retrospectiva.

1.3.3 Justificación

La educación es un tema de relevante para el desarrollo de nuestro país; actualmente, las instituciones educativas son el primer ambiente donde los peruanos comenzamos nuestra formación. Teniendo en cuenta este alto nivel de importancia, este proyecto plantea brindar una herramienta que pueda apoyar a los encargados de guiarnos (docentes y directores) en la recopilación de información de sus instituciones educativas con la finalidad que puedan tener un amplio panorama de lo que ocurre dentro de estas y así tomar decisiones a tiempo.

Esta solución plantea utilizar datos que se usan actualmente como indicadores independientes (calificaciones de estudiantes y monitoreo docente) para evitar problemas de cotejamiento manual, y aprovechar la disponibilidad de las nuevas tecnologías automatizando el proceso de recolección y procesamiento manual de datos, lo que permite obtener información oportuna y exacta del estado de las instituciones educativas pudiendo de esta manera obtener mejores resultados.

1.4 Alcances

Este proyecto se orienta a describir el análisis, diseño e implementación de un datamart para dar soporte a Instituciones Educativas en la captura y análisis de información basado en las calificaciones de sus estudiantes y el monitoreo docente. Además de este documento, se obtendrán los diferentes modelos y la documentación necesaria que permitan el desarrollo del proyecto.

El sistema será capaz de obtener datos de las instituciones educativas públicas (caso de estudio: IE Santa Rosa del Sauce – San Juan de Lurigancho) de Lima – Perú a ser procesados para obtener información que permita a la administración de instituciones educativas su análisis para la corrección y guía en el proceso educativo de los estudiantes, apoyando en la comparación histórica de información.

1.5 Objetivos

A continuación, se detallan los objetivos que se intentan alcanzar para la culminación del presente proyecto. Se especifican dos tipos de objetivos entorno a los cuales gira el proyecto principalmente. Los objetivos generales, que determinan el logro del proyecto de tesis a su culminación y los objetivos específicos que detallan los logros que se realizarán a lo largo del proyecto y que contribuyen a la obtención del objetivo general.

1.5.1 Objetivo General

Implementación de un sistema de consultas analíticas utilizando un datamart para la captura y análisis de información sobre los resultados de las evaluaciones de estudiantes y monitoreos docentes que disminuya el tiempo de su recopilación y explotación y facilite el acceso a datos históricos, caso de estudio I.E. N° 170 Santa Rosa del Sauce.

1.5.2 Objetivos Específicos

- Evaluar las distintas metodologías para el desarrollo de un datamart existentes que puede ser usada en este proyecto.
- Evaluar cuál de las herramientas de procesamiento analítico existentes en el mercado que puede ser usada con mayor facilidad en este proyecto.
- Determinar qué modelo lógico de datamart se adecúa mejor para su utilización en este proyecto.
- Analizar los trabajos realizados en el estado del arte que dan solución a problemas relacionados.
- Analizar los procesos de evaluación de estudiantes y monitoreos docentes (en el contexto de la I.E. N° 170 Santa Rosa del Sauce) para poder modelarlos en el datamart e implementar un sistema de consultas analíticas utilizando un datamart para la captura y análisis de información de estudiantes y docentes en la I.E. N° 170 Santa Rosa del Sauce.

CAPÍTULO II. MARCO TEÓRICO

En esta sección se definirán los conceptos base que son necesarios para incursionar en el tema propuesto a tratar. La siguiente sección consta de definiciones comunes en el contexto de datamart y marco normativo para la solución planteada.

2.1 Definiciones teóricas

En este apartado veremos algunas de las definiciones básicas que nos permita entender la solución planteada al problema, además de familiarizarnos con la labor del desarrollo de data warehouse y datamart.

2.1.1 Data Warehouse

En (Nagabhushana, 2006) se define un data warehouse como un componente de arquitectura que convierte e integra grandes masas de datos de los sistemas operacionales en una empresa. Con este componente se busca obtener todos los datos centralizados, ordenados, resumidos e históricos que facilite información a quienes toman decisiones.

(Nagabhushana, 2006) recomienda que las aplicaciones de soporte a la toma de decisiones deben obtener datos del data warehouse; el cual, tiene todos los datos obtenidos de los sistemas transaccionales y otros datos externos a la empresa, como se muestra en Ilustración 2. De esta forma, las aplicaciones de apoyo a la toma de decisiones se alimentan con esta información, la analizan y la proveen a sus usuarios.

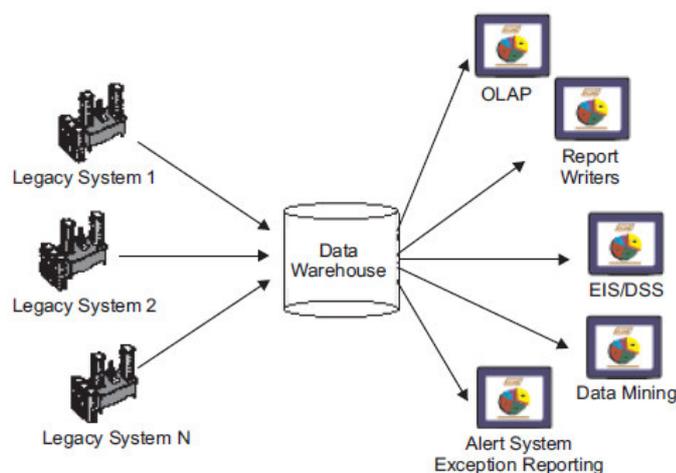


Ilustración 2 Aplicaciones transaccionales y de toma de decisiones.

Fuente: (Nagabhushana, 2006)

2.1.2 Datamart

Un datamart es un concepto de la inteligencia de negocios que simplifica el concepto de data warehouse debido a su complejidad y dificultad de implementación. Según (Nagabhushana, 2006), un datamart puede ser considerada como una porción del Data Warehouse para un área específica de la empresa que por motivos de tiempo, dinero u otros decide realizar una implementación por etapas de un data warehouse y comienza utilizando una estrategia bottom up (implementa datamart en áreas distintas hasta que completa todas y las integra en un data warehouse). Gracias al alcance limitado del presente proyecto, se define realizar una implementación de un datamart y no de un data warehouse

2.1.3 OLAP

OLAP es el acrónimo para Procesamiento Analítico en Línea (On-Line Analytical Processing) que representa una forma de ver los datos resumidos históricos en estructuras multidimensionales (conocidos como cubos OLAP) que permitan el acceso rápido a los mismos. Las aplicaciones OLAP proveen información de datos históricos y actuales de la empresa. OLAP permite a sus usuarios la creación de presupuestos, simulaciones y pronósticos. Como se menciona en (Bernabeu, 2010), existen varias formas de implementar un sistema OLAP, como lo son, el Sistema Relacional Clásico (ROLAP), el Multidimensional (MOLAP) y el Híbrido (HOLAP).

2.1.3.1 Cubos Multidimensionales

De acuerdo con (Bernabeu, 2010), los cubos multidimensionales son estructuras de datos almacenados en un sistema gestor de base de datos que consisten en tablas de hechos y dimensiones que permiten realizar el análisis de los indicadores, en las tablas de hechos, a través de diferentes dimensiones (tablas de dimensiones) y las relaciones que existen entre dichas dimensiones. Los cubos multidimensionales pueden ser implementados en diferentes esquemas (Estrella, Copo de Nieve y Constelación). Como sabemos, los datos que encontremos en el sistema transaccional serán extraídos, transformados y cargados en un nuevo sistema sobre el cual se está implementando el datamart, para que a partir de estos, se pueda

realizar las distintas consultas. Para esto la metodología escogida define tres posibles modelos para la base de datos multidimensionales (Bernabeu, 2010).

a) Modelo Estrella

El modelo en estrella consiste en tener en una tabla central del datamart, la tabla de hechos y las tablas de dimensiones relacionadas a esta mediante las claves foráneas que identifiquen cada indicador con sus respectivas cualidades en las tablas de dimensiones. En el modelo en estrella se considera importante la de-normalización de las tablas de dimensiones, evitando de esta forma los tiempos que toma realizar sentencias JOIN excesivas, en la Ilustración 3 se muestra un ejemplo de este modelo.

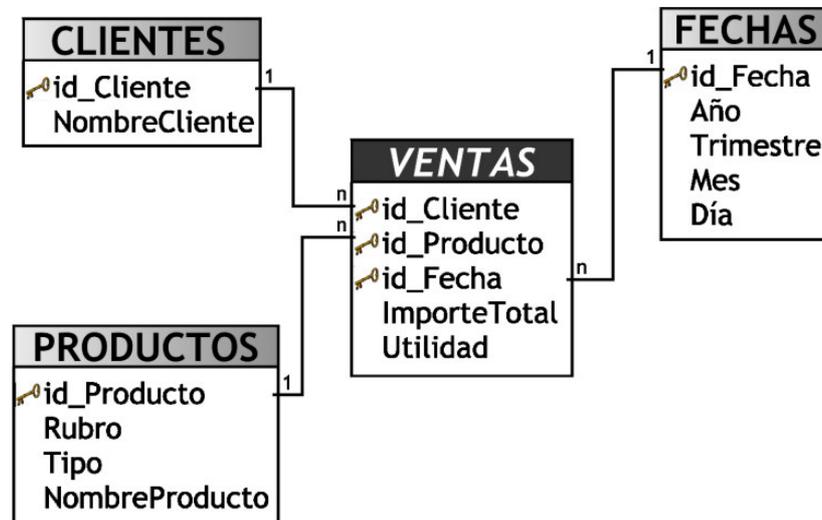


Ilustración 3 Ejemplo de Modelo estrella.

Fuente: (Bernabeu, 2010)

b) Modelo Copo de Nieve

El modelo de copo de nieve es una extensión del modelo estrella con la diferencia que las dimensiones se encuentran organizadas en más niveles. Este tipo de modelo usualmente se usa en modelos cambiantes y optimiza uso del espacio gracias a la normalización, en la Ilustración 4 se muestra un ejemplo del modelo de copo de nieve.

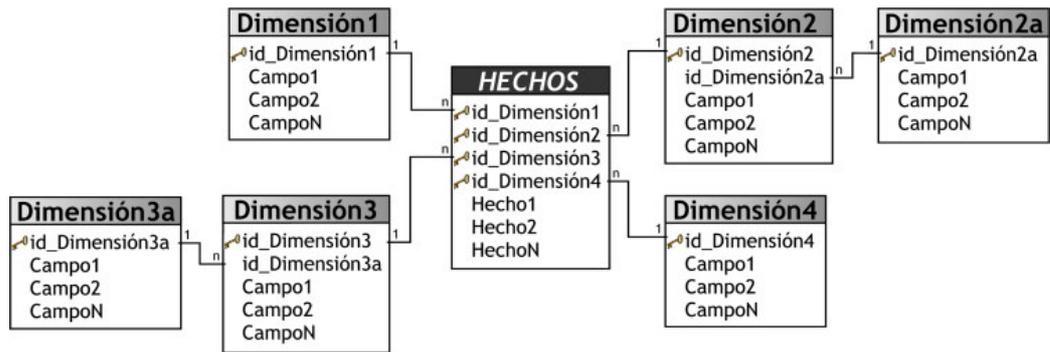


Ilustración 4 Ejemplo de modelo de copo de nieve.

Fuente: (Bernabeu, 2010)

c) Modelo Constelación o copo de estrellas

Este modelo es una extensión del modelo estrella con la diferencia que puede tener más de una tabla de hechos, en la Ilustración 5 se muestra un ejemplo del modelo de constelación.

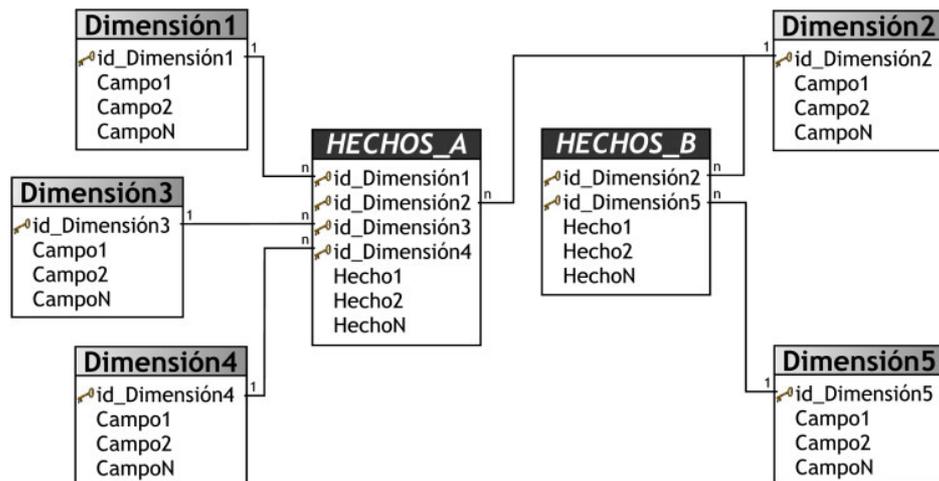


Ilustración 5 Ejemplo de modelo de Constelación.

Fuente: (Bernabeu, 2010)

2.1.4 Proceso ETL (Extracción, Transformación y Carga)

En (Kimball & Caserta, The Data Warehouse ETL Toolkit, 2004), los autores explican que un sistema ETL extrae datos desde los sistemas fuente, verifica los estándares de calidad y consistencia que los datos requieren para ser procesados y los entrega de tal manera que a partir de ellos se pueda construir las herramientas para la toma de decisiones. Este proceso no solo se refiere al transporte de los datos de una estructura a otra, por el contrario, durante este proceso se agrega valor a los datos en las siguientes actividades: (1) Elimina errores y corrige datos faltantes, (2) Integra diferentes fuentes de información y (3) Estructura los datos para que puedan

ser usadas por las herramientas de análisis. Durante este proceso los resultados parciales de la transformación necesitan ser guardados en disco en su totalidad o parcialmente, a este proceso se le conoce como staging; por esta razón, el proceso ETL también se conoce como “staging area”. En la Ilustración 6 los autores muestran la importancia del proceso ETL dentro del diseño de un data warehouse, en el primer bloque de gestión de datos se realiza la extracción y transformación de datos convirtiéndolos en información que se pone a disposición de la capa de presentación para que pueda realizar las consultas necesarias por los reportes.

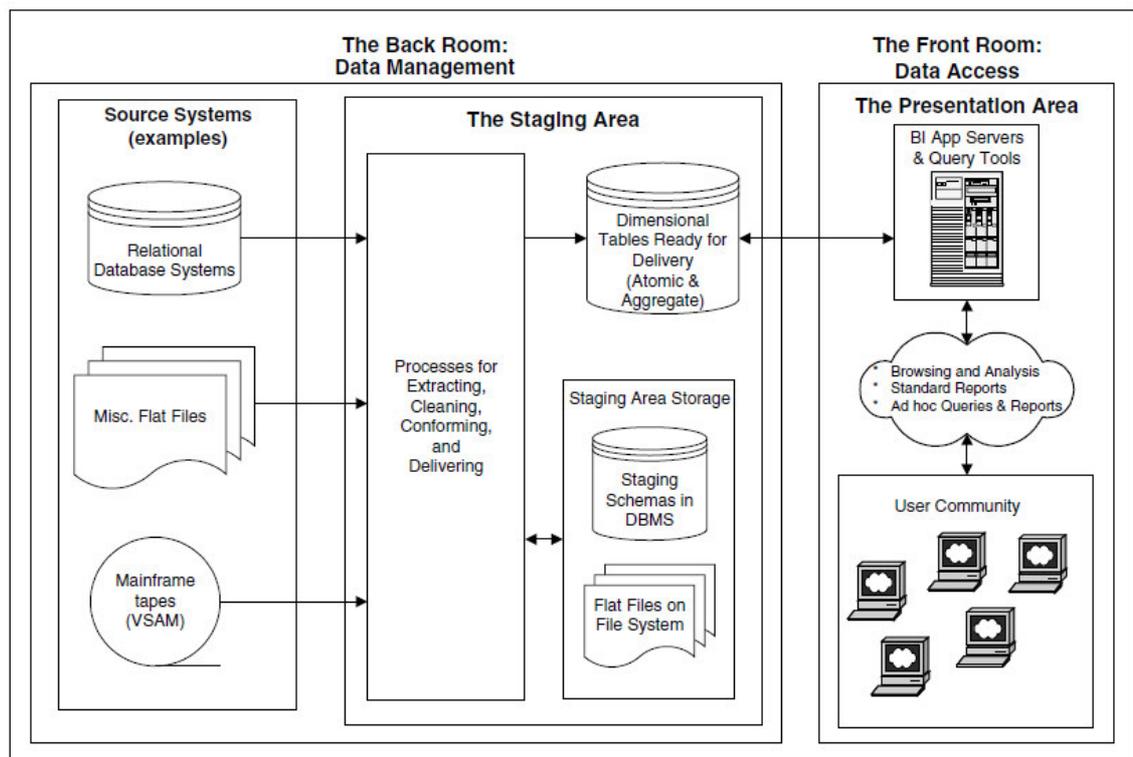


Ilustración 6 División entre gestión de datos y capa de presentación de un data warehouse.

Fuente: (Kimball & Caserta, *The Data Warehouse ETL Toolkit*, 2004)

Los autores (Kimball & Caserta, *The Data Warehouse ETL Toolkit*, 2004) proponen cuatro pasos que el proceso de staging debe seguir, primero la (1) extracción de datos obtiene datos crudos desde las diferentes fuentes establecidas, la (2) limpieza de datos, paso en el que los datos son validados, se verifica su consistencia y remueven duplicados, (3) integración de datos, paso en el que los datos de distintas fuentes son integradas de acuerdo con los estándares de medición establecidos en la organización y finalmente, (4) entrega de datos que hace referencia a estructurar los datos recolectados en los esquemas seleccionados para

dar soporte al data warehouse. Podemos ver que estos pasos son realizados uno después de otro en la Ilustración 7.

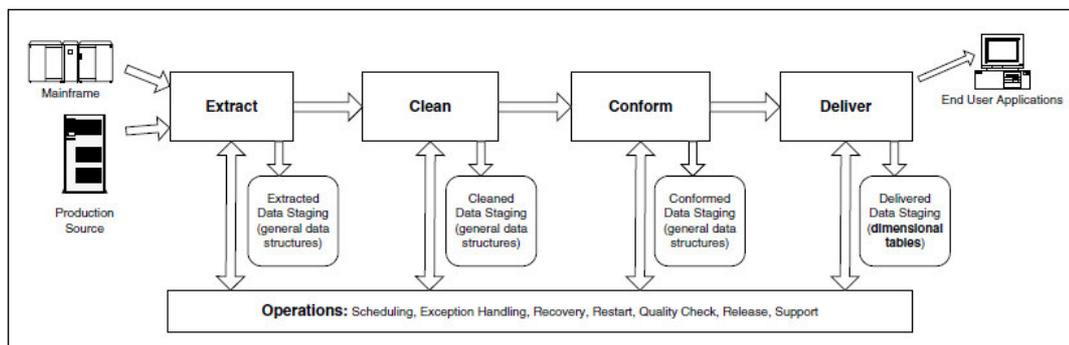


Ilustración 7 Pasos del proceso de staging.

Fuente: (Kimball & Caserta, The Data Warehouse ETL Toolkit, 2004)

2.2 Marco normativo

En esta sección veremos algunos de las restricciones y alcances dados por dos instituciones del estado que rigen el proceso educativo y de tecnologías, el Ministerio de Educación y el ONGEI - INEI (Oficina Nacional de Gobierno Electrónico e Informática), respectivamente.

2.2.1 Diseño curricular nacional de la educación básica regular

El diseño curricular de la educación básica regular es un documento que contiene elaborado por el Ministerio de Educación, este tiene como objetivo documentar los aprendizajes que los estudiantes deben desarrollar en cada nivel educativo. Se realiza este documento ya que de acuerdo con la ley general de la educación N° 28044, los currículos deben ser comunes en todo el país. El documento consta de tres partes: la primera parte tiene los objetivos de la educación básica regular; la segunda, presenta las áreas curriculares, los lineamientos de diversificación curricular y la evaluación de los aprendizajes; finalmente, la tercera parte, presenta las áreas curriculares por nivel educativo.

2.2.1.1 Nivel de Logro

El nivel de logro hace referencia a las competencias, capacidades, conocimientos y actitudes previstas en la programación curricular alcanzados por los estudiantes al final de un periodo o año académico.

Durante el proceso educativo, se necesita realizar una serie de evaluaciones para verificar que los conocimientos que se están impartiendo.

2.2.2 Manual de construcción de un Data Warehouse

Desarrollado por INEI en septiembre de 1997 (INEI, 1997), este manual desarrolla los puntos básicos y genéricos del proceso de desarrollo de un data warehouse. El documento explica los conceptos básicos relacionados con el data warehouse y además define un proceso modelo de gestión del proyecto de elaboración de un data warehouse. El proceso de elaboración del data warehouse, según esta metodología, de cuatro fases. La primera fase, la de organización, contempla la adaptación de la estrategia a seguir para llevar a cabo el desarrollo del data warehouse. Las siguientes dos fases, la de desarrollo e implementación definen los alcances del desarrollo y la estrategia de la implementación a utilizar. Finalmente se realiza, en la cuarta fase, una evaluación respecto al rendimiento de la inversión.

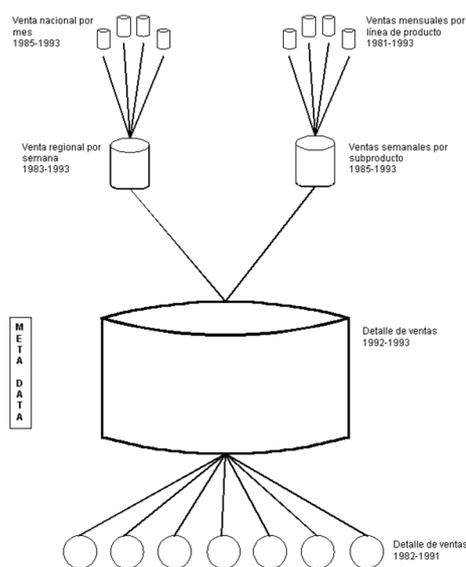


Ilustración 8 Ejemplo de niveles de esquematización que podría encontrarse en un data warehouse.

Fuente (INEI, 1997)

2.3 Metodologías de desarrollo de Data warehouse

En este apartado se comentará las metodologías existentes para el desarrollo de un data warehouse.

2.3.1 Diseño del data warehouse por Bill Inmon

Para (Inmon, 2002) los requisitos del data warehouse no se pueden conocer hasta que no tenga datos y sea usado, de igual manera, no se puede generalizar el diseño del data warehouse. El autor explica que el ciclo de vida del desarrollo de sistemas

(SDLC – System Development Life Cycle) convencional o algunas veces conocido como “cascada” no es compatible con ciclo de vida del desarrollo de un data warehouse, mientras que SDLC inicia con un conjunto de requisitos a partir de los cuales se realiza el diseño y desarrollo, la implementación de un data warehouse debe realizarse de la manera opuesta, (1) se comienza con la obtención de datos necesarios para poblar el data warehouse, luego (2) se integran los datos y (3) se realizan pruebas para ver que tendencias hay en los datos, a continuación (4) se desarrollan los programas necesarios y se analizan sus resultados. Finalmente (5) los requisitos del sistema son comprendidos, esta metodología es conocida también como espiral.

Este contraste entre el ciclo de vida propuesto en (Inmon, 2002) y el ciclo de vida tradicional se aprecia en Ilustración 9, donde las tareas comprensión de requisitos, análisis de resultados y diseño de sistemas de soporte a las decisiones son realizadas al final del ciclo de vida, y no al comienzo como se realizaría en una metodología en “cascada”.

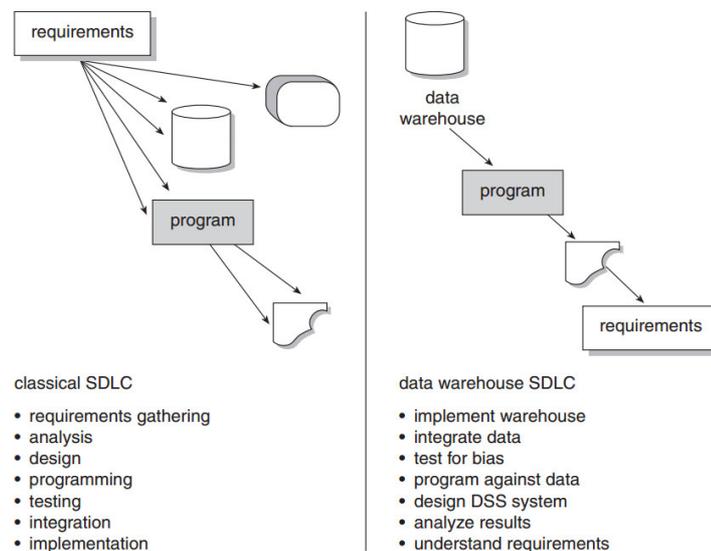


Ilustración 9 Contraste entre el ciclo de vida del desarrollo de sistemas (SDLC) y la propuesta de Inmon (CLDS).

Fuente: (Inmon, 2002).

(Inmon, 2002) comienza con el análisis de datos ya que las grandes organizaciones tienen un conjunto de aplicaciones que fueron desarrolladas en su momento con diferentes requisitos y sin pensar en una futura integración de los datos, por lo que los datos existen en diferentes sistemas, tal vez con diferentes nombres o con

diferentes formas de expresarlos, como por ejemplo el sexo de una persona puede ser representado como x/y en un sistema, mientras que en otro como 1/0 o M/F, siendo este un campo básico en los sistemas, requiere estandarizarse, siendo un proceso complejo tomando en cuenta todos los campos y sistemas que debe consultarse. Cabe recalcar que el proceso propuesto por el autor consta de diferentes iteraciones en los que se desarrollan aplicaciones guiándose de un modelo de datos lo que permite mantener los diferentes esfuerzos de desarrollo alineado con el objetivo del data warehouse.

2.3.2 Ciclo de vida de Ralph Kimball

En (Kimball & Ross, The Data Warehouse Toolkit 2ed., 2002) se hace referencia al proceso de implementación de un Data Warehouse partiendo de unidades pequeñas conocidas como Datamart, que son implementadas individualmente por cada área de acuerdo a las necesidades encontradas para al final de la construcción de cada Datamart en distintas áreas de la empresa encontrar un data warehouse.

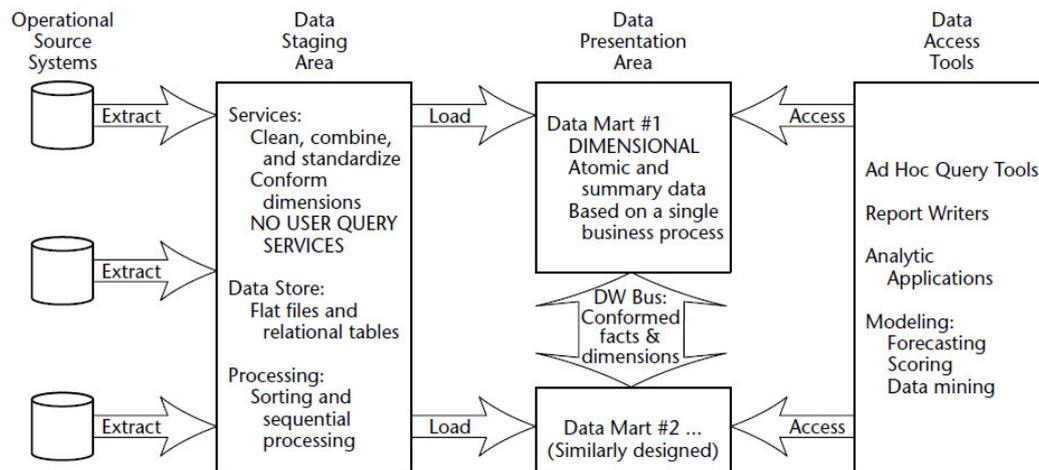


Ilustración 10 Elementos básicos del data warehouse.

Fuente: (Kimball & Ross, The Data Warehouse Toolkit 2ed., 2002)

Según (Kimball & Ross, The Data Warehouse Toolkit 2ed., 2002), se definen tres componentes básicos de un data warehouse, como vemos en la Ilustración 10, entre ellos encontramos los Sistemas Fuente Operacionales, Data Staging Area o ETL, Área de Presentación de datos y Herramientas de acceso de datos. La fuente de Sistemas Operacionales son los sistemas transaccionales que existen actualmente en una empresa, que se encargan de ayudar a sus usuarios en sus operaciones diarias. El elemento de data Staging es también conocido como ETL, durante este

proceso se tiene como finalidad extraer los datos desde los distintos sistemas transaccionales, formalizarlos (limpiarlos) con la finalidad de que los datos de diferentes sistemas tengan los mismos formatos y signifiquen lo mismo y finalmente, cargarlos al modelo dimensional. El tercer elemento, el área de presentación es donde encontramos los modelos dimensionales. Finalmente, el elemento de Herramientas de acceso a datos, incluye todas las herramientas que usan los usuarios finales para acceder a los datos, estas herramientas no solo se limitan a facilitar las consultas de datos a los modelos dimensionales, sino también pueden servir de retroalimentación para los sistemas transaccionales o alguno de los otros elementos del datamart.

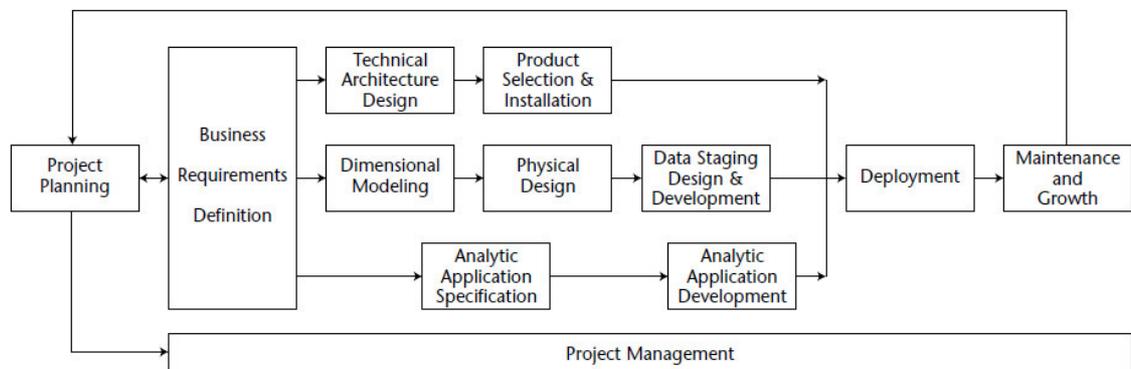


Ilustración 11 Diagrama del ciclo de vida del negocio dimensional.

Fuente: (Kimball & Ross, The Data Warehouse Toolkit 2ed., 2002)

A continuación, se mostrará de manera breve el ciclo de vida mostrado en la Ilustración 11 de acuerdo con lo revisado en (Kimball & Ross, The Data Warehouse Toolkit 2ed., 2002).

1. Planeamiento y Gestión del Proyecto

Durante esta etapa del desarrollo del data warehouse según (Kimball & Ross, The Data Warehouse Toolkit 2ed., 2002), se realiza el planeamiento del proyecto que se realizará, para esto brinda un conjunto de recomendaciones que apoyarán a mejorar las posibilidades de éxito del proyecto, por ejemplo, deben existir interesados dentro de la empresa comprometidos que comprendan los beneficios de la construcción del data warehouse, se debe definir un alcance real, además, se debe justificar debidamente la construcción de un data warehouse, especialmente, económicamente. Se debe entender los beneficios de un data warehouse y

considerar la posible expansión del mismo. Finalmente, dentro de un proyecto de desarrollo de un data warehouse debe existir un equipo multidisciplinario que lleve adelante el data warehouse.

2. Definición de Requerimientos

Durante esta etapa, (Kimball & Ross, The Data Warehouse Toolkit 2ed., 2002) recomienda identificar los usuarios que serán claves en el desarrollo del data warehouse e identificar al equipo encargado de realizar la especificación de los requerimientos. Durante este proceso, se recomienda realizar reuniones con los administradores de datos (DBA) junto con los usuarios de negocio de tal forma que se pueda evaluar la posibilidad de que el data warehouse sea realizado con los datos existentes. Una vez que los requerimientos han sido captados, se realiza la priorización de los mismos para poder seleccionar con cuál de los procesos de negocio comenzar, se recomienda escoger es el que tenga altas posibilidades de ser realizado y un alto impacto en el negocio de acuerdo con la Ilustración 12.

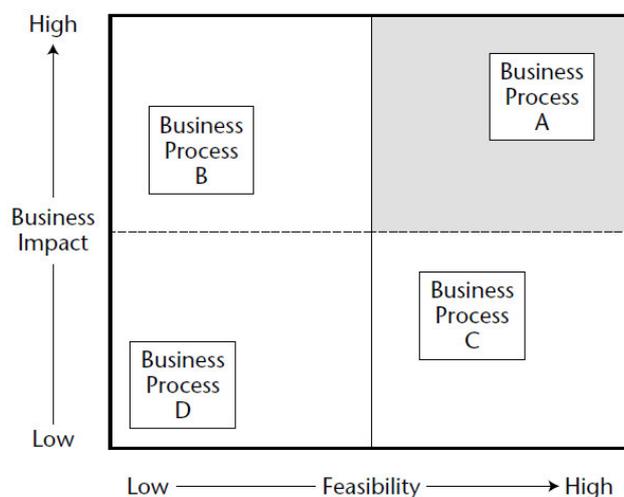


Ilustración 12 Cuadrante de priorización.

Fuente: (Kimball & Ross, The Data Warehouse Toolkit 2ed., 2002)

3. Definición de Tecnologías

En esta etapa dentro de la metodología planteada por (Kimball & Ross, The Data Warehouse Toolkit 2ed., 2002), se resaltan dos etapas, primero en la fase de diseño de arquitectura, se define la arquitectura que hará posible la implementación del data warehouse. En el siguiente punto, se realiza la selección de la herramienta que está acorde con la arquitectura planteada.

4. Definición de la Información

De acuerdo con Kimball, una vez que se tienen claros los requerimientos de negocio, se procede a realizar el modelo dimensional que proveerá la información solicitada, el modelo físico y se define el proceso ETL. Es en esta fase que se desarrolla fuertemente el proceso ETL, comenzando con el plan para la extracción, definiendo las transformaciones (limpieza de datos, corrección de data) necesarias para que la data pueda ser registrado en los modelos definidos anteriormente, finalmente se realiza la carga de la data en los modelos dimensionales y completando la información que se guardará las tablas de hechos.

5. Definición de las Aplicaciones Analíticas

En esta sección se cubre el desarrollo de aplicaciones que se alimentarán de los modelos definidos anteriormente, del hardware y software, de la inversión realizada en el data warehouse, para mostrar información que apoye a las diferentes áreas de la empresa.

2.3.3 Metodología para la construcción de un Data Warehouse (HEFESTO)

La metodología HEFESTO, (Bernabeu, 2010), es un proyecto comunitario liderado por el Ing. Bernabeu Ricardo Dario, actualmente se encuentra en la versión 2 publicada en Córdoba, Argentina en Julio de 2010. La idea principal de esta metodología, según lo expresa su autor, es “comprender cada paso que se realizará, para no caer en el tedio de tener que seguir un método al pie de la letra sin saber exactamente qué se está haciendo ni por qué”. El autor explica en la documentación de la metodología que la metodología HEFESTO puede ser embebida en cualquier ciclo de vida de desarrollo de software, siempre y cuando sea adaptada para trabajar con la metodología seleccionada. Además resalta la importancia de tener fases de desarrollo cortos que puedan demostrar las ventajas de usar un data warehouse y de esta forma motivar a los usuarios.

La metodología se basa en las fases del proceso de BI, para esto el autor propone un conjunto de pasos y actividades, comenzando con el (1) Análisis de requerimientos, en este paso se realiza la identificación de preguntas, indicadores y perspectivas que se implementarán, además se realiza el primer modelo conceptual del datamart. En el paso (2) Análisis de los OLTP se definen los indicadores, establecen correspondencias, define el nivel de granularidad y se obtiene un modelo

conceptual ampliado. A continuación, el autor sugiere la creación del (3) modelo lógico del DW en la que se selecciona el modelo lógico que soportará los requisitos de los usuarios y se crearán las tablas de dimensiones y hechos. Finalmente se realizará la (4) Integración de datos, realizando la carga inicial de datos y se establece el proceso de actualización de datos. En la Ilustración 13 se muestra una imagen donde con los pasos propuestos por esta metodología y seguidamente detallamos cada uno de ellos.



Ilustración 13 Metodología HEFESTO, pasos.

Fuente: (Bernabeu, 2010)

1. Análisis de Requerimientos

El primer paso propuesto por el autor es el análisis de requerimientos, este paso se conforma de tres actividades principales, (a) la identificación de preguntas, (b) la identificación de indicadores y perspectivas y (c) un modelo conceptual. A continuación, detallaremos brevemente cada una de las actividades a seguir en este paso.

La primera actividad, identificación de preguntas, busca captar las necesidades de información en la organización de acuerdo con sus metas y estrategias haciendo uso de diferentes técnicas como encuestas,

cuestionarios, observaciones, etc. El objetivo de esta actividad es identificar las necesidades de información en un alto nivel.

Una vez descubiertas las necesidades de información de la organización, se procede a su análisis para obtener los indicadores y las perspectivas que se tomarán en cuenta durante la implementación del datamart. Los indicadores son, generalmente, valores numéricos que se desean analizar y las perspectivas son los puntos de vista sobre los cuales se examinan los indicadores para responder las preguntas analizadas durante la primera actividad.

Finalmente, el modelo conceptual permitirá reunir la información hasta el momento obtenida y presenta un primer bosquejo de alto nivel del alcance del proyecto. Este modelo se representa a través de un gráfico que consta de tres partes, la primera son las perspectivas sobre las cuales se realizará el análisis, en la posición central se grafica la relación que tienen todas estas perspectivas y finalmente como resultado de esta relación se miden los indicadores resultantes como se aprecia en la Ilustración 14.

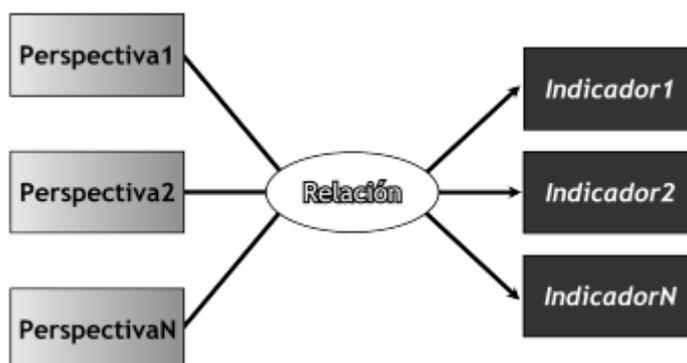


Ilustración 14 Ejemplo de modelo conceptual.

Fuente: (Bernabeu, 2010)

2. Análisis de los OLTP

En este paso, se realiza el análisis de las fuentes de datos de las cuales se proveerá de datos al modelo planteado en el punto anterior. Para esto se llevan a cabo las siguientes actividades: (a) Conformar indicadores, (b) establecer correspondencias, (c) definir el nivel de granularidad y finalmente la (d) creación del modelo conceptual ampliado.

Para comenzar, en la actividad conformar indicadores, se define las funciones de cálculo para los indicadores. Los indicadores están compuestos de hechos y una función de sumarización para agregar los resultados.

Seguido, se realiza la correspondencia de los datos requeridos por el datamart y con los datos que existen en el OLTP, esto facilitará su próxima extracción y asegura que se cuente con los datos requeridos para la construcción del datamart.

A continuación, se definirán los campos que tendrán cada perspectiva y el nivel de detalle (granularidad) con el que los datos serán importados al datamart.

Finalmente obtenemos un modelo conceptual ampliado que incluye los campos de las perspectivas y las fórmulas de cálculo de cada indicador definido.

3. Modelo lógico del DW

En este paso se realiza el pase del modelo conceptual a un modelo lógico que definirá la estructura del data warehouse. Para esto las actividades desarrolladas durante este paso son: (a) Selección del tipo del modelo lógico del data warehouse, (b) definición de las tablas dimensionales, (c) las tablas de hechos y las (d) relaciones entre ambas.

En la actividad Selección el tipo de modelo lógico del data warehouse el equipo encargado del desarrollo seleccionará objetivamente el esquema a utilizar (entre los esquemas estrella, constelación o copo de nieve).

A continuación, se definen las tablas de dimensiones y hechos que serán usados en el datamart a partir de los modelos conceptuales definidos en pasos anteriores, así como las relaciones existentes entre estas tablas formando el modelo de la base de datos basado en el esquema escogido para la estructura del data warehouse.

4. Integración de datos

Durante este paso se ven dos actividades, (a) la carga inicial y la (b) actualización. En estas actividades se desarrolla el proceso ETL que se encargará de alimentar el data warehouse con los datos necesarios para su uso. Para esto, en la primera actividad, carga inicial, se prueba el modelo de datos diseñado en pasos anteriores asegurándonos que los datos cargados no

estén incompletos o anómalos, además que contengan las restricciones necesarias para responder las preguntas definidas en la etapa de requerimientos. El paso de actualización, consistirá en definir la política y estrategias de refresco de datos que contemple lo aprendido en la etapa de carga inicial y que tenga las mismas consideraciones.

Los pasos mencionados anteriormente definen la estructura del data warehouse que se usará, sin embargo, se requiere que se defina la manera de explotación de este datamart, para esto se crean cubos dimensionales, indicadores, atributos, y jerarquías, que permitirán la explotación de la información y que usualmente está atado a una herramienta de análisis.

2.3.4 Metodología DWEP (Data Warehouse Engineering Process)

Esta metodología, según (Luján Mora & Trujillo, 2004), está basada en la metodología de Unified Process (UP) y hace uso de los artefactos propuestos por UML, propuesto por Sergio Lujan-Mora y Juan Trujillo. Esta metodología hace uso del ciclo de vida propuesto por UP (Incepción, Elaboración, Construcción y Transición) además de sus flujos de trabajo (Requerimientos, Análisis, Diseño, Implementación y Pruebas) y agrega dos flujos de trabajo, Mantenimiento y Revisión post-desarrollo.

1. Requerimientos

Durante esta fase, se obtienen los indicadores, dimensiones de análisis, la frecuencia de actualización, etc. En esta fase se hace uso de los diagramas de caso de uso para modelar los requerimientos de usuario, además con su especificación respectiva.

2. Análisis

En esta fase se busca refinar y estructurar los requerimientos de usuario, definir las fuentes de datos y de ser necesario documentarlas con la finalidad de tener claras las fuentes de datos. Resultando en un Esquema de Fuentes Conceptual (Lógico y Físico) que modela las diferentes fuentes que alimentaran el data warehouse.

3. Diseño

Durante esta fase se diseña el data warehouse conceptualmente y el mapeo que será usado para generar la data que se guardará en el data warehouse

desde los sistemas fuente. Para esto se desarrollan los siguientes artefactos: Data Warehouse Conceptual Schema (DWCS), Client Conceptual Schema (CCS), y Data Mapping (DM), siendo este último el encargado de identificar las relaciones entre las fuentes de datos y el data warehouse.

4. Implementación

Durante este proceso se realiza la construcción del data warehouse, se realiza la carga inicial de datos y se optimiza el performance del data warehouse. En este proceso se obtienen los siguientes artefactos: Data Warehouse Logical (Physical) Schema, Cliente Logical (Physical) Schema, ETL Process, Exportation Process y Transportation Diagram.

5. Pruebas

Se realizan las pruebas del data warehouse creado para verificar que el diseño sea el correcto, en esta fase no se crean nuevos artefactos, pero si se pueden modificar los ya existentes debido a las correcciones realizadas.

6. Mantenimiento

Esta fase busca mantener el data warehouse actualizado con los nuevos datos generados en las fuentes de data. Además se captan nuevos requerimientos y realizan nuevos desarrollos.

7. Revisión post-desarrollo

Esta fase busca la obtención de buenas prácticas y lecciones aprendidas, así como la adecuación del proceso de desarrollo para mejorar el mismo.

CAPÍTULO III. ESTADO DEL ARTE

En esta parte se desarrollará el estado del arte del problema que se intenta resolver. Este apartado consta de dos partes, la primera describe las herramientas que actualmente existen para apoyar a la solución del problema; la siguiente sección describe algunas soluciones encontradas que aplican la inteligencia de negocios en el mercado y nombra algunas, brindando una breve descripción de las mismas. Finalmente se muestra unos cuadros comparativos que se usaron para realizar la selección de la metodología a aplicar en este proyecto.

3.1 Herramientas para la construcción de datamart

Actualmente, nos encontramos con diferentes soluciones en el mercado que permiten la implementación de un datamart en una empresa. A continuación, veremos algunas soluciones encontradas, que en este trabajo han sido consideradas relevantes como alternativas de uso para la implementación del proyecto.

3.1.1 Oracle Business Analytics

Como sabemos, los sistemas de inteligencia de negocios se basan en los sistemas de bases de datos para realizar las operaciones de minería de datos y otros. En este sentido, empresas que brindan un sistema de base de datos, implementan módulos de inteligencia de negocios. Oracle Business Analytics es una solución que ayuda a las organizaciones de todos los tamaños permitiéndoles encontrar nuevas formas de realizar estrategias, planear y optimizar operaciones de negocio. La información de este producto se encuentra en su sitio oficial (Oracle, 2016)

3.1.2 Pentaho

De acuerdo con el desarrollador (Pentaho, 2016), Pentaho es una plataforma de integración de datos y análisis de negocio que centraliza el acceso, integración, visualización y explotación de datos. Pentaho provee una solución empresarial que consiste en guiar a los usuarios en la implementación de la solución. Por otro lado, Pentaho es una solución construida por una comunidad de código libre y tiene los siguientes proyectos puestos a disponibilidad de los usuarios que deseen usar o desarrollar para sus propios negocios:

1. Pentaho BI Platform and Server

La plataforma provee la arquitectura necesaria para construir soluciones de inteligencia de negocios. Es un framework modular que permite la

integración de distintas herramientas para el desarrollo de una solución de Inteligencia de Negocios.

2. Pentaho Reporting

Esta es una herramienta que proporciona la habilidad de realizar reportes de diferentes fuentes como PDF, Excel, HTML, Texto, RTF, XML y CSV.

3. Pentaho Data Integration (Kettle)

Este proyecto se enfoca en la creación de una herramienta ETL (Extracción, Transformación y Carga) basada en estándares y dirigido por meta data.

4. Pentaho Analysis Services (Mondrian)

Mondrian es un servidor OLAP escrito en java que permite analizar datos en bases de datos SQL sin necesidad de escribir sentencias SQL.

5. Pentaho Data Mining (Weka)

Weka es un conjunto de herramientas para la minería de datos usados para el análisis predictivo.

3.1.3 Microsoft SQL Integration Services y Analysis Services

Microsoft SQL Integration Services (Microsoft, 2016) es una plataforma de extracción, transformación y carga de datos que provee la habilidad de realizar operaciones ETL desde cualquier fuente de Información ya sean de Microsoft o no.

Características

- Fuentes de datos: SQL Server, Oracle, Teradata, DB2, ODBC, OLE DB, Archivos de texto, y Microsoft Entity Framework.
- Sincronización de datos en la nube con Data Sync.
- Soporta grandes cantidades de información
- Integración con software de desarrollo como visual estudio.

Microsoft SQL Analysis Services (Microsoft, 2016) permite a sus usuarios el desarrollo de soluciones para el análisis predictivo y explotación de datos. Como podemos ver en la Ilustración 15, la arquitectura planteada por Microsoft de una solución BI, describe que a partir de los datos obtenidos de los sistemas transaccionales, se construye el datamart/data warehouse, de los cuales se obtienen cubos. Finalmente se realiza la presentación de los datos por distintos medios, e incluso mediante web.

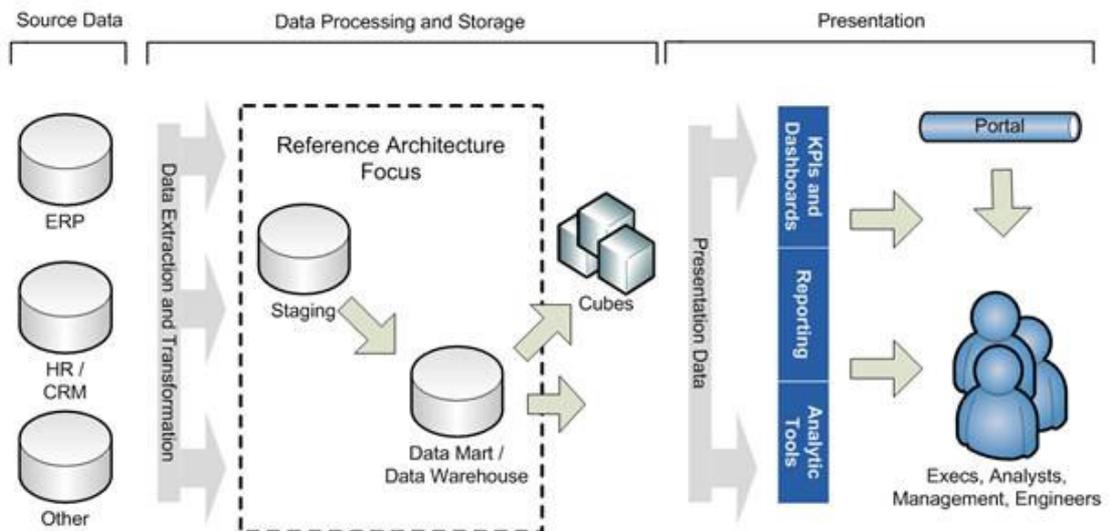


Ilustración 15 Arquitectura utilizada por una solución en MSSQL Server.

Fuente: (Microsoft, 2016)

3.1.4 MicroStrategy

MicroStrategy (Microstrategy, 2010) es una empresa dedicada a la comercialización de su solución de inteligencia de negocios que se centra en el uso de análisis avanzado y predictivo, escalable e incluso en tiempo real. En la industria de la educación, MicroStrategy provee las siguientes características.

- Dashboards de matricula
- Gestión del ciclo de vida del estudiante. Esta funcionalidad permite a la entidad acceder a la información biográfica, financiera y académica de cada uno de los estudiantes individualmente, apoyando el proceso de promoción de los servicios (como atraer al estudiante).
- Contabilidad educativa. Esta característica permite al usuario el acceso rápido de datos estadísticos de estudiantes, docente e investigadores.

La arquitectura planeada por MicroStrategy, a diferencia de la MSSQL Server, determina que existe una interfaz entre los datos fuente y la capa de prestación, el servidor de inteligencia de negocios de MycroStrategy, esta solución se encarga de integrar los datos y ponerlos a disposición de los diferentes tipos de presentación, para usuarios diferentes, ya que brinda una arquitectura orientada a servicios.

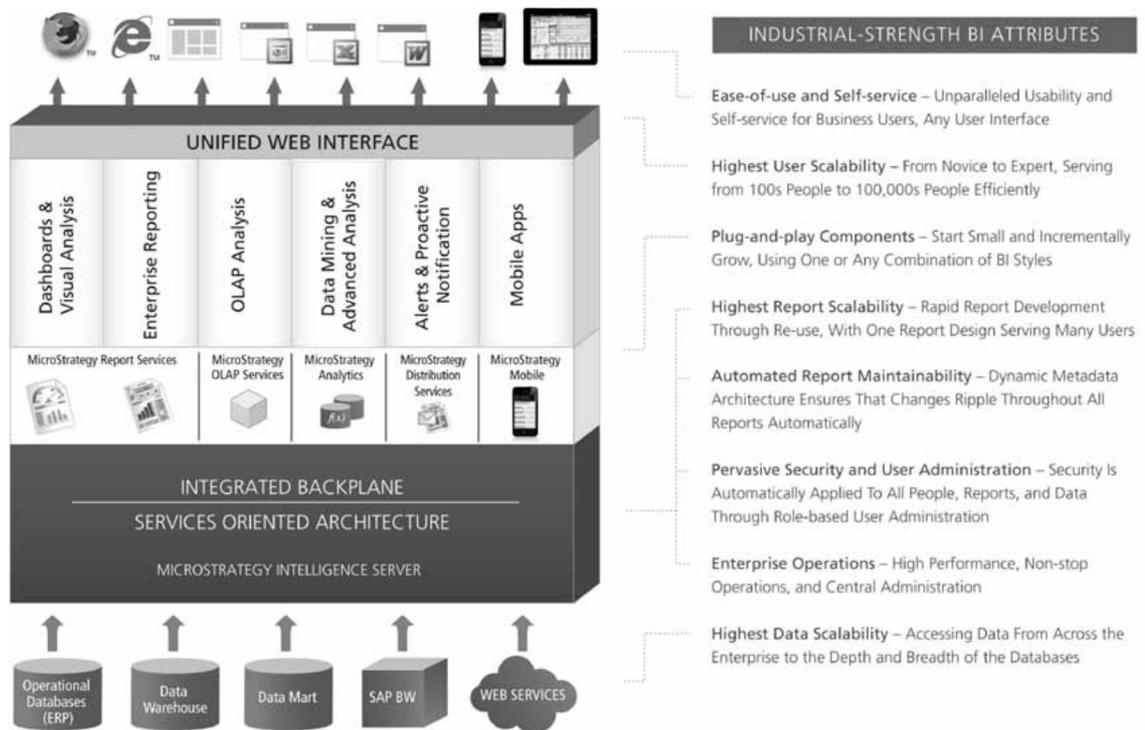


Ilustración 16 Arquitectura planteada por MicroStrategy.

Fuente: (Microstrategy, 2010)

3.1.5 Jaspersoft

Jaspersoft (Jaspersoft, 2016), es una compañía que ha desarrollado una solución de inteligencia de negocios que, de manera análoga a MicroStrategy, se ha enfocado en la educación, dentro de los puntos que propone Jaspersoft podemos resaltar.

- La solución provee reportes de logro de los estudiantes, lo que permite a los docentes realizar el análisis y la toma de decisiones fácilmente.
- Integración con sistemas web que faciliten la integración de los estudiantes al uso de la herramienta.

La empresa (Jaspersoft, 2016), provee una arquitectura parecida a la proporcionada por MicroStrategy (basada en servicios) pero con un número menor de opciones, sin embargo, implementa un módulo de extensiones para que el usuario pueda definir su aplicación front-end de acuerdo a su preferencia como se puede ver en la Ilustración 17.

JasperServer Architecture

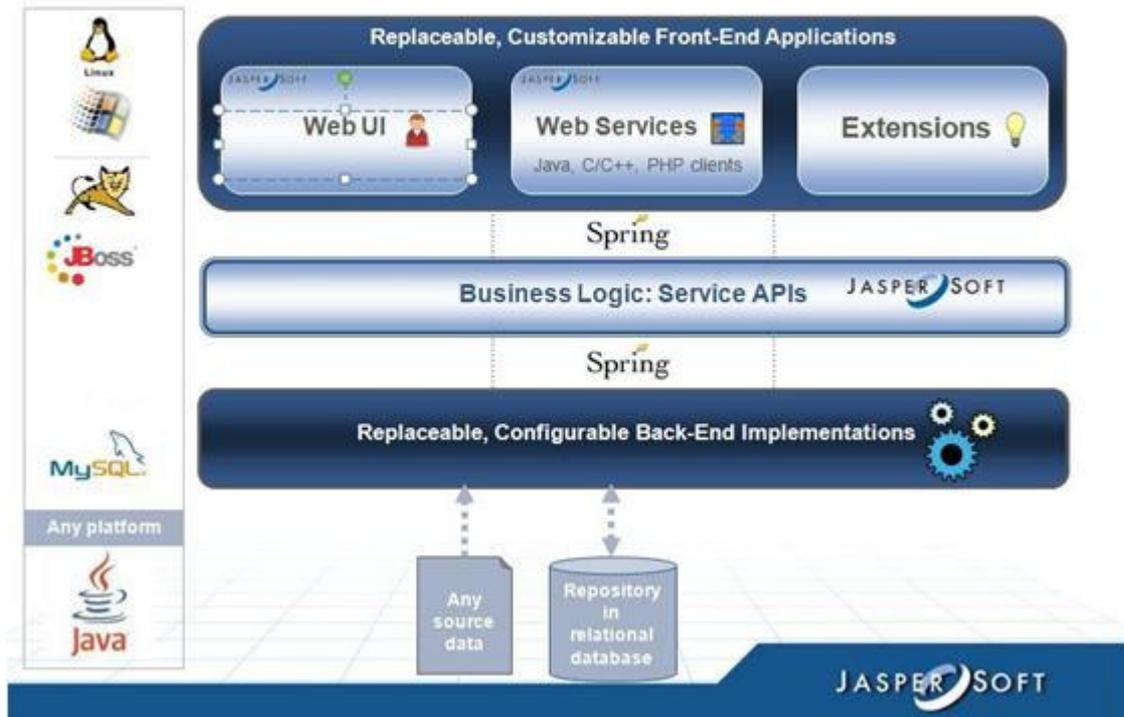


Ilustración 17 Arquitectura propuesta por jaspersoft.

Fuente: (Jaspersoft, 2016)

3.1.6 QlikView

“La plataforma de QlikView provee un verdadero auto servicio de inteligencia de negocios que posibilita a los usuarios la toma de decisiones innovadora” – Fuente: (Qlik, 2016)

QlikView proporciona una arquitectura aún más encapsulada que las anteriores, QlikView, define un data cloud, que hace de intermediario entre la interfaz gráfica de usuario y la fuente de datos.



Ilustración 18 Arquitectura QlikView.

Fuente: (Qlik, 2016)

3.2 Trabajos realizados

En la literatura he encontrado diferentes trabajos relacionados con los sistemas de apoyo a la toma de decisiones. Entre los diferentes trabajos encontrados, resalto los siguientes:

3.2.1 Construcción de una herramienta de desarrollo de soluciones para inteligencia de negocios – explotación (Falcón Cisneros, Palma Stanciuc, & Celi Peña, 2007)

Como se ha señalado en puntos anteriores, existen diferentes soluciones de diversas empresas para implementar Inteligencia de Negocios. (Falcón Cisneros, Palma Stanciuc, & Celi Peña, 2007) señalan que muchas veces, las implementaciones de estas soluciones son costosas y se relacionan con una sola base de datos, que también es un producto del distribuidor del software.

Por esta razón, los autores proponen una solución para la explotación de datos que puede ser usado, en teoría, con cualquier manejador de bases de datos. En la Tabla 2, se demuestra que gracias a la modularidad de la aplicación diseñada con un conector de base de datos genérico y el uso del servidor OLAP mondrian, del proyecto Pentaho, se logró desarrollar una aplicación de inteligencia de negocios que permite conectarse a diferentes motores de bases de datos como MS SQL server, Oracle, MySQL y MS Access en distintas versiones. El software fue desarrollado en java web con servidor apache Tomcat.

Tabla 2 Motores de bases de datos soportados por la solución.

Fuente: (Falcón Cisneros, Palma Stanciuc, & Celi Peña, 2007)

Motor de base de datos	Versión
Microsoft SQL Server	7.0
Oracle	8
MySQL	5.0
Microsoft Access	11.66 SP2

3.2.2 Solución de inteligencia de negocios en el área de finanzas en la municipalidad de metropolitana de Lima

De acuerdo con la investigación realizada por (Núñez Soto, 2010), en el 2010, el área de finanzas de la municipalidad de Lima, sufría de un gran problema respecto a el procesamiento y análisis de información crítica para la toma de decisiones. En su trabajo, los autores, proponen una solución de inteligencia de negocios para resolver este problema, y de esta forma lograr sus objetivos. En este trabajo de investigación se hace uso de las herramientas de Microsoft SQL Server Integration Services y Server Analysis Services, elección realizada tomando en cuenta dos consideraciones: la herramienta cuenta con una gran cantidad de documentación accesible para realizar la implementación y en el ambiente de trabajo de la Municipalidad de Lima ya se había implementado proyectos de Datamart anteriormente y se tenía proyectado la implementación de un Data Warehouse con tecnologías Microsoft, por lo que la elección de Microsoft SQL Server era la más indicada.

3.2.3 Escuelas públicas de Tacoma

De acuerdo con uno de los casos de estudio reportados por Microsoft (Microsoft, 2011), las escuelas públicas de Tacoma optaron por una solución de inteligencia de negocios con los siguientes objetivos:

- Disminuir las diferencias entre los grupos étnicos y económicos
- Incrementar los porcentajes de graduados
- Disminuir los números de estudiantes que abandonan la educación

Entre las barreras más difíciles de romper, según reporta Microsoft, se encontraban la adaptación al uso de tecnologías de información y sobreponerse a la cultura tradicional del sistema de educación. La solución se implementó usando Microsoft

SharePoint® 2010, Microsoft SQL Server® 2008, SQL Server Integration Services (SSIS), SQL Server Reporting Services (SSRS), SQL Server Analysis Services (SSAS), Microsoft Excel®, PowerPivot, and PerformancePoint®; todas herramientas de Microsoft.

Beneficios

- Los directores y profesores tienen acceso a información crítica de rendimiento.
- Brinda probabilidades de integrarse con tecnologías en la nube.
- “Los directores pueden saber tan pronto como entran en su oficina que estudiantes están faltando a clases; que clases podrían no tener profesores, y que estudiantes se portan mal” – Taylor, un administrador de sistemas beneficiado con esta solución.

3.2.4 Soluciones de Inteligencia de negocios en la práctica: apoyo a la toma de decisiones en proyectos educativos para población infantil vulnerable en el caribe Colombiano

Este trabajo, (Galvis Lista, Gonzales Zabala, García Avendaño, & Benavides Fawcet, 2011), de la Universidad Industrial de Magdalena que presenta el proceso y los resultados obtenidos de la construcción de una solución de inteligencia de negocios para el apoyo a la toma de decisiones en proyectos educativos para población infantil vulnerable en el caribe colombiano. Este estudio hace uso de la metodología de Moss y Atre y programación extrema para el desarrollo de la solución. El problema al que se presenta la solución de inteligencia de negocios trata sobre el análisis de los resultados de los círculos de aprendizaje que realiza el gobierno colombiano como proyecto educativo en Colombia. Esta solución hace uso de Microsoft SQL Server 2008. El proyecto resalta el apoyo significativo a la toma de decisiones en los temas de asignación de tutores y la compra de útiles. Los objetivos, dificultades y beneficios de este trabajo de investigación se muestran en la Tabla 3.

Tabla 3 Cuadro resumen Objetivos, Dificultades y Beneficios del proyecto.

Fuente: (Galvis Lista, Gonzales Zabala, García Avendaño, & Benavides Fawcet, 2011)

Objetivos	<ul style="list-style-type: none"> • Gestión del proyecto y procesamiento de transacciones en línea • Generación de consultas y reportes
Dificultades	<ul style="list-style-type: none"> • Inconsistencias encontradas en los datos • Inexperiencia del equipo
Beneficios	<ul style="list-style-type: none"> • Fortalecimiento de del trabajo realizado por la universidad

En este artículo se realizó una comparación entre las distintas metodologías existentes que tomaremos como válida y se muestra en la Tabla 4.

Tabla 4 Metodologías para BI y sus componentes.

Fuente: (Galvis Lista, Gonzales Zabala, García Avendaño, & Benavides Fawcet, 2011)

Metodologías y Técnicas / características	Modelo de Ciclo de vida	Nivel de detalle	Alcance
Ciclo de vida de Ralph Kimball	Cascada	Alto	Data warehouse
Método de desarrollo de sistemas dinámicos para bodegas de datos	Iterativo	Bajo	Data warehouse
Metodología para la construcción de un Data Warehouse (HEFESTO)	Cascada	Bajo	Data warehouse
Metodología DWEP	Iterativo	Alto	Data warehouse
MBD	Ágil	Bajo	Data warehouse
SEMMA (Sample, Explore, Modify, Model and Assess)	Cascada	Medio	Minería de datos
CRISP-DM (Cross Industry Standard Process for Data Mining)	Cascada	Medio	Minería de datos
BI Road Map	Cascada	Bajo	Solución BI
BIEP	Iterativo	Alto	Solución BI
Metodología y modelo de ciclo de vida de una solución de inteligencia de negocios (Gartner)	Iterativo	Bajo	Solución BI
Pentaho BI	Ágil	Medio	Solución BI

CAPÍTULO IV. APOORTE TEÓRICO

En esta sección se define la metodología y herramientas a utilizar, además la arquitectura de la propuesta de este proyecto las herramientas necesarias y especificaciones técnicas.

4.1 Metodología de desarrollo

Con la finalidad de seleccionar la metodología de desarrollo para la implementación del datamart se analizaron las necesidades del proyecto y se encontró que se requiere una metodología que (1) permita trabajar a una sola persona con etapas bien definidas, además que (2) no requiera una documentación exhaustiva debido a que el periodo de tiempo designado para este proyecto es relativamente corto y finalmente que (3) permita una implementación de un data warehouse incremental con la finalidad que en futuros proyecto se pueda expandir el alcance y continuar con otras investigaciones. Teniendo en cuenta los aspectos mencionados anteriormente, se definen los siguientes criterios en la Tabla 5 sobre los cuales se evalúan las metodologías “Ciclo de vida de Ralph Kimball”, HEFESTO y DWEP haciendo uso un sistema de puntajes. En este caso se utilizará el valor 1 cuando determinada metodología cumple con un criterio y 0 en caso contrario.

Tabla 5 Comparación de metodologías.

Fuente: (Elaboración propia)

Criterio	Diseño del data warehouse por Bill Inmon	Ciclo de vida de Ralph Kimball	HEFESTO	Metodología DWEP
Considera una metodología de desarrollo por etapas definidas haciendo uso de un modelo en cascada.	Si	Sí	Sí	Sí
Requiere un nivel de detalle mediano – bajo.	No	Sí	Sí	No
Tiene un alcance mediano y puede ser usado para expandir el sistema.	No	Sí	Sí	Sí
Busca una implementación del data warehouse incremental.	Si	Sí	Sí	No
Puntuación total	2	4	4	2

De los resultados obtenidos en la Tabla 5, se escoge la metodología HEFESTO debido a que se adecúa mejor a las necesidades en tiempo y alcance de este proyecto. Basándonos en la metodología seleccionada, se define el proceso para la implementación del datamart como el conjunto de actividades mostradas en la Tabla 6, esta propuesta incluye el análisis de negocio realizado al inicio del proyecto para identificar los requisitos tanto funcionales como no funcionales, y además un paso más al finalizar la definición del datamart que representa su implementación con las herramientas seleccionadas.

Tabla 6 Procesos para la implementación del datamart basado en la metodología HEFESTO.

Fuente: (Elaboración propia)

Nro.	Proceso	Actividades, Entradas y salidas
1	Análisis del negocio	Actividades 1. Definir casos de uso del negocio
		Salidas - Diagrama de casos de uso del negocio - Diagramas de actividades - Requisitos del proyecto
2	Análisis de requerimientos	Entradas - Requisitos del proyecto
		Actividades 1. Identificar preguntas 2. Identificar indicadores y perspectivas 3. Modelo conceptual
		Salidas - Requisitos de información - Indicadores y perspectivas - Modelo conceptual
3	Análisis de los OLTP	Entradas - Modelo conceptual inicial y requisitos de información
		Actividades 1. Conformar indicadores 2. Establecer correspondencias 3. Nivel de granularidad 4. Modelo conceptual ampliado
		Salidas - Modelo conceptual

		<ul style="list-style-type: none"> - Indicadores definidos - Mapeo entre base de datos OLTP y el datamart
4	Modelo lógico del DW	Entradas <ul style="list-style-type: none"> - Modelo conceptual con perspectivas
		Actividades <ol style="list-style-type: none"> 1. Tipo de modelo lógico del datamart. 2. Tablas de dimensiones. 3. Tablas de hechos. 4. Relación entre tablas de hechos y dimensiones.
		Salidas <ul style="list-style-type: none"> - Modelo físico del datamart
5	Integración de datos	Entradas <ul style="list-style-type: none"> - Datos para la carga inicial - Modelo físico del datamart
		Actividades <ol style="list-style-type: none"> 1. Carga Inicial 2. Actualización
		Salidas <ul style="list-style-type: none"> - Scripts de consulta para ETL - Procedimientos de ejecución para actualización de datos
6	Creación de cubos multi dimensionales	Entradas <ul style="list-style-type: none"> - Modelo físico del datamart
		Actividades <ol style="list-style-type: none"> 1. Creación de cubos multidimensionales 2. Implementación de consultas
		Salidas <ul style="list-style-type: none"> - Consultas pre definidas para usuarios

4.2 Herramienta de procesamiento analítico

Para seleccionar la herramienta de procesamiento analítico se debe tomar en cuenta la situación actual en la que se encuentran las instituciones educativas públicas por lo que después de realizar un análisis de la situación se considera los siguientes aspectos relevantes (1) debe poder ser implementado en plataformas libres o basadas en GNU/Linux, (2) debe permitir hacer uso de las herramientas sin un costo adicional de una licencia, (3) no debe tener dependencias que requieran de una licencia que signifique un costo y finalmente, (4) debe proporcionar una suite de

herramientas que facilite el desarrollo del datamart (herramientas ETL, diseño). Estos aspectos están marcados mayoritariamente por querer implementar una herramienta que no signifique un costo adicional para las instituciones educativas.

Teniendo en cuenta los aspectos requeridos en este proyecto se elaboró un cuadro comparativo entre las herramientas disponibles mencionadas anteriormente, donde se evalúan un conjunto de criterios tal como se muestra en la Tabla 7, en este caso se utilizará el valor 1 cuando determinada herramienta cumple con un criterio y 0 en caso contrario.

Tabla 7 Comparación de herramientas.

Fuente: (Elaboración propia)

Criterio	Oracle Business Analytics	Pentaho	MSSQL Integration Services y Analysis Services	MicroStrategy	Jaspersoft	QlikView
Permite su uso multiplataforma (Windows/Linux)	Sí	Sí	No	Sí	Sí	No
Tiene versión comunitaria que puede ser usada sin la compra de una licencia	No	Sí	No	No	Sí	Sí
No tiene dependencias propietarias	No	Sí	No	Sí	Sí	Sí
Posee integración de proceso ETL y utilitarios de visualización de datos	Sí	Sí	Sí	Sí	Sí	Sí
Puntaje total	2	4	1	3	4	3

Como podemos observar, las herramientas de código abierto permiten proveer alternativas a sistemas de pago, como los MSSQL y Oracle BA, en este caso, tenemos dos alternativas a utilizar, Pentaho y Jaspersoft. Para este proyecto, se utilizará el Pentaho y su herramienta de procesamiento analítico Mondrian que será complementado con una herramienta de consulta Saiku BI (Meteorite BI, 2016).

4.3 Arquitectura de la solución

La solución propuesta trabaja con los datos de las calificaciones de estudiantes y monitoreo de docente, para esto se hará uso de una base de datos transaccional donde se reunirán ambos. Los datos de las calificaciones de estudiantes provenientes desde el sistema SIAGIE serán importados o registrados manualmente a esta base de datos y sobre esta última se realizarán los trabajos ETL.

En la Ilustración 19 mostramos las dos fuentes (datos del SIAGIE – importados y datos de monitoreos) que serán procesados mediante un proceso de integración de datos (ETL) y puestos a disposición por medio de cubos multidimensionales para que puedan ser usados por el sistema de seguimiento docente.

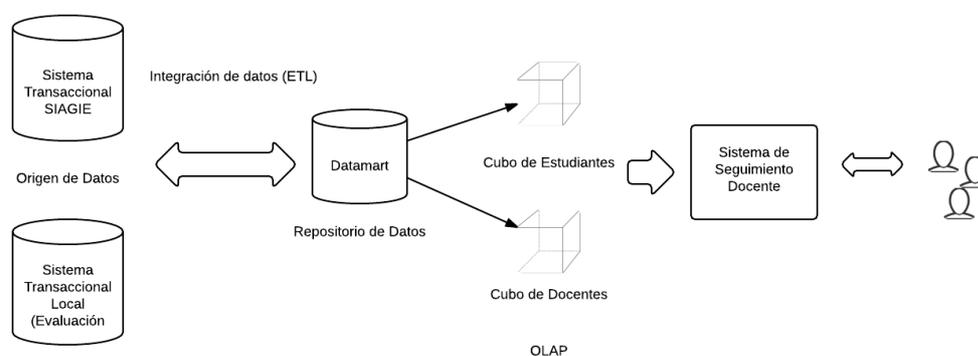


Ilustración 19 Arquitectura planteada para la implementación de la solución.

Fuente: (Elaboración propia)

De acuerdo con lo planteado en el punto anterior se seleccionó la suite de BI de Pentaho como herramienta de procesamiento analítico, en la Ilustración 20 se muestra cómo se utiliza esta suite dentro de las tecnologías usadas para el desarrollo de este proyecto. La solución planteada tendrá una base de datos en la que registrará los datos de la institución educativa, incluyendo los datos de monitoreo docente y de las calificaciones de los estudiantes. El proceso ETL, que definiremos en puntos más adelante, se encargará de proveer de datos al data mart. Para la explotación de los datos se hará uso del motor OLAP Mondrian, que forma parte de la suite de Pentaho y de una herramienta externa (Saiku Server) para hacer consultas a los cubos multidimensionales definidos en el data mart.

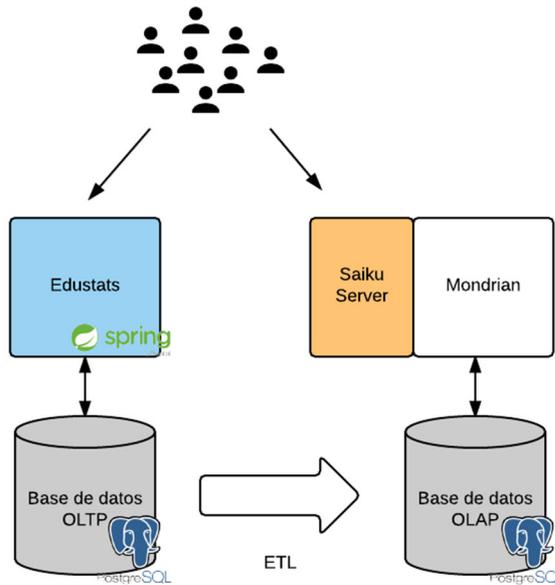


Ilustración 20 Topología de la propuesta.

Fuente: (Elaboración propia)

4.4 Herramientas tecnológicas

Para desarrollar este trabajo se hizo uso de distintas herramientas tecnológicas, por un lado, tenemos las herramientas que se utilizaron para la aplicación que sirve de interfaz con el usuario y por otro para la explotación de los datos, a continuación, sus especificaciones.

- Motor Gestor de Base de Datos: PostgreSQL 9.4
- Servidor de aplicaciones: Tomcat 7.0

Herramientas de inteligencia de negocios para poder facilitar la explotación de datos:

- Pentaho Data Integration: Se utiliza esta herramienta para automatizar el proceso ETL.
- Saiku BI y Pentaho Mondrian, el aplicativo Saiku, es una herramienta que facilita la integración del servidor OLAP Mondrian, brindando una API de consulta.

4.5 Especificaciones técnicas

- El sistema está optimizado para ser ejecutado en buscadores Firefox 45.0.1, Internet Explorer 9 y Chrome 49.0.26.

- El sistema debe ser alojado en un servidor de aplicaciones que soporte Java Servlets y con un motor gestor de base de datos PostgreSQL, facilitando la conectividad necesaria.

CAPÍTULO V. APOORTE PRÁCTICO

En este capítulo se expondrá el caso de estudio seleccionado para esta tesis, se realizará una evaluación de la organización seleccionada y la implementación del datamart. El caso de estudio es el nivel secundario de la Institución Educativa Santa Rosa del Sauce N° 170 de San Juan de Lurigancho.

5.1 Evaluación de la organización

En la siguiente sección se realiza una breve evaluación de la organización con la finalidad de conocerla mejor.

5.1.1 Contexto del Negocio

La Institución Educativa Santa Rosa del Sauce se desenvuelve en el ámbito educativo en el nivel inicial, primario y secundario. Ubicada en el distrito de San Juan de Lurigancho, esta institución educativa pública tiene como objetivo brindar la mejor formación a sus estudiantes.

5.1.2 Visión

“La I.E. N° 170 “Santa Rosa del Sauce” al 2015 es un centro de desarrollo humano; donde se incentiva, forma y potencia las habilidades y capacidades de sus estudiantes; capaces de integrarse a la sociedad y convivir en armonía consigo mismos, con los demás y su ambiente; aportando en la resolución de problemas de su comunidad; generando su propio trabajo y la de los demás, camino al desarrollo sostenible y sustentable de nuestro distrito y país.” – (Fuente: PEI SRS)

5.1.3 Misión

“Somos una Institución Educativa integral que contribuye a que los niños y jóvenes del AA.HH. Santa Rosa del Sauce desarrollen competencias básicas que le garanticen el cultivo del pensamiento crítico y creativo para la solución de problemas la toma de decisiones en su vida y la de su comunidad; basándonos en el desarrollo de una cultura de paz, ética, intelectual, artística-cultural, afectiva, física y espiritual.” – (Fuente: PEI SRS)

5.1.4 Factores Externos

5.1.4.1 La comunidad

La comunidad del AA.HH. es el factor externo más importante de la institución educativa ya que depende de esta el comportamiento de los estudiantes, sus costumbres y su desarrollo como estudiantes. La comunidad espera que la

institución educativa logre brindar calidad educativa y que sus estudiantes salgan adelante.

5.1.4.2 Entidades Externas

Dentro de estos factores tenemos dos factores importantes, las Unidades de Gestión Educativa (UGEL) y otras Instituciones Educativas públicas. Las UGEL, tienen como finalidad la regularización de las Instituciones Educativas, controlar que la normativa vigente se cumpla en las instituciones educativas y que no se encuentren irregularidades. Otras Instituciones Educativas públicas tienen un rol importante en la compartición de información y de experiencias en solución de problemas.

5.1.5 Factores Internos

5.1.5.1 Procesos

A continuación, veremos los dos procesos que tiene un rol importante en este proyecto.

5.1.5.1.1 Proceso de monitoreo

El proceso de monitoreo es un proceso de control pedagógico que tiene como finalidad mejorar los métodos que usan los docentes durante el desarrollo de clases. El proceso comienza con la definición de una asamblea de docentes en la cual se realiza la selección de los docentes quienes realizarán el monitoreo y además se asignarán los monitoreos que se deben realizar. Durante esta asamblea, también se elabora una ficha de monitoreo derivada de una plantilla proporcionada por la Dirección Regional de Educación (DRE). Una vez organizado los monitoreos, se procede a realizar la evaluación en el aula por parte del docente monitor asignado y además por el director de la institución educativa, de acuerdo con la asignación que se realizó en la asamblea. Al terminar cada monitoreo, el docente se compromete con ciertas opciones de mejoras que debe implementar. Seguidamente, se procede en una fase de Evaluación, donde los docentes recopilan los datos que tienen de los distintos monitoreos que realizaron y se prepara un documento de resultados de monitoreo. Finalmente se organiza una reunión en la que se evalúan los resultados y plantean mejorar a seguir, como se muestra en la Ilustración 21.

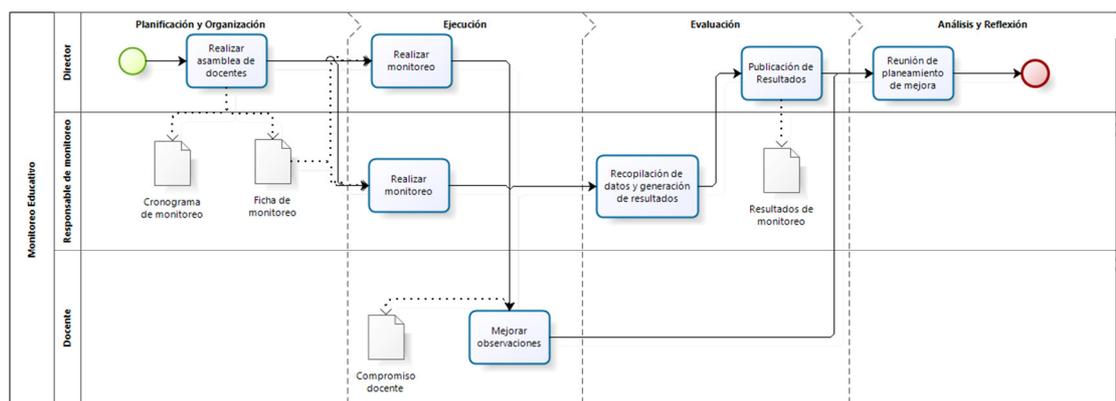


Ilustración 21 Monitoreo Educativo.

Fuente: (Elaboración propia)

5.1.5.1.2 Proceso Desarrollo de Año Escolar

Este proceso es realizado cada año y tiene la finalidad de realizar la organización y ejecución del año escolar. El proceso comienza con la planificación del siguiente año escolar, en esta planificación se define el número aproximado de estudiantes por grado y sección, el número de vacantes que se tendrá y la cantidad de docentes necesarios para cubrir la demanda, esto es conocido como metas de atención y son procesadas por la UGEL para que pueda gestionar el techo presupuestal de la institución educativa. Durante la matrícula, se realiza la inscripción de los estudiantes, tanto automática (estudiantes que han pasado de año), de nuevos estudiantes (inicial, primero de primaria) o traslado, además se realiza el planeamiento de la infraestructura necesaria para cubrir la demanda (asignación de salones de clase y mobiliario). Finalmente, el desarrollo de clases y todos los ajustes que se ven durante esta actividad, en la Ilustración 22 se muestra este flujo.

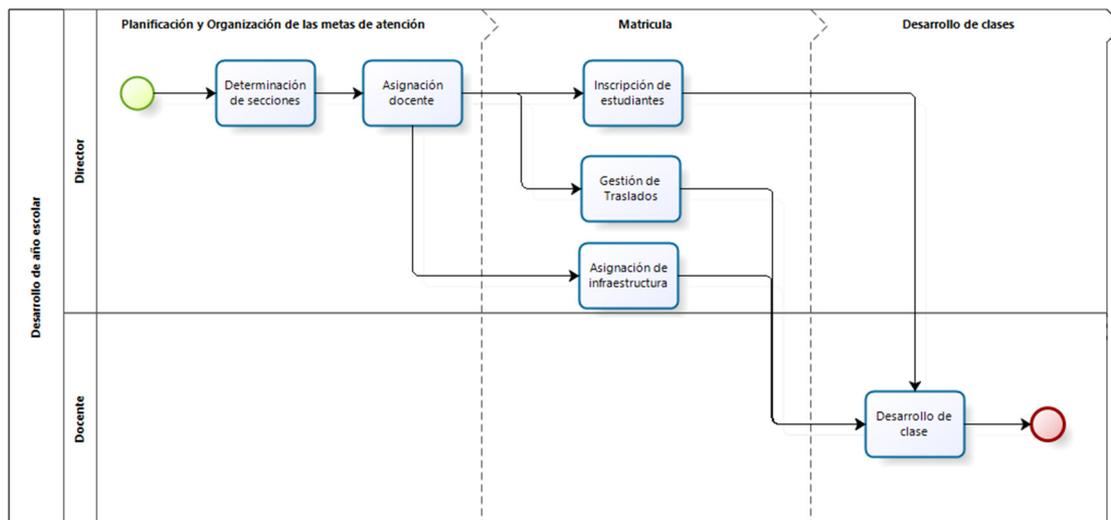


Ilustración 22 Desarrollo de año escolar.

Fuente: (Elaboración propia)

5.2 Análisis del negocio

Con la finalidad de explicar el contexto en el que se desarrolla este proyecto, en esta sección, realizamos el análisis del negocio y sus requisitos para implementar el datamart. Primero se muestran los casos de uso del negocio, seguido, se exponen los requisitos del proyecto.

5.2.1 Casos de uso del negocio

El proyecto se desarrolla en el contexto de las instituciones educativas públicas y la necesidad de información histórica de forma oportuna, para esto se analiza dos procesos dentro de la institución educativa, el proceso de evaluación de estudiantes y el de monitoreo de la práctica pedagógica. En la Tabla 8 se codifican los casos de uso de negocio identificados.

Tabla 8 Casos de uso del negocio identificados.

Fuente: (Elaboración propia)

Código	Nombre
CUN001	Monitoreo Docente
CUN002	Proceso de Evaluación
CUN003	Procesamiento de Resultados

A continuación, detallaremos brevemente cada uno de los casos de uso del negocio.

1. Monitoreo docente, es el proceso mediante el cual un comité de monitoreo se visita el desarrollo de una clase y evalúa el desempeño del docente haciendo uso de una ficha de monitoreo.
2. El proceso de evaluación tiene como objetivo evaluar el desempeño de los estudiantes mediante el uso de exámenes que evalúan diferentes criterios que podemos encontrar en la DCN. Las evaluaciones son agrupadas por periodos de evaluación que usualmente son divididos en los tres trimestres durante un periodo académico (un año).
3. El proceso de resultados es desarrollado periódicamente, usualmente antes de iniciar las clases, en las vacaciones de julio y en las reuniones de coordinaciones con los docentes, en el que se hace talleres y se reflexiona sobre el desempeño obtenido hasta el momento.

5.2.2 Diagrama de casos de uso del negocio

A continuación, se identifican los principales actores dentro de la organización y su relación con los casos de uso de negocio identificados en el punto anterior.

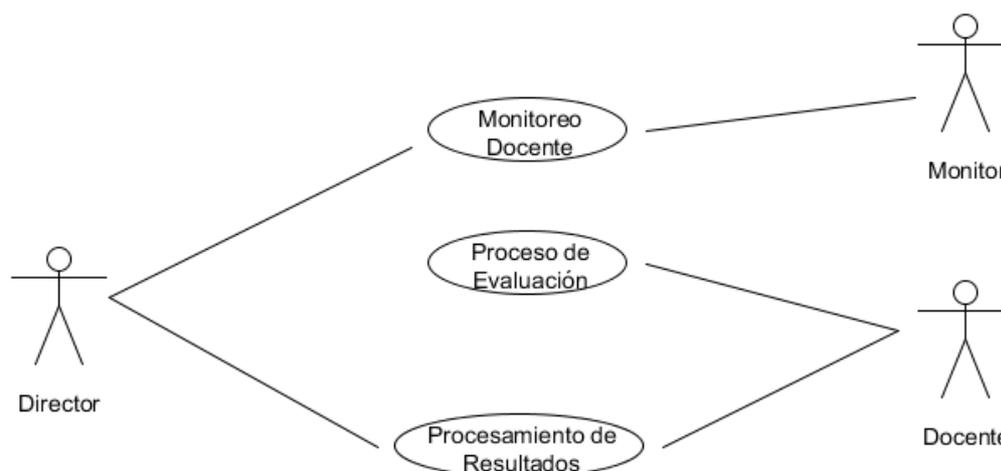


Ilustración 23 Diagrama de casos de uso del negocio.

Fuente: (Elaboración propia)

5.2.3 Modelo de análisis del negocio

Dentro de este proyecto identificamos tres involucrados principales, el director es el encargado de gestionar las actividades académicas, así como velar por la mejora del desempeño de la institución educativa, el docente tiene como función principal el dictado de clases y la evaluación de los estudiantes y finalmente el monitor se encarga de evaluar al docente. De los procesos mencionados del paso anterior encontramos tres entidades importantes en nuestro proyecto, (1) la ficha de monitoreo es llenada por el docente monitor y contiene los criterios de evaluación para el monitoreo, (2) la evaluación del estudiante contiene la calificación obtenida por el estudiante de acuerdo con los criterios de evaluación y los (3) resultados del monitoreo es el puntaje obtenido por un docente resultado del monitoreo, en la Ilustración 24 se muestran los trabajadores del negocio, las entidades y las realizaciones de caso de uso que se utilizarán en este proyecto para explicar las necesidades del negocio.

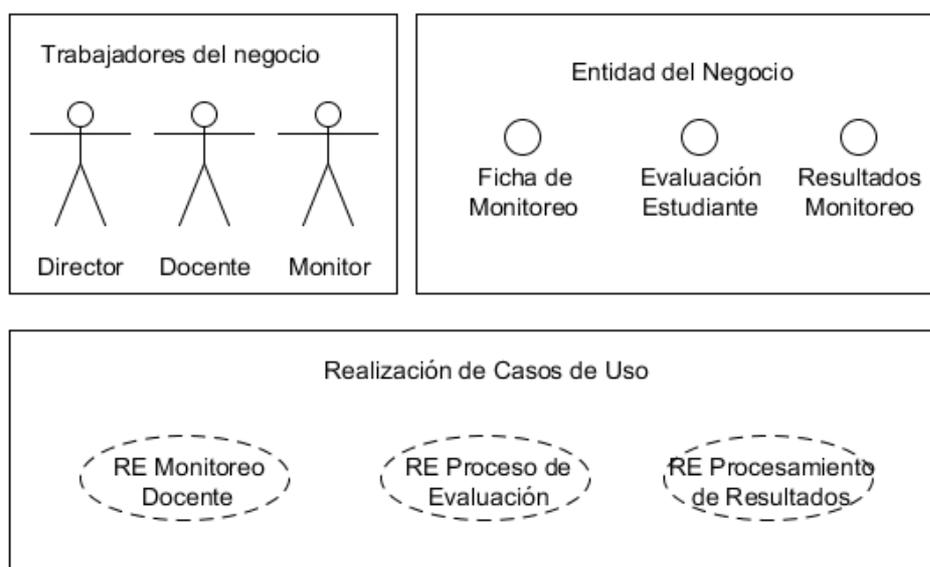


Ilustración 24 Modelo de análisis del negocio.

Fuente: (Elaboración propia)

5.2.4 Diagramas de actividades

A continuación, se presentan el diagrama de actividades de los tres casos de uso de negocio identificados.

En la Ilustración 25 se muestra las actividades que desarrolla cada actor del negocio y cómo se relacionan entre ellos para lograr realizar el monitoreo de la práctica

pedagógica en la institución educativa, además de las entidades generadas durante esta actividad.

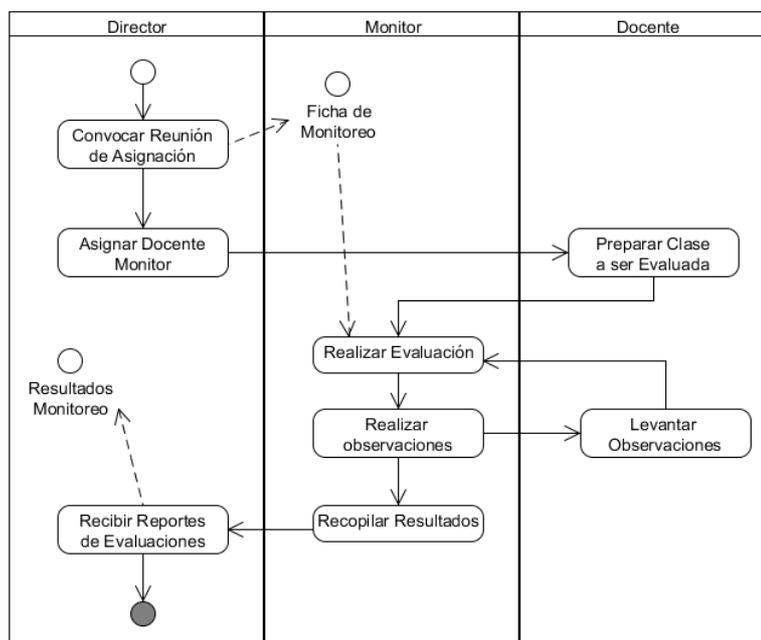


Ilustración 25 Diagrama de actividades de monitoreo docente.

Fuente: (Elaboración propia)

En la Ilustración 26, se muestra cuáles son las actividades que sigue el docente para realizar la evaluación a sus estudiantes y finalmente, en la Ilustración 27, se muestra el procedimiento para realizar una evaluación del desempeño de la institución educativa.

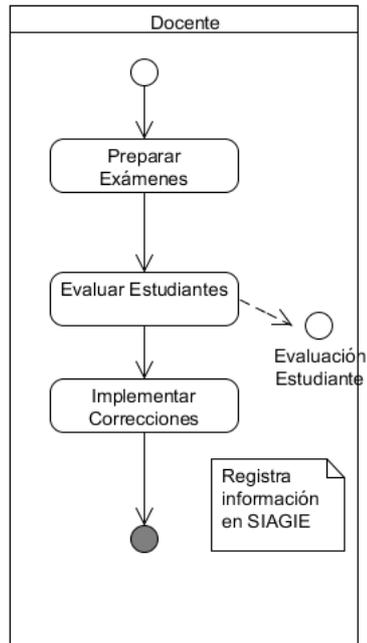


Ilustración 26 Diagrama de actividades de la evaluación a los estudiantes

Fuente: (Elaboración propia)

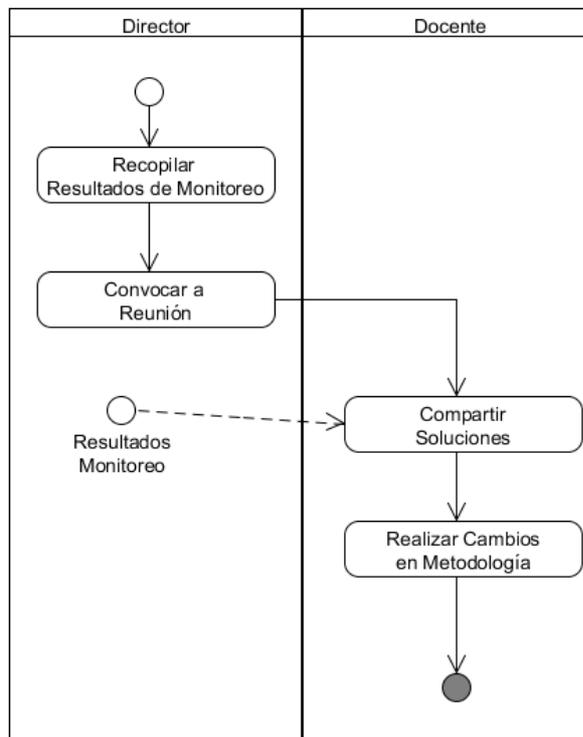


Ilustración 27 Diagrama de actividades de procesamiento de resultados.

Fuente: (Elaboración propia)

5.2.5 Requisitos del proyecto

Con el análisis de negocio realizado anteriormente, definimos los requisitos que el sistema debe soportar con la finalidad que la información pueda ser procesada en el datamart. En la Tabla 9 mostramos la lista de requisitos funcionales que debe soportar el sistema del que se extraerán los datos para ser procesados por el datamart.

Tabla 9 Requisitos funcionales.

Fuente: (Elaboración propia)

Código	Descripción
RF001	El sistema debe permitir el ingreso de los usuarios mediante un usuario y contraseña.
RF002	El sistema debe permitir la asignación del docente a determinado monitor.
RF003	El sistema debe permitir registrar los formularios de monitoreo.
RF004	El sistema debe permitir registrar los monitoreos realizados.
RF005	El sistema debe permitir procesar los resultados del monitoreo.
RF006	El sistema debe permitir la carga desde SIAGIE de datos de las calificaciones a estudiantes.
RF007	El sistema debe permitir el registro de los datos ingresados hacia los modelos de análisis.
RF008	El sistema debe permitir la obtención de reportes que le permita al usuario saber el estado de sus docentes.

En la Tabla 10 se enumeran los requisitos no funcionales que debe cubrir la aplicación para poder dar el servicio a las instituciones educativas. Finalmente se muestra la relación entre los usuarios del sistema y los casos de uso que ejecutan en la Ilustración 28.

Tabla 10 Requisitos no funcionales.

Fuente: (Elaboración propia)

Código	Descripción
RNF001	El sistema debe presentar una interfaz que muestre la información relevante para el usuario de manera organizada y cada una de las metáforas que se use en su interfaz debe ser única y comunicar la acción sin redundancia.
RNF002	El sistema debe encontrarse disponible durante las horas de oficina de las instituciones educativas (L-V de 8 a.m. a 7 p.m.) como mínimo.
RNF003	El tiempo de espera de respuesta del sistema debe ser no mayor de 5 segundos en caso de operaciones complejas.
RNF004	El sistema debe poder ser accedido desde buscadores Google Chrome v. 42 en adelante y Mozilla Firefox v 37 en adelante.

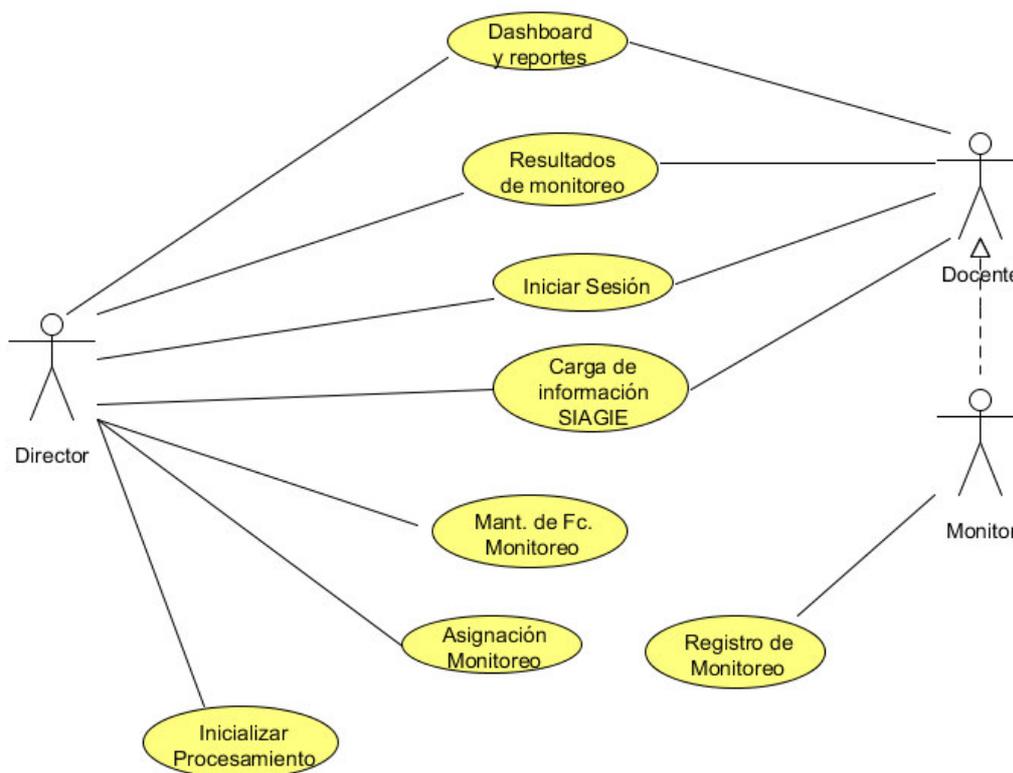


Ilustración 28 Diagrama de casos de uso.

Fuente: (Elaboración propia)

5.3 Diseño del datamart - Aplicación de la Metodología HEFESTO

En esta sección seguiremos los pasos de la metodología seleccionada en el punto 4.1, organizando los siguientes puntos según las actividades especificadas por la metodología HEFESTO.

5.3.1 Análisis de requerimientos

Para comenzar con la aplicación de la metodología definiremos un modelo conceptual que represente las necesidades de información de los usuarios, para lograr nuestro objetivo seguiremos tres pasos importantes, (1) identificaremos las preguntas que deben ser respondidas por el datamart planteado, (2) identificaremos los indicadores y perspectivas necesarios para responder las preguntas planteadas y (3) crearemos un modelo conceptual de alto nivel que será la base de nuestro proyecto. En las siguientes secciones se detalla cada una de las actividades mencionadas.

5.3.1.1 Identificar preguntas

Dentro de las conversaciones en la institución educativa, se encontró que se desea obtener información respecto a las calificaciones de los estudiantes y los monitoreos de la práctica pedagógica. Durante el análisis de las necesidades de la institución educativa, se encontró que el ministerio de educación solicita a los directores de las instituciones educativas envíen información sobre el desempeño durante el año académico, para enviar esta información se hace el procesamiento manual de las notas de los distintos grados/secciones y se completa un Excel (proveído por el ministerio de educación y disponible en la siguiente URL: www.minedu.gob.pe/campanias/pdf/matriz-de-monitoreo-de-los-cge.xls), en este documento se solicita la siguiente información, entre otros, número de docentes que cumplen con los indicadores que se obtienen a través del monitoreo como se muestra en Ilustración 29 y en la Ilustración 30.

RESULTADOS del Monitoreo de Inicio de año					
Número y porcentaje de docentes que cumplen con los indicadores ("nivel logrado")		Uso del tiempo en el aula	Uso de herramientas metodológicas	Materiales y recursos educativos	Total
Nivel INICIAL	Nro	1	0	0	0
	Porcentaje	50%	0%	0%	0%
Nivel PRIMARIA	Nro	0	0	0	0
	Porcentaje				
Nivel SECUNDARIA	Nro	0	0	0	0
	Porcentaje				
Toda la IE	Nro	1	0	0	0
	Porcentaje	50%	0%	0%	0%

Ilustración 29 Información solicitada por el ministerio de educación en la matriz de monitoreo - consolidado. Fuente: Ministerio de Educación

1. INICIO DE AÑO - Visitas realizadas en los meses de Marzo, Abril y Mayo

CONSOLIDADO	Nivel de enseñanza		INICIAL			CONSOLIDADO	Nivel de enseñanza		PRIMARIA		
	Nro de docentes en el nivel		0				Nro de docentes en el nivel		0		
	Nro de docentes monitoreados		2				Nro de docentes monitoreados		0		
	Aspecto observado		Desarrollo de la sesión				Aspecto observado		Desarrollo de la sesión		
	Uso del tiempo en el aula	Uso de herramientas metodológicas	Materiales y recursos educativos	Total		Uso del tiempo en el aula	Uso de herramientas metodológicas	Materiales y recursos educativos	Total		
	Pje. Máximo	12	24	12	10.0	Pje. Máximo	12	24	12		
	Pje. Promedio del nivel	10.0	0.0	0.0		Pje. Promedio del nivel					

Ilustración 30 Detalle de resultados de monitoreo solicitado por el ministerio de educación.

Fuente: Ministerio de educación

De igual forma, como parte de las metas para las instituciones educativas se solicita información relacionada a las calificaciones de los estudiantes como “El porcentaje de estudiantes, del nivel secundaria, que alcanza el nivel satisfactorio (18-20; 14-

17) en rendimiento en Ciencia, Tecnología y Ambiente”, entre otras, como se muestra en la Ilustración 31.

Compromiso de Gestión 8.

Indicador: **Porcentaje de actividades planificadas en el PAT que fueron implementadas.**

[Ir a Tabla de contenido](#)

Monitoreo del cumplimiento de metas 2015

¿Qué hacer?
 En muchos de los casos esta matriz se completará automáticamente a medida que completes los datos solicitados en las pestañas anteriores.
 En las columnas "Datos 2015" y "Meta 2015" deberá trasladar los datos desde el otro archivo "Matriz para la elaboración del PAT" sea manualmente o con la función "copiar y pegar" (debes seleccionar la opción de pegar solo los valores).

1	Progreso anual de todas y todos los estudiantes de la Institución Educativa.	Porcentaje de estudiantes, de los demás grados, que alcanzan nivel satisfactorio en rendimiento.	La institución educativa demuestra un incremento en el porcentaje de estudiantes que logran un nivel satisfactorio de aprendizajes en todos los grados, respecto al año anterior.	Primaria	Porcentaje de estudiantes, del nivel inicial, que alcanzan el nivel satisfactorio (A) en rendimiento en Personal Social.			0%	0%
					Porcentaje de estudiantes, del nivel primaria, que alcanzan el nivel satisfactorio (AD y A) en rendimiento en Comunicación.			0%	0%
					Porcentaje de estudiantes, del nivel primaria, que alcanzan el nivel satisfactorio (AD y A) en rendimiento en Matemática.			0%	0%
					Porcentaje de estudiantes, del nivel primaria, que alcanzan el nivel satisfactorio (AD y A) en rendimiento en Ciencia y Ambiente.			0%	0%
					Porcentaje de estudiantes, del nivel primaria, que alcanzan el nivel satisfactorio (AD y A) en rendimiento en Personal Social.			0%	0%
					Porcentaje de estudiantes, del nivel secundaria, que alcanzan el nivel satisfactorio (18-20; 14-17) en rendimiento en Comunicación.			0%	0%
					Porcentaje de estudiantes, del nivel secundaria, que alcanzan el nivel satisfactorio (18-20; 14-17) en rendimiento en Matemática.			0%	0%
					Porcentaje de estudiantes, del nivel secundaria, que alcanzan el nivel satisfactorio (18-20; 14-17) en rendimiento en Ciencia, Tecnología y Ambiente.			0%	0%
				Secundaria	Porcentaje de estudiantes, del nivel secundaria, que alcanzan el nivel satisfactorio (18-20; 14-17) en rendimiento en Historia, Geografía y Economía.			0%	0%

Ilustración 31 Datos solicitados por el ministerio de educación relacionada a las calificaciones de los estudiantes.

Fuente: Ministerio de educación

Con esta información, se procedió a realizar el análisis de la institución educativa y se encontró que respecto a la calificación de los estudiantes, se requiere clasificar las calificaciones dentro de los rangos siguientes rangos (1) 0-10, (2) 11-13, (3) 14-17 y (4) 18-20, siendo los últimos dos considerados por el ministerio de educación como un nivel satisfactorio del rendimiento de los estudiantes; además la institución educativa necesita conocer el promedio de las calificaciones en los distintos grados, estos requisitos fueron plasmados en los puntos RQ1, 2 y 3 de la Tabla 11. De igual forma, la institución educativa desea conocer la cantidad de docentes que cumplen o no con determinado aspecto observado durante el monitoreo de desempeño pedagógico, para esto se proponen los RQ4, 5 y 6 en la Tabla 11.

Tabla 11 Requerimientos de información.

Fuente: (Elaboración propia)

REQ	DESCRIPCION
RQ1	Se desea conocer cuál es la calificación promedio por área, grado, nivel, periodo de evaluación en un cuadro y gráfico comparativo que permita evidenciar el cambio o mejora.
RQ2	Se desea conocer cuál que cantidad de estudiantes tienen una calificación por debajo y por encima del promedio de determinado periodo.
RQ3	Se desea conocer cuál es la cantidad de estudiantes que alcanzan un nivel de rendimiento en los siguientes rangos de calificaciones: 0-10, 11-13, 14-17, 18-20 para cada periodo, grado/sección y curso
RQ4	Se desea conocer cuál es la cantidad y el porcentaje de docentes que cumplen con determinado aspecto observado en el monitoreo docente por periodo, grado y curso monitoreado.
RQ5	Se desea conocer cuál es la cantidad y porcentaje de docentes que no cumplen con determinado aspecto observado en el monitoreo docente por periodo, grado y curso monitoreado.
RQ6	Se desea conocer el puntaje promedio de todos los docentes en determinado aspecto observado por periodo, grado y sección.
RQ7	Se desea conocer el número de docentes que cumplen con los indicadores (nivel logrado) para determinado aspecto observado en un periodo, grado y sección.

5.3.1.2 Identificar indicadores y perspectivas

En este paso de la metodología HEFESTO se procede a identificar las perspectivas necesarias para cubrir las necesidades de información. Se identificó 11 indicadores a partir de los 7 requerimientos de información identificados en la Tabla 11, además se analizó las perspectivas necesarias para cada uno de los indicadores como se muestra en la Tabla 12.

Tabla 12 Indicadores y perspectivas.

Fuente: (Elaboración propia)

REQ	INDICADORES	PERSPECTIVAS
RQ1	Calificación promedio de estudiantes	Periodo académico, periodo de evaluación, grado, sección, curso
RQ2	Cantidad de estudiantes con una calificación por debajo del promedio	Periodo académico, periodo de evaluación, grado, sección, curso
RQ2	Cantidad de estudiantes con una calificación por encima del promedio	Periodo académico, periodo de evaluación, grado, sección, curso
RQ3	Cantidad de estudiantes que tienen una calificación entre 0-10	Periodo académico, periodo de evaluación, grado, sección, curso
RQ3	Cantidad de estudiantes que tienen una calificación entre 11-13	Periodo académico, periodo de evaluación, grado, sección, curso
RQ3	Cantidad de estudiantes que tienen una calificación entre 14-17	Periodo académico, periodo de evaluación, grado, sección, curso
RQ3	Cantidad de estudiantes que tienen una calificación entre 18-20	Periodo académico, periodo de evaluación, grado, sección, curso
RQ4	Cantidad de docentes que cumplen los aspectos observados en el monitoreo	Periodo académico, periodo de evaluación, grado, sección, aspecto observado
RQ5	Cantidad de docentes que no cumplen los aspectos observados en el monitoreo	Periodo académico, periodo de evaluación, grado, sección, aspecto observado
RQ6	Puntaje promedio de un aspecto observado durante el monitoreo	Periodo académico, periodo de evaluación, grado, sección, aspecto observado
RQ7	Número de docentes que cumplen con los indicadores del monitoreo	Periodo académico, periodo de evaluación, grado, sección, aspecto observado

5.3.1.3 Modelo Conceptual

De acuerdo con la metodología HEFESTO, se define los modelos conceptuales que representen gráficamente los indicadores y perspectivas. Esto permitirá el desarrollo de una delimitación del trabajo a realizar. A continuación se muestra los modelos conceptuales obtenidos, en la Ilustración 32 se muestra el modelo conceptual para las calificaciones de estudiantes, en la parte izquierda se encuentran las perspectivas identificadas y en la derecha los indicadores definidos, de igual manera, en la Ilustración 33 se muestra el modelo conceptual para el monitoreo docente.

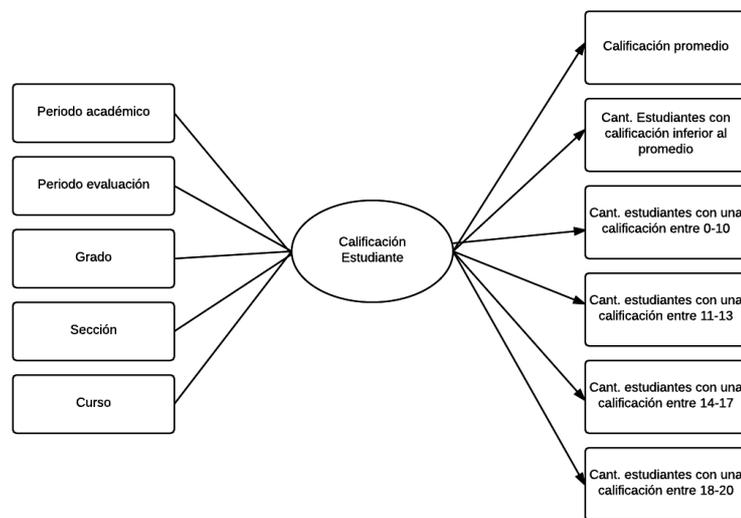


Ilustración 32 Modelo Conceptual Calificación Estudiante.

Fuente: (Elaboración Propia)

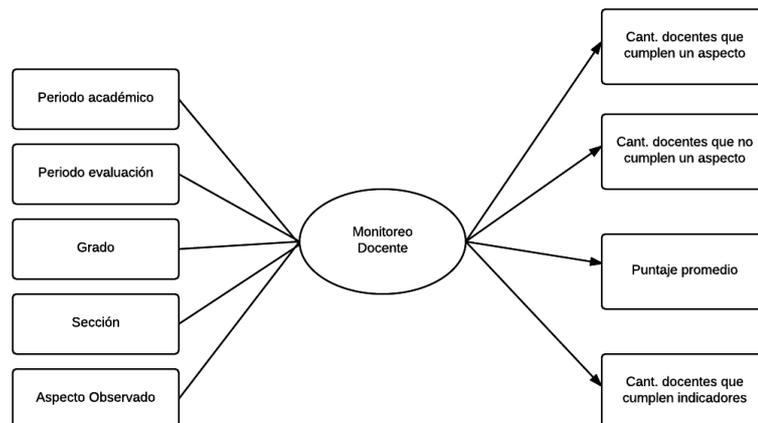


Ilustración 33 Modelo Conceptual Calificación Monitoreo.

Fuente: (Elaboración Propia)

5.3.2 Análisis de los OLTP

En este paso se hace el análisis de los OLTP que contienen los datos a ser usados en el datamart, para esto se especifica cómo serán compuestos los indicadores definidos en pasos anteriores, luego se establece la relación que existe entre la base de datos fuente (OLTP) y las perspectivas deseadas.

5.3.2.1 Conformación de indicadores

En la Tabla 13 se definen los indicadores y como se obtendrá su valor.

Tabla 13 Conformación de indicadores.

Fuente: (Elaboración propia)

INDICADOR	Fórmula de cálculo
Calificación promedio de estudiantes	<i>AVG (CALIFICACIÓN)</i>
Cantidad de estudiantes con una calificación por debajo del promedio	<i>SUM (IF CALIFICACIÓN < PROMEDIO_CURSO THEN 1 ELSE 0)</i>
Cantidad de estudiantes con una calificación por encima del promedio	<i>SUM (IF CALIFICACIÓN > PROMEDIO_CURSO THEN 1 ELSE 0)</i>
Cantidad de estudiantes que tienen una calificación entre 0-10	<i>SUM (IF CALIFICACIÓN BETWEEN 0 AND 10 THEN 1 ELSE 0)</i>
Cantidad de estudiantes que tienen una calificación entre 11-13	<i>SUM (IF CALIFICACIÓN BETWEEN 11 AND 13 THEN 1 ELSE 0)</i>
Cantidad de estudiantes que tienen una calificación entre 14-17	<i>SUM (IF CALIFICACIÓN BETWEEN 14 AND 17 THEN 1 ELSE 0)</i>
Cantidad de estudiantes que tienen una calificación entre 18-20	<i>SUM (IF CALIFICACIÓN BETWEEN 18 AND 20 THEN 1 ELSE 0)</i>
Cantidad de docentes que cumplen los aspectos observados en el monitoreo	<i>COUNT(IF ASPECTO = TRUE)</i>
Cantidad de docentes que no cumplen los aspectos observados en el monitoreo	<i>COUNT(IF ASPECTO = FALSE)</i>
Puntaje promedio de un aspecto observado durante el monitoreo	<i>AVG(PUNTAJE_MONITOREO)</i>
Número de docentes que cumplen con los indicadores del monitoreo	<i>COUNTIF (PUNTAJE > PUNTAJE_MIN)</i>

5.3.2.2 Establecer correspondencias

A continuación, mostramos la correspondencia entre las tablas OLTP desde las cuales se extraerán los datos y las perspectivas definidas en pasos anteriores. En la Ilustración 34 se muestra el diagrama de tablas construido para soportar la información obtenida en la institución educativa. En la Tabla 14 se muestra las correspondencias entre las perspectivas analizadas y las tablas existentes en el OLTP.

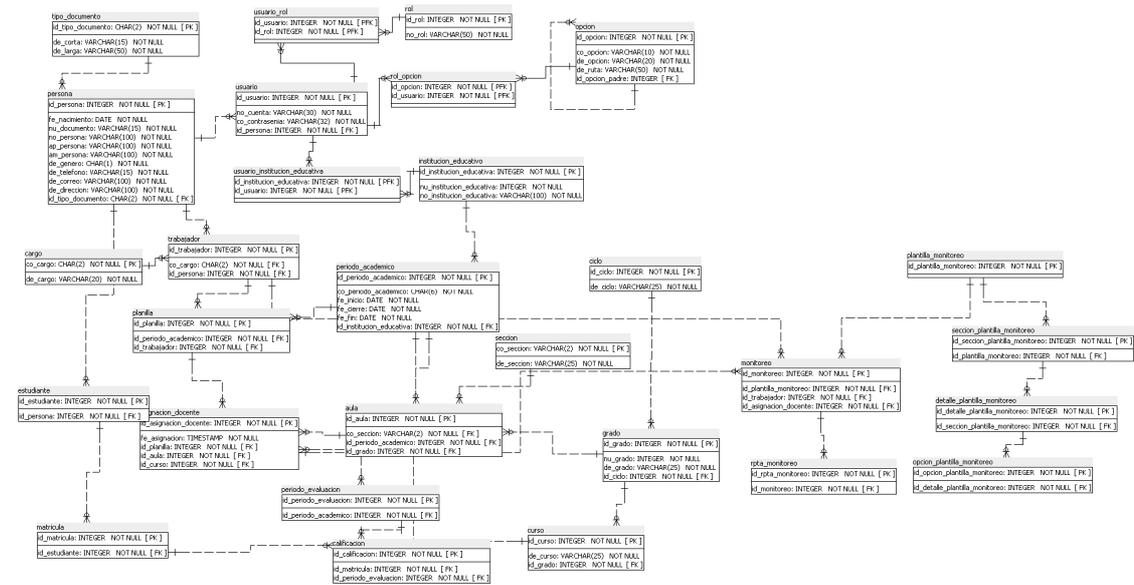


Ilustración 34 Modelo OLTP

Fuente: (Elaboración propia)

Tabla 14 Correspondencias de datos OLTP – ETL.

Fuente: (Elaboración propia)

TABLA OLTP	CAMPO	PERSPECTIVA	CAMPO
Periodo académico	ID_PERIODO_ACADEMICO	Periodo académico	Identificador
	CO_PERIODO_ACADEMICO		Código de periodo académico
Periodo calificación	ID_PERIODO_CALIFICACION	Periodo calificación	Identificador
	NO_PERIODO_CALIFICACION		Nombre del periodo de calificación
Curso	ID_CURSO	Curso	Identificador

	DE_CURSO		Nombre del curso
Grado	ID_GRADO	Grado	Identificador
	DE_GRADO		Nombre grado
Ciclo	ID_CICLO		Identificador ciclo
	DE_CICLO		Nombre ciclo
Sección	CO_SECCION	Sección	Código de sección
Sección plantilla monitoreo	ID_SECCION_PLANTILLA	Aspecto observado	Identificador
	DE_SECCION_PLANTILLA		Descripción

5.3.2.3 Nivel de granularidad

A continuación, vamos a definir los campos que se requieren para poder realizar las labores de análisis y que son considerados en el modelo de datos. Pasando a definir cada uno de ellos.

- Con respecto a la perspectiva Periodo los datos que se pueden utilizar son los siguientes:
 - Identificador del periodo académico: El periodo académico en el que se desarrolla el actual año lectivo.
 - Nombre del periodo de calificación: Trimestre en el que se da la evaluación del estudiante. Usualmente tiene valores de los trimestres en los que los docentes registran las calificaciones de sus estudiantes.
 - Descripción_trimestre: Descripción del trimestre en que se realizó la evaluación, nombre largo.
 - Código de periodo académico: Año en el que se desarrolla la evaluación.
- Con respecto a la perspectiva Curso los datos que se pueden utilizar son los siguientes:
 - Nombre del curso: Descripción del grado, de acuerdo a como se encuentra en el DCN
 - Identificador: Identificador del área que se utiliza en el sistema.
- Con respecto a la perspectiva Aspecto observado, los datos que se pueden utilizar son los siguientes.
 - Descripción: Descripción del criterio que se establece para el monitoreo.

- Identificador: Identificador del criterio que se usa en el sistema.

5.3.3 Modelo Lógico del datamart

A continuación, se procederá a realizar el modelo lógico a utilizar en el datamart. En esta sección se definirá el tipo de modelo a utilizar y posteriormente se procederá a la definición formal de las tablas.

5.3.3.1 Tipo de modelo lógico del datamart

Para este proyecto requerimos un esquema que (1) responda rápidamente a las consultas realizadas, además que sea (2) simple y (3) escalable. Dentro de nuestras opciones encontramos los esquemas de copo de nieve, estrella y constelación, se seleccionó el esquema de constelación ya que es un esquema fácil de entender, no normalizado y facilita la interacción con la herramienta de consulta y análisis. Además, se seleccionó este modelo sobre el modelo en estrella debido a que las perspectivas requeridas son muy similares para los dos procesos que se están analizando, calificaciones de estudiantes y monitoreo docente.

5.3.3.2 Tablas dimensionales

En relación con las perspectivas definidas anteriormente, se definen las tablas de dimensiones, en la Tabla 15 se listan las dimensiones consideradas para este trabajo.

Tabla 15 Descripción de Tablas de dimensiones-

Fuente: (Elaboración propia)

Dimensión	Descripción
Área	Esta dimensión engloba todas las áreas del proceso educativo. Tiene como principal funcionalidad filtrar los datos de las tablas de hechos por área.
Periodo	Esta dimensión engloba los periodos de evaluación y académicos de evaluación de los estudiantes.
Aspecto observador	Esta dimensión engloba los criterios de evaluación que se utilizan durante un monitoreo.
Grado	Esta dimensión ayuda a definir los grados del sistema educativo en los que una educación educativa desarrolla sus funciones, puede variar de institución en educación.
Sección	Esta dimensión incluye las diferentes secciones que tiene la institución educativa.

5.3.3.3 Tablas de hechos

Las tablas de hechos se han construido basándose en los requerimientos para los indicadores definidos anteriormente, se han identificado dos tablas de hechos (1) HEC_CALIFICACION y (2) HEC_MONITOREO. En la Tabla 16 se muestra un listado con las tablas de hechos que se han tomado en cuenta.

Tabla 16 Descripción de Tabla de hechos.

Fuente: (Elaboración propia)

Tabla de hechos	Descripción
HEC_CALIFICACION	En esta tabla se tiene los indicadores a ser usados para responder las preguntas respecto a las notas de los estudiantes y el desempeño que estos han tenido en los diferentes periodos.
HEC_MONITOREO	En esta tabla se tiene los indicadores a ser usados para responder las preguntas respecto al estado general de los docentes en determinado monitoreo.

5.3.3.4 Relación entre las tablas de hechos y las dimensiones

Con las tablas de hechos y de dimensiones definidas en los puntos anteriores, procedemos a definir las relaciones entre estas dos, para esto en la Ilustración 35 se muestra el diagrama diseñado para los datos de las calificaciones y su relación con las dimensiones encontradas, de igual manera en la Ilustración 36 se muestra el diagrama para los hechos de monitoreo.

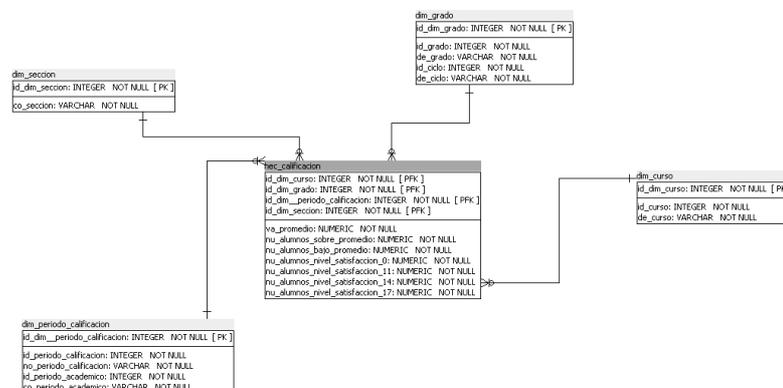


Ilustración 35 Diagrama estrella para los hechos de calificaciones.

Fuente: (Elaboración propia)

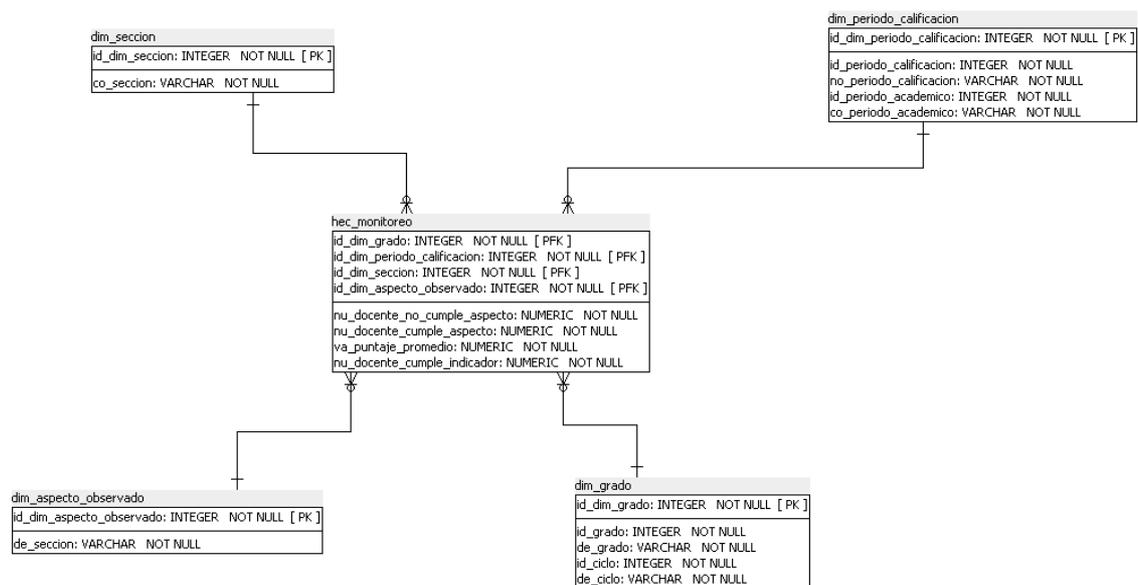


Ilustración 36 Diagrama en estrella para los hechos de monitoreo.

Fuente (Elaboración propia)

5.3.4 Definición del proceso ETL

En esta sección se define el proceso ETL con el que operará el datamart. Para esto se describe el proceso de carga inicial con el que se realiza la primera migración y luego el proceso que se realizará periódicamente para actualizar los datos en el datamart.

5.3.4.1 Carga Inicial

Se realizará la carga inicial de información desde el sistema SIAGIE, para esto se obtuvieron las calificaciones de los estudiantes en formato Excel y se procesaron con un script en Python para formatear y estandarizar estos datos, y llenar nuestra base de datos OLTP. La Ilustración 37 muestra el formato que fue procesado durante la carga inicial, durante el proceso de estandarización se hizo uso de la librería faker (Faker, 2016) con la finalidad de generar nombres y DNI aleatorios para proteger los datos de los estudiantes. De igual manera la información de las calificaciones fue extraída desde las boletas de los estudiantes exportadas desde el SIAGIE en formato PDF haciendo uso de la librería PDFMiner (PDFMiner, 2016), en la Ilustración 38 podemos ver el formato de estos documentos.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	textbox6	textbox9	textbox10	textbox8	textbox11	textbox12	textbox13	textbox14											
2	Nivel y Mo	Secundaria	Turno:	TARDE	Grado:	CUARTO	Sección:	A											
3																			
4	textbox5	textbox39	textbox4	textbox7	textbox15	textbox19	textbox21	textbox23	textbox25	textbox27	textbox29	textbox31	textbox33	textbox35	textbox37	textbox43	textbox44	textbox45	textbox4
5	Código del Apellidos	1	0	0	0	0	0	0	0	7	5	2	6	4	5	6	9	AGUEDO CHAVEZ ASTRID SARITA	F
6	Código del Apellidos	2	0	5	0	0	5	1	9	6	9	4	0	0	0	1	0	APAZA CCOROPUNA ANA BECILIA	F
7	Código del Apellidos	3	0	0	0	0	0	0	0	7	5	2	6	4	5	6	4	APAZA QUISPE ALEX	M
8	Código del Apellidos	4	0	0	0	0	0	0	0	7	0	5	8	8	5	6	6	AYTE UÑADILLO LEYDI ROCIO	F
9	Código del Apellidos	5	0	0	0	0	0	0	0	7	6	5	4	2	1	0	3	CANCHARI GABRIEL STEFANY	F
10	Código del Apellidos	6	0	0	0	0	0	0	0	7	5	6	6	9	4	0	4	CCALLOCONTO CAYETANO LINCOLN RENE	M
11	Código del Apellidos	7	0	0	0	0	0	0	0	7	5	2	9	1	1	0	6	CHAVEZ LOPEZ VALENTIN MIGUEL	M
12	Código del Apellidos	8	0	4	1	3	0	3	3	3	8	0	0	0	3	0	0	HERRERA SALDAÑA ROYLE	M
13	Código del Apellidos	9	0	0	0	0	0	0	0	7	6	6	5	9	0	2	2	HUILCA CHOQUEHUANCA IANNED	F
14	Código del Apellidos	10	0	0	0	0	0	0	0	7	6	4	1	5	4	1	8	LAVADO DE LA CRUZ CRISTINA CAROLINE	F
15	Código del Apellidos	11	0	6	0	7	7	7	1	1	0	0	0	0	6	4	0	LEON RAFAELE JHON LENIN	M
16	Código del Apellidos	12	0	0	0	0	0	0	0	7	4	9	9	9	6	4	4	LEON CUETO FLAVIO CESAR	M
17	Código del Apellidos	13	0	0	0	0	0	0	0	7	6	5	4	2	7	8	6	LEONARDO LOAYZA SAYDERS EDGAR	M
18	Código del Apellidos	14	0	0	0	0	0	0	0	7	2	3	7	8	7	6	6	MAMANI TUNQUI NATALY INES	F
19	Código del Apellidos	15	0	4	1	5	0	1	6	8	3	0	0	0	5	0	0	MARCOS ROMERO AHIBETH YESENIA	F
20	Código del Apellidos	16	0	6	0	5	1	8	5	8	9	0	0	0	4	0	0	QUISPE CHIPAYO ANA LIZBETH	F
21	Código del Apellidos	17	1	1	1	4	6	6	3	9	0	0	0	8	4	8	8	RIVERA ZUÑIGA RENZO DAVID	M
22	Código del Apellidos	18	0	0	0	0	0	0	0	7	5	6	0	3	3	2	4	SUNQUILPA CARDENAS LUIS ANDERSON	M
23	Código del Apellidos	19	0	0	0	0	0	0	0	7	3	8	7	8	7	1	9	VASQUEZ GIL JORDYN FRANKLIN	M

Ilustración 37 Formato de archivos excel procesados para la carga inicial.

Fuente: SIAGIE

ÁREA	CRITERIOS DE EVALUACIÓN	BIMESTR	TRIMESTR	Calif. Final del Área	Eval. de Recuperación
MATEMÁTICA	Razonamiento y demostración	16	16	17	17
	Comunicación matemática	17	17	17	
	Resolución de problemas	17	17	18	
	Actitud ante el área	17	17	18	
CALIF. PROMEDIO AREA		17	17	18	
COMUNICACIÓN	Expresión y comprensión oral	16	14	13	16
	Comprensión de textos	19	16	16	
	Producción de textos	18	17	14	
	Actitud ante el área	17	16	14	
CALIF. PROMEDIO AREA		17	16	14	
INGLÉS	Expresión y comprensión oral	17	16	20	18
	Comprensión de textos	17	17	20	
	Producción de textos	20	20	20	
	Actitud ante el área	18	15	20	
CALIF. PROMEDIO AREA		18	17	20	
ARTE	Expresión artística	16	15	16	15
	Apreciación artística	16	16	14	
	Actitud ante el área	17	15	14	
	CALIF. PROMEDIO AREA	16	16	16	
HISTORIA GEOGRAFÍA Y ECONOMÍA	Manejío de información	10	12	15	15
	Comprensión espacio-temporal	12	16	15	
	Juicio crítico	14	16	16	
	Actitud ante el área	14	16	17	
CALIF. PROMEDIO AREA		13	15	16	
FORMACIÓN CIUDADANA Y CÍVICA	Construcción de la cultura cívica	18	18	18	18
	Ejercicio ciudadano	16	18	17	
	Actitud ante el área	16	16	19	
	CALIF. PROMEDIO AREA	18	18	18	
PERSONA, FAMILIA Y RELACIONES HUMANAS	Construcción de la autonomía	15	17	17	16
	Relaciones Interpersonales	12	15	18	
	Actitud ante el área	14	16	16	
	CALIF. PROMEDIO AREA	14	17	17	
EDUCACIÓN FÍSICA	Comprensión y desarrollo de la Corp. y la Salud	18	17	18	17
	dominio corporal y expresión creativa	18	18	17	
	Convivencia e interacción sociomotriz	19	16	14	
	Actitud ante el área	16	16	16	
CALIF. PROMEDIO AREA		18	17	17	
EDUCACIÓN RELIGIOSA	Comprensión doctrinal cristiana	17	19	18	19
	Discernimiento de fe	18	19	19	
	Actitud ante el área	19	20	19	
	CALIF. PROMEDIO AREA	18	19	19	
CIENCIA, TECNOLOGÍA Y AMBIENTE	Comprensión de información	15	16	16	16
	Indagación y experimentación	14	15	15	
	Actitud ante el área	17	17	17	
	CALIF. PROMEDIO AREA	15	16	16	
EDUCACIÓN PARA EL TRABAJO	Gestión de procesos	15	17	15	15
	Ejecución de procesos	15	15	16	
	Comprensión y aplicación de tecnologías	12	15	17	
	Actitud ante el área	15	16	16	
CALIF. PROMEDIO AREA		14	16	16	
COMPORTAMIENTO*		A	A	A	A

Ilustración 38 Formato PDF de calificaciones exportado desde el SIAGIE.

Fuente: SIAGIE

El proceso ETL se dividirá en dos partes, la primera se encargará de realizar la carga de tablas dimensionales y la segunda se encargará de cargar las tablas de hechos. En la Ilustración 39 se muestra el proceso definido para la carga de información por primera vez, en este caso se inicia con la carga de cada una de las dimensiones y finalmente se cargan las medidas definidas para las tablas de hechos.

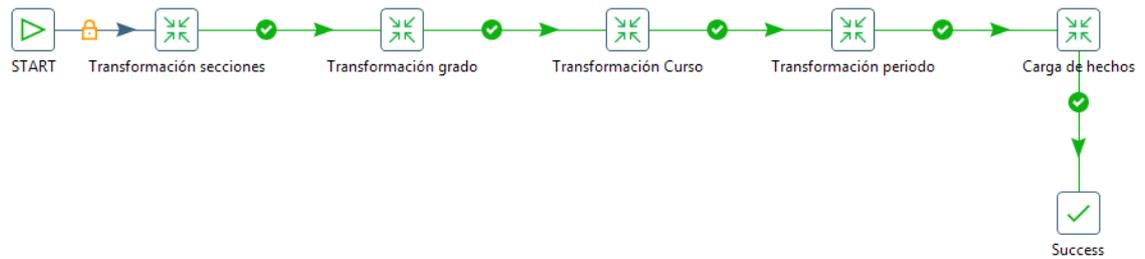


Ilustración 39 Proceso ETL con Pentaho Data Integration.

Fuente: Elaboración propia

El Job definido en la Ilustración 39 define 5 transformaciones de las cuales 4 cargan datos de dimensiones, estas transformaciones son implementadas haciendo uso de consultas SQL que extraen datos directamente desde la base de datos transaccional y actualizan la información necesaria en las tablas del datamart. Como se puede ver en la Ilustración 40, las transformaciones para la carga de datos en las tablas de dimensiones consisten en extraer los datos de la base de datos OLTP (Table Input) y actualizarla en la tabla de dimensiones (Update/Insert), una de las consultas utilizadas para la transformación de la dimensión periodo se puede apreciar en la Ilustración 41. Este proceso es similar en las demás transformaciones.

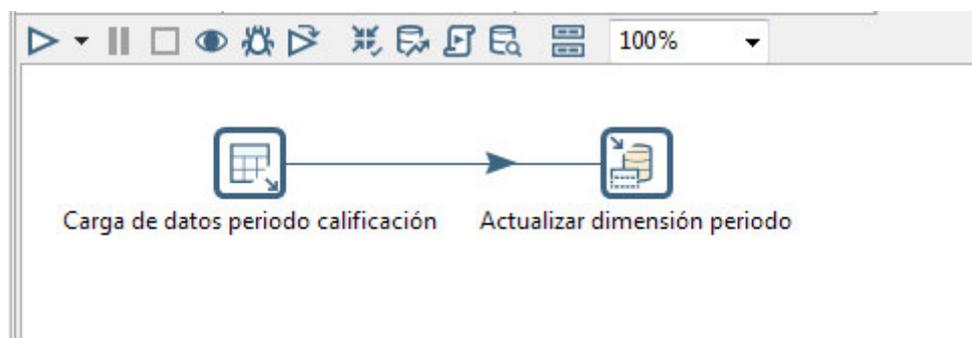


Ilustración 40 Transformación Carga Periodo

Fuente: Elaboración propia

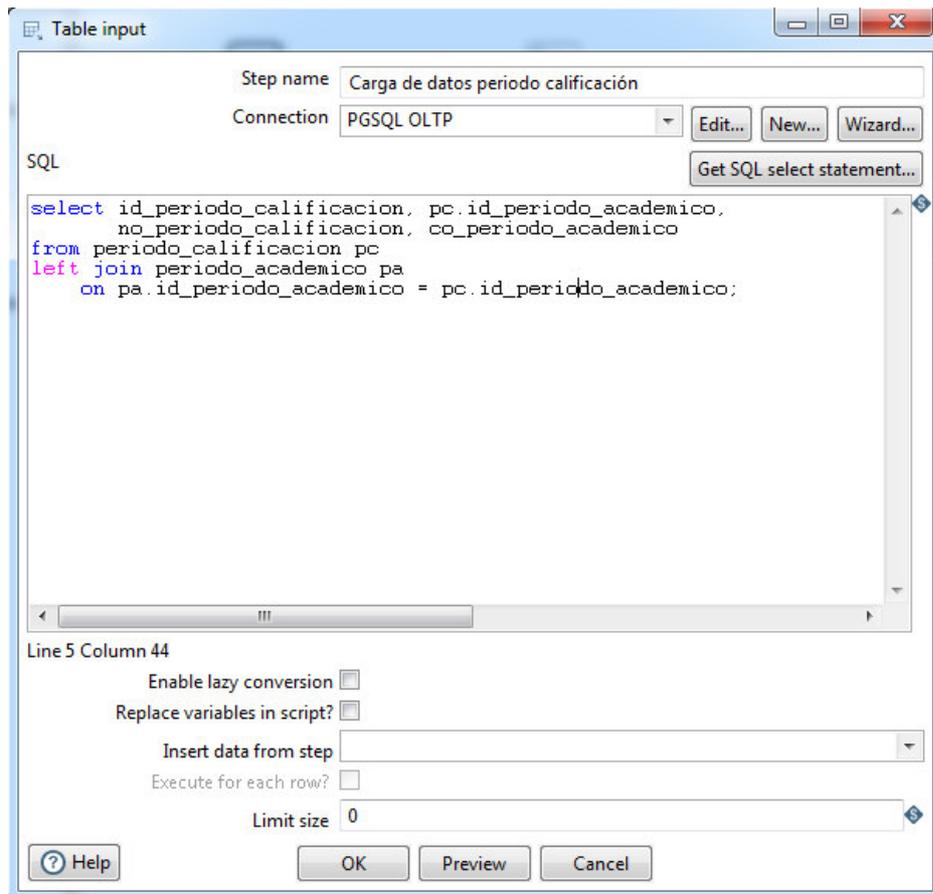


Ilustración 41 Consulta de carga de datos para ETL en Pentaho Data Integration.

Fuente: Elaboración propia

Para la tabla de hechos se hace uso del mismo método utilizado en la carga de datos en las dimensiones del datamart, con la diferencia que para la tabla de hechos se tiene que utilizar funciones de agregación y agrupar los resultados de acuerdo a los requisitos de información de la institución educativa. En la Ilustración 42 podemos ver el procedimiento definido para la carga de la tabla de hechos, en este procedimiento se hizo uso de un paso más, el “Lookup”, que tiene la finalidad de buscar la correspondencia entre el identificador que pertenece a la base de datos OLTP y el nuevo identificador que fue asignado al momento de hacer la migración en el datamart.



Ilustración 42 Procedimiento definido para la carga de datos en la tabla de hechos de calificaciones en Pentaho Data Integration.

Fuente: Elaboración propia

El procedimiento de extracción de datos para las tablas de hechos es, de igual forma, mediante una consulta SQL pero siendo más complejo debido al procesamiento que se debe realizar para cumplir con los requisitos de información como se puede ver en la Ilustración 43 e Ilustración 44.

```
select k.id_periodo_calificacion, k.id_grado, k.co_seccion, k.id_curso, k.va_not_promedio_curso_grado,
sum(case when p.va_not_promedio_alumno > k.va_not_promedio_curso_grado then 1 else 0 end) nu_sobre_avg,
sum(case when p.va_not_promedio_alumno < k.va_not_promedio_curso_grado then 1 else 0 end) nu_bajo_avg,
sum(case when va_not_promedio_alumno < 11 then 1 else 0 end) nu_nv1_st_0,
sum(case when va_not_promedio_alumno >= 11 and va_not_promedio_alumno < 14 then 1 else 0 end) nu_nv1_st_11,
sum(case when va_not_promedio_alumno >= 14 and va_not_promedio_alumno < 18 then 1 else 0 end) nu_nv1_st_14,
sum(case when va_not_promedio_alumno >= 18 then 1 else 0 end) nu_nv1_st_17
from
(select id_periodo_calificacion, id_grado, co_seccion, id_curso, avg(va_not_promedio_alumno) va_not_promedio_alumno,
avg(va_not_promedio_curso_grado) va_not_promedio_curso_grado
from
(select id_periodo_calificacion, id_grado, co_seccion, cu.id_curso, c.id_matricula, avg(va_not_promedio_alumno) va_not_promedio_alumno
from calificacion c
left join matricula m on m.id_matricula = c.id_matricula
left join aula a on a.id_aula = m.id_aula
left join criterio_evaluacion ce on ce.id_criterio_evaluacion = c.id_criterio_evaluacion
left join curso cu on cu.id_curso = ce.id_curso
group by id_periodo_calificacion, id_grado, co_seccion, cu.id_curso, c.id_matricula) q
group by id_periodo_calificacion, id_grado, co_seccion, id_curso) k
left join
(select id_periodo_calificacion, id_grado, co_seccion, cu.id_curso, c.id_matricula, avg(va_not_promedio_alumno) va_not_promedio_alumno
from calificacion c
left join matricula m on m.id_matricula = c.id_matricula
left join aula a on a.id_aula = m.id_aula
left join criterio_evaluacion ce on ce.id_criterio_evaluacion = c.id_criterio_evaluacion
left join curso cu on cu.id_curso = ce.id_curso
group by id_periodo_calificacion, id_grado, co_seccion, cu.id_curso, c.id_matricula) p
on p.id_periodo_calificacion = k.id_periodo_calificacion
and p.id_grado = k.id_grado and p.co_seccion = k.co_seccion and p.id_curso = k.id_curso
group by k.id_periodo_calificacion, k.id_grado, k.co_seccion, k.id_curso, k.va_not_promedio_curso_grado
order by id_periodo_calificacion, id_grado, co_seccion, id_curso;
```

Ilustración 43 Consulta para carga de hechos de calificaciones.

Fuente: (Elaboración propia)

```

SELECT
  k.id_seccion_plantilla_monitoreo,
  m3.id_perido_calificacion,
  a3.id_grado,
  a3.co_seccion,
  avg(puntaje)          va_puntaje_promedio,
  sum(cumple_indicador) nu_cumple_indicador,
  sum(cumple_aspecto)  nu_cumple_aspecto,
  sum(no_cumple_aspecto) nu_no_cumple_aspecto
FROM (SELECT
  m2.id_monitoreo,
  q.id_seccion_plantilla_monitoreo,
  puntaje,
  (CASE WHEN puntaje > spm2.va_puntaje_minimo
    THEN 1
    ELSE 0 END) cumple_indicador,
  (CASE WHEN puntaje > 0
    THEN 1
    ELSE 0 END) cumple_aspecto,
  (CASE WHEN puntaje > 0
    THEN 0
    ELSE 1 END) no_cumple_aspecto
FROM (SELECT
  m.id_monitoreo,
  spm.id_seccion_plantilla_monitoreo,
  sum(va_respuesta) puntaje
FROM respuesta_monitoreo rm
LEFT JOIN monitoreo m ON m.id_monitoreo = rm.id_monitoreo
LEFT JOIN detalle_plantilla_monitoreo dpm
  ON dpm.id_detalle_plantilla_monitoreo = rm.id_detalle_plantilla_monitoreo
LEFT JOIN seccion_plantilla_monitoreo spm
  ON spm.id_seccion_plantilla_monitoreo = dpm.id_seccion_plantilla_monitoreo
LEFT JOIN asignacion_docente ad ON ad.id_asignacion_docente = m.id_asignacion_docente
LEFT JOIN aula a ON a.id_aula = ad.id_aula
LEFT JOIN periodo_calificacion pc ON pc.id_perido_calificacion = a.id_perido_academico
GROUP BY m.id_monitoreo, spm.id_seccion_plantilla_monitoreo) q
LEFT JOIN seccion_plantilla_monitoreo spm2
  ON spm2.id_seccion_plantilla_monitoreo = q.id_seccion_plantilla_monitoreo
LEFT JOIN monitoreo m2 ON m2.id_monitoreo = q.id_monitoreo
LEFT JOIN asignacion_docente ad2 ON ad2.id_asignacion_docente = m2.id_asignacion_docente
LEFT JOIN aula a2 ON a2.id_aula = ad2.id_aula) k
LEFT JOIN monitoreo m3 ON m3.id_monitoreo = k.id_monitoreo
LEFT JOIN asignacion_docente ad3 ON ad3.id_asignacion_docente = m3.id_asignacion_docente
LEFT JOIN aula a3 ON a3.id_aula = ad3.id_aula
GROUP BY id_seccion_plantilla_monitoreo, id_perido_calificacion, a3.id_grado, a3.co_seccion
ORDER BY id_seccion_plantilla_monitoreo, id_perido_calificacion, a3.id_grado, a3.co_seccion;

```

Ilustración 44 Consulta para carga de hechos monitoreo.

Fuente: (Elaboración propia)

5.3.4.2 Actualización

Los requisitos de información no requieren una carga diaria del datamart ya que las calificaciones y los monitoreos no son actividades que se realicen todos los días, es por esto que se han convenido las siguientes políticas de actualización:

- La información se actualizará a demanda de acuerdo con las necesidades de los usuarios.
- La información será actualizada de ser necesaria en caso que existan modificaciones ya sea de alguna dimensión o de una calificación, la información de los hechos de calificaciones y monitoreos serán actualizada

siempre y cuando el periodo académico al que pertenecen los datos se encuentra aperturado. No se podrá actualizar datos de periodos anteriores.

- El procedimiento de carga de datos será similar al establecido en la carga inicial con la diferencia que se modificará la consulta SQL de la carga de la tabla de hechos para reflejar que sólo se carguen periodos aperturados.

5.3.4.3 Limpieza de datos

Como parte del proceso ETL se define la actividad de limpieza de datos con la finalidad de mantener la confiabilidad de los datos presentados por el datamart, este proceso se ejecuta en dos ocasiones, primero se hace una revisión de los datos de las calificaciones que se encuentra en el sistema y luego se realiza una verificación de los datos que serán migrados, tanto en las dimensiones con la finalidad de migrar datos consistentes, los datos detectados fuera de los parámetros establecidos serán registrados en un archivo de texto para que puedan ser revisados en la fuente de datos, corregidos y migrados posteriormente. Con la finalidad de lograr la confiabilidad del datamart, se establecen los siguientes criterios que los datos a migrar deben cumplir para que puedan ser migrados.

Tabla 17 Criterios para limpieza de datos.

Fuente: (Elaboración propia)

Elemento	Restricción
Calificación de estudiantes	La calificación de los estudiantes que será migrada debe encontrarse en el rango de 0 a 20 puntos.
Sección	Los valores migrados no deben ser vacíos.
Grado	Los valores migrados no deben ser vacíos.
Curso	Los valores migrados no deben ser vacíos.
Periodo	Los valores migrados no deben ser vacíos.
Hechos de calificaciones	La calificación promedio obtenida debe encontrarse en el rango de 0 y 20, además, los valores calculados para las cantidades de estudiantes que cumplen un criterio deben ser enteros y mayores a cero.
Hechos de monitoreo	Los puntajes de los monitoreos son validados para asegurarnos que sean valores numéricos y que sean mayores a 0, además la cantidad de docentes que cumplen o no ciertos criterios son verificados para que sean valores enteros.

Para implementar la limpieza de datos en el proceso ETL definido en 5.3.4.1 se realizaron cambios en el job que controla la carga de datos agregando un paso de

validación de calificaciones, y posteriormente, se agregó un paso de validación en cada una de las transformaciones ejecutadas en el job para que realicen la validación correspondiente con la data de la dimensión o hecho migrado. En la Ilustración 45 se muestra la tarea de migración de calificaciones en la que se incluye la transformación de calificaciones de calificación con la finalidad de asegurar que la información de calificaciones sea la correcta.

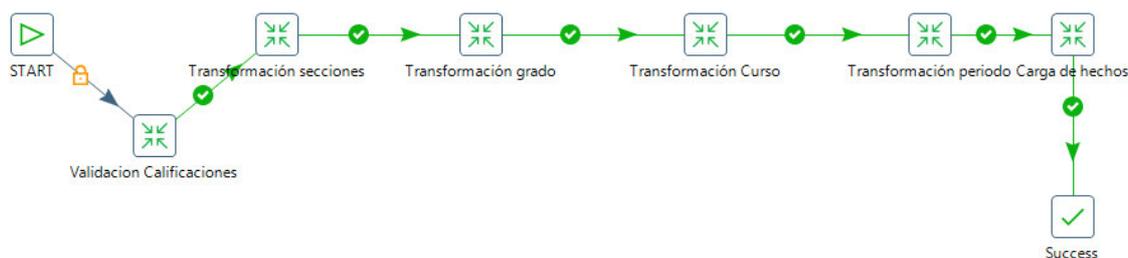


Ilustración 45 Job de migración de calificaciones modificado con validación de datos.

Fuente: (Elaboración propia)

Como podemos ver en la Ilustración 46 la actividad aborta cuando se encuentra con valores irregulares en las calificaciones de los estudiantes y estos valores son registrados en un archivo log para que puedan ser revisados en la fuente de datos y de esta forma asegurar la confiabilidad de los resultados mostrados en este proyecto.



Ilustración 46 Validación de calificaciones.

Fuente: (Elaboración propia)

De igual manera, durante la carga de datos a las dimensiones del datamart, se agregó un paso previo de validación, respecto al modelo mostrado en la Ilustración 40, mediante el cual se refuerza la fiabilidad de los datos migrados. Se muestra en la Ilustración 47 que la validación permite que se migren los datos de la dimensión periodo, registrando en un log de errores, los datos que no cumplen con los criterios de validación para que sean corregidos en la fuente de datos.

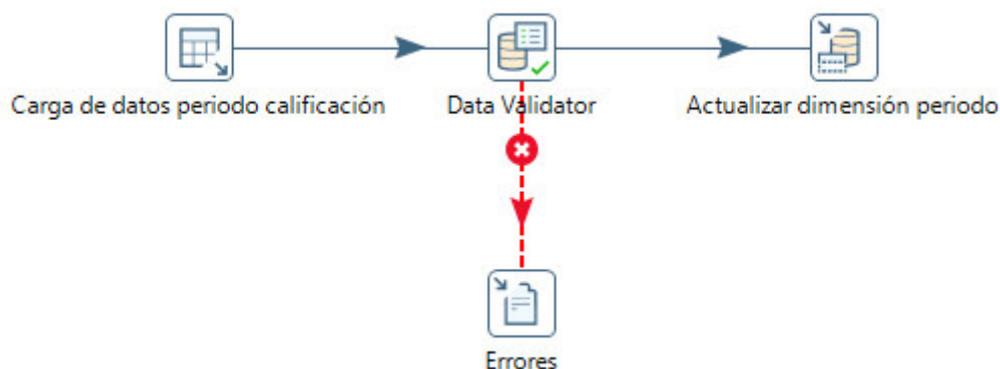


Ilustración 47 Validación de datos en migración de dimensiones.

Fuente: (Elaboración propia)

5.3.5 Creación de cubos multidimensionales

Para la creación de los cubos multidimensionales se hará uso de la herramienta Schema Workbench de Pentaho, que nos permitirá crear archivos XML que son leídos en el servidor SAIKU en el que podremos hacer las consultas. En esta sección se crearán dos cubos, la primera lo llamaremos “Cubo de calificaciones” y el segundo “Cubo de monitoreos”.

5.3.5.1 Creación de indicadores

En esta sección se crea el schema con las medidas para que puedan ser procesadas por el motor de consulta. Para esto se crearon los indicadores por cada uno de los cubos y se asignó la fórmula correspondiente como se especifica a continuación:

- Cubo de calificaciones
 - Calificación promedio:
 - AVG (VA_NOTA_PROMEDIO)
 - Estudiantes bajo el promedio:
 - SUM (NU_ESTUDIANTES_BAJO_PROMEDIO)
 - Estudiantes sobre el promedio:
 - SUM (NU_ESTUDIANTES_SOBRE_PROMEDIO)
 - Estudiantes con calificación entre 0-10:
 - SUM (NU_ALUMNOS_NIVEL_SATISFACCION_0)
 - Estudiantes con calificación entre 11-13:
 - SUM (NU_ALUMNOS_NIVEL_SATISFACCION_11)
 - Estudiantes con calificación entre 14-17:
 - SUM (NU_ALUMNOS_NIVEL_SATISFACCION_14)
 - Estudiantes con calificación entre 18-20:

- SUM (NU_ALUMNOS_NIVEL_SATISFACCION_17)
 - Cubo de monitoreo
 - Docentes que cumplen con un aspecto
 - SUM(NU_DOCENTE_CUMPLE_ASPECTO)
 - Docentes que no cumplen con un aspecto
 - SUM(NU_DOCENTE_NO_CUMPLE_ASPECTO)
 - Docentes que cumplen indicador
 - SUM(NU_DOCENTE_CUMPLE_INDICADOR)
 - Puntaje promedio
 - SUM(VA_PUNTAJE_PROMEDIO)

5.3.5.2 Creación de dimensiones

En el schema work bench se pueden crear las dimensiones del cubo relacionándolas con las dimensiones definidas por el datamart. En este proyecto se crearon 4 dimensiones por cada cubo, para el cubo de calificaciones se crearon las dimensiones Sección, Grado, Curso y Periodo calificación y para el cubo de Monitoreo se crearon las dimensiones Sección, Grado, Periodo de calificación y Aspecto observado como se muestra en Ilustración 48, de igual forma se definió la jerarquía correspondiente al periodo de calificación como se muestra en la Ilustración 49.

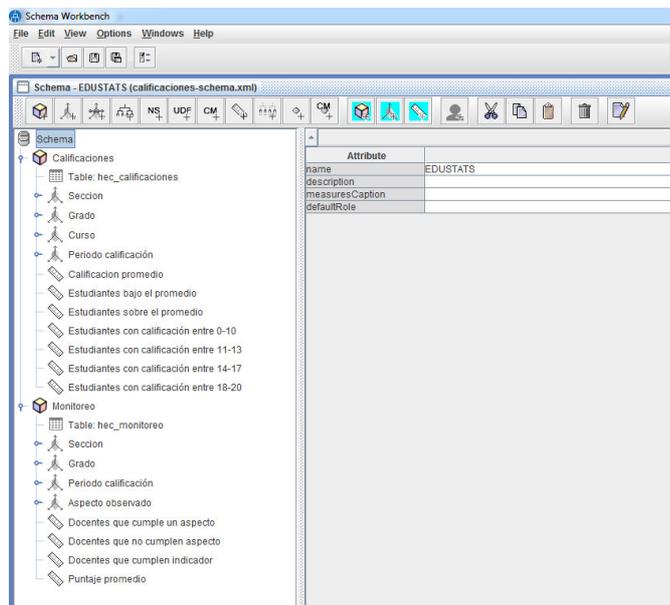


Ilustración 48 Schema de definición de cubos en Schema Workbench de Pentaho.

Fuente: (Elaboración propia)

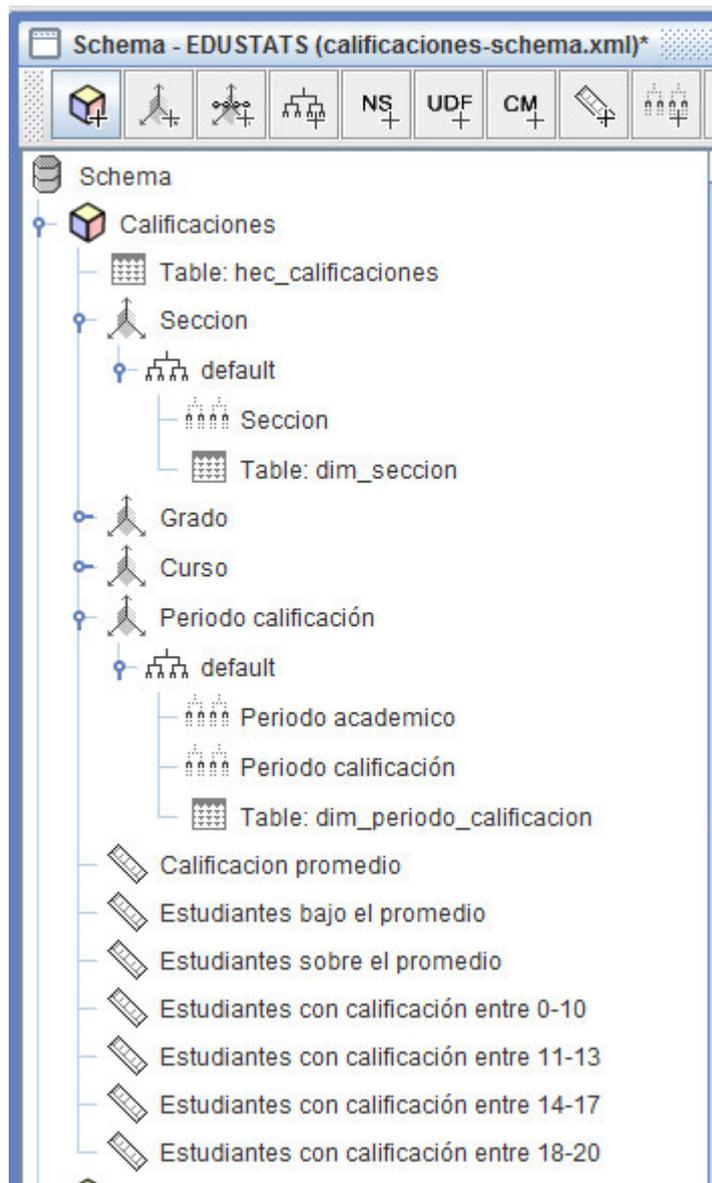


Ilustración 49 Definición de jerarquías para el periodo de calificación.

Fuente: (Elaboración propia)

5.4 Implementación de consultas con SAIKU BI

Para poder realizar las consultas se utilizó el servidor de SAIKU BI (Meteorite BI, 2016) que es una aplicación web que hace uso de Pentaho Mondrian internamente para ejecutar consultas sobre cubos multidimensionales. En esta plataforma se elaboraron consultas predeterminados solicitadas por los usuarios como se puede ver en las ilustraciones siguientes. Además, se permite que el usuario pueda modificar y/o crear nuevas consultas para poder explorar sus datos en los cubos definidos.

A continuación, presentamos el sistema de consultas analíticas implementado con SAIKU BI, internamente los datos de los estudiantes y docentes son consultados a los cubos multidimensionales definidos en los pasos anteriores, en la Ilustración 50, podemos observar la vista de las calificaciones promedio de los estudiantes por periodo, grado y sección en las filas y en las columnas de los cursos seleccionados. En la Ilustración 51, se muestra un cuadro comparativo entre las calificaciones promedio de dos secciones para el curso de matemática y su evolución a través del tiempo.

Periodo académico	Grado	Sección	MATEMÁTICA	COMUNICACIÓN	INGLÉS	ARTE	HISTORIA, GEOGRAFÍA Y ECONOMÍA	FORMACIÓN CÍVICA Y CÍVICA	PERSONA, FAMILIA Y RELACIONES HUMANAS	EDUCACIÓN FÍSICA	EDUCACIÓN RELIGIOSA	CENCIA, TECNOLOGÍA Y AMBIENTE	EDUCACIÓN PARA EL TRABAJO	
2013	PRIMERO SECUNDARIA	A	12.462	12.795	12.795	13.793	11.709	15.115	14.255	14.473	14.110	12.051	13.660	
		B	11.925	13.619	13.619	13.381	11.774	13.937	14.201	14.651	14.28	12.048	13.714	
		C	11.22	12.185	12.185	13.127	10.393	14.167	12.692	12.649	12.622	11.349	12.682	
	SEGUNDO SECUNDARIA	A	13.169	12.668	12.668	13.993	10.931	14.369	14.302	14.074	13.933	12.469	14.125	
		B	13.127	12.167	12.167	13.665	11.745	14.241	14.012	14.539	14.485	12.316	13.095	
		C	12.318	12.542	12.542	13.669	10.217	14.273	14.671	14.172	12.015	12.924		
	TERCERO SECUNDARIA	A	12.95	11.61	11.61	13.353	13.354	13.321	13.679	12.47	13.233	12.211	13.586	
		B	11.959	11.174	11.174	13.059	13.86	12.726	14.104	13.495	12.595	11.427	12.706	
		C	12.293	13.162	13.162	12.744	14.647	10.105	14.077	14.612	14.208	12.966	13.33	
	CUARTO SECUNDARIA	A	12.187	13.313	13.313	14.329	15.393	15.173	14.762	14.477	14.704	13	13.007	
		B	10.833	11.979	11.979	14.179	12.567	13.472	14.344	14.456	13.979	12.317	13.433	
		C	11.656	11.806	11.806	14.156	12.033	13.319	14.719	14.211	14.415	12.319	13.605	
	2014	PRIMERO SECUNDARIA	A	12.38	12.167	12.167	12.644	13.937	12.762	16.755	14.947	14.62	12.769	12.269
			B	12.132	13.52	13.52	13.011	14.049	12.948	10.623	14.376	15.467	12.498	12.322
			C	12.116	13.055	13.055	13.08	12.778	13.438	17.84	14.403	16.204	13.772	12.319
SEGUNDO SECUNDARIA		A	11.946	13.031	13.031	13	12.246	13.351	16.925	14.513	15.467	13.926	11.909	
		B	12.969	12.244	12.244	13.369	12.229	13.316	16.959	14.162	15.035	13.961	12.195	
		C	13.443	12.744	12.744	13.17	12.149	13.883	16.289	14.67	14.769	14.439	12.272	
TERCERO SECUNDARIA		A	13.652	12.889	12.889	13.778	12.111	13.026	17.019	14.107	15.090	13.706	11.917	
		B	13.949	13.37	13.37	13.722	12.708	14.265	17.632	14.102	15.753	13.716	13.324	
		C	11.644	12.622	12.622	14.356	13.181	13.493	16.946	14.669	15.193	12.919	13.184	
CUARTO SECUNDARIA		A	13.967	12.286	12.286	13.679	13.266	12.657	16.147	14.027	15.907	12.147	12.601	
		B	11.083	12.891	12.891	13.671	13.047	16.912	16.939	15.011	16.266	12.946	13.466	
		C	11.566	13.644	13.644	14.134	13	16.826	15.456	15.003	16.644	12.369	13.696	
2015		PRIMERO SECUNDARIA	A	12.716	13.57	12.798	13.996	13.255	12.926	16.999	15.313	14.68	12.854	12.93
			B	12.634	13.906	13.253	14.29	13.995	12.471	16.972	16.966	15.322	13.405	12.966
			C	11.745	12.463	14.634	14.764	12.696	14.471	17.393	15.294	15.21	13.026	13.261
	SEGUNDO SECUNDARIA	A	11.764	12.105	14.205	14.474	14	14.246	15.342	14.953	15.269	13.417	13.668	
		B	12.229	12.289	14.439	15.417	12.764	14.666	17.323	15.431	15.719	14.014	14.014	
		C	13.146	12.053	13.525	14.479	12.999	13.295	15.743	14.334	14.611	12.939	12.939	
	TERCERO SECUNDARIA	A	12.765	11.726	13.046	14.432	12.143	12.472	17.379	14.666	14.533	12.961	12.972	
		B	11.662	13.275	14.997	15.616	13.941	14.471	16.029	14.961	15	13.966	14.266	
		C	10.721	12.277	13.233	14.017	13.35	12.4	15.433	14.722	14.217	13.479	13.656	
	CUARTO SECUNDARIA	A	12.13	12.826	13.802	14.681	13.636	13.764	17.722	15.029	14.429	13.764	13.483	
		B	11.462	13.452	15.072	15.37	14.397	14.619	17.945	15.903	15.03	13.067	14.106	
		C	10.41	11.474	11.474	14.44	14.116	11.464	14.416	15.444	15.444	13.444	13.444	

Ilustración 50 Vista de calificaciones promedio por periodo, grado, sección y curso en saiku.

Fuente: (Elaboración propia)

MATEMATICA			
PRIMERO SECUNDARIA			
	A	B	
Periodo academico	Calificacion promedio	Calificacion promedio	
2013	12.242	11.925	
2014	12.38	12.132	
2015	12.716	12.824	

Ilustración 51 Vista comparativa de calificaciones promedio en estudiantes por año en secciones diferentes. Fuente: (Elaboración propia)

En la Ilustración 52, podemos ver la evolución que tienen los estudiantes en los resultados de calificación promedio en diferentes periodos y grados lo que nos ofrece una vista de la evolución de los estudiantes durante su paso por la institución educativa, por ejemplo, los alumnos que iniciaron el primer año de secundaria en el periodo 2013 incrementaron su calificación promedio en periodos posteriores obteniendo 12.107 en el segundo año de secundaria en el periodo del 2014 y un último incremento cuando el mismo grado paso al siguiente año, 2015, obteniendo un promedio de 12.971 (las calificaciones mostradas han sido filtradas para el curso de matemática), esta información es mostrada en un gráfico en la Ilustración 53.

	PRIMERO SECUNDARIA	SEGUNDO SECUNDARIA	TERCERO SECUNDARIA	
Periodo academico	Calificacion promedio	Calificacion promedio	Calificacion promedio	
2013	11.798	11.195	11.819	
2014	12.258	12.107	13.748	
2015	12.67	11.91	12.971	

Ilustración 52 Cuadro comparativo que evidencia la calificación promedio en los estudiantes en tres periodos académicos.

Fuente: (Elaboración propia)

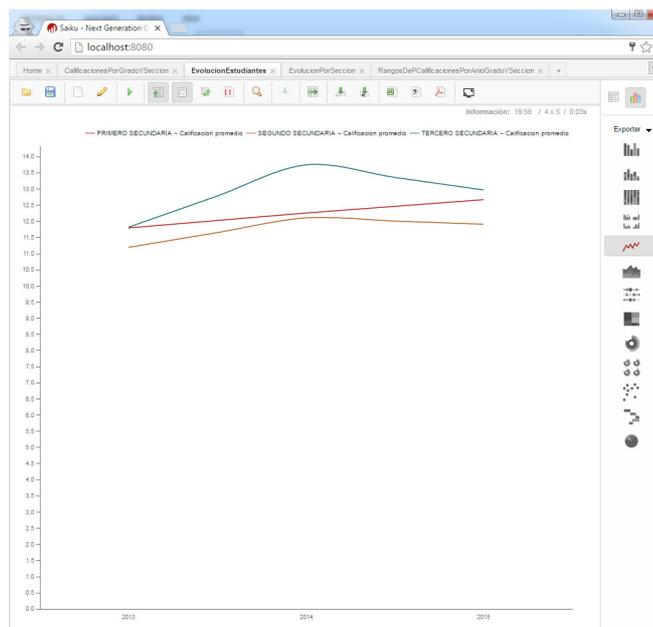


Ilustración 53 Gráfico de las calificaciones de los estudiantes en diferentes periodos.

Fuente: (Elaboración propia)

Periodo académico	Periodo calificación	Grado	Sección	Curso	Estudiantes con calificación entre 0-10	Estudiantes con calificación entre 11-13	Estudiantes con calificación entre 14-17	Estudiantes con calificación entre 18-20
2013	PRIMER TRIMESTRE	PRIMERO SECUNDARIA	A	MATEMÁTICA	6	13	3	0
				COMUNICACIÓN	5	6	11	0
				B	MATEMÁTICA	9	6	6
			COMUNICACIÓN	0	10	0	0	
			C	MATEMÁTICA	8	6	0	0
			COMUNICACIÓN	3	9	2	0	
	SEGUNDO TRIMESTRE	PRIMERO SECUNDARIA	A	MATEMÁTICA	9	6	7	0
				COMUNICACIÓN	5	12	9	0
				B	MATEMÁTICA	10	7	6
			COMUNICACIÓN	1	12	7	1	
			C	MATEMÁTICA	9	4	1	0
			COMUNICACIÓN	6	7	1	0	
TERCER TRIMESTRE	PRIMERO SECUNDARIA	A	MATEMÁTICA	9	9	4	0	
			COMUNICACIÓN	1	15	8	0	
			B	MATEMÁTICA	12	6	3	0
		COMUNICACIÓN	1	13	9	2		
		C	MATEMÁTICA	7	6	1	0	
		COMUNICACIÓN	0	12	2	0		
2014	PRIMER TRIMESTRE	PRIMERO SECUNDARIA	A	MATEMÁTICA	3	19	3	0
				COMUNICACIÓN	3	19	3	0
				B	MATEMÁTICA	4	20	5
			COMUNICACIÓN	1	14	14	0	
			C	MATEMÁTICA	3	14	8	0
			COMUNICACIÓN	6	17	2	0	
	SEGUNDO TRIMESTRE	PRIMERO SECUNDARIA	A	MATEMÁTICA	2	24	3	0
				COMUNICACIÓN	3	9	17	0
				B	MATEMÁTICA	6	19	0
			COMUNICACIÓN	12	10	3	0	
			C	MATEMÁTICA	9	19	1	0
			COMUNICACIÓN	1	15	13	0	
TERCER TRIMESTRE	PRIMERO SECUNDARIA	A	MATEMÁTICA	10	11	6	0	
			COMUNICACIÓN	2	19	6	0	
			B	MATEMÁTICA	12	12	3	0
		COMUNICACIÓN	0	22	7	0		
		C	MATEMÁTICA	11	8	2	1	
		COMUNICACIÓN	0	13	14	0		
2015	PRIMERO TRIMESTRE	PRIMERO SECUNDARIA	A	MATEMÁTICA	11	10	7	1
				COMUNICACIÓN	0	6	21	0
				B	MATEMÁTICA	0	6	0

Ilustración 54 Resultados obtenidos por rangos en tres periodos académicos por curso, grado y sección.

Fuente: (Elaboración propia)

En la Ilustración 54 se muestra una tabla comparativa de los resultados obtenidos por los estudiantes de acuerdo con los rangos establecidos por el ministerio de educación en tres periodos diferentes. En la Ilustración 55 mostramos los puntajes promedio que fueron obtenidos en los monitoreos realizados a los docentes durante los periodos 2014-2015, en los distintos aspectos evaluados y en la

Ilustración 56, podemos ver los docentes que lograron cumplir con los indicadores propuestos por el ministerio de educación (superan el puntaje mínimo).

Información: 12:45 / 4 x 7 / 0.09s

Aspecto observado	2013	2014	2015
Estrategias metodológicas empleadas por el docente (Datos a ser registrados mediante la OBSERVACIÓN)	Puntaje promedio 12.8	Puntaje promedio 13.407	Puntaje promedio 12.138
Uso de materiales y recursos didácticos por el docente (Registrar estos datos a partir de la OBSERVACIÓN)	6.171	6.47	6.391
Gestión del tiempo para los aprendizajes --Registrar estos datos a partir de la OBSERVACION	5.988	5.822	5.489
Clima de aula-- Registrar estos datos a partir de la OBSERVACION	4.987	5.349	5.128
Planificación-- Datos a ser registrados CONSULTANDO al docente	0.842	1.059	0.971

Ilustración 55 Puntajes promedios de los monitoreos por periodo y aspecto observado en SAIKU.

Fuente: (Elaboración propia)

Información: 12:47 / 7 x 11 / 0.09s

Periodo académico	Periodo e calificación	PRIMERO SECUNDARIA	SEGUNDO SECUNDARIA	TERCERO SECUNDARIA	CUARTO SECUNDARIA	QUINTO SECUNDARIA
		Docentes que cumplen indicador				
2013	PRIMER TRIMESTRE	3		0	3	7
	SEGUNDO TRIMESTRE	1	1		1	3
	TERCER TRIMESTRE	0	1	3	2	0
2014	PRIMER TRIMESTRE	0	2	7	4	
	SEGUNDO TRIMESTRE	4	2	4	5	1
	TERCER TRIMESTRE		4	4	4	2
2015	PRIMER TRIMESTRE	3	3	2	2	3
	SEGUNDO TRIMESTRE	0	4	2	0	5
	TERCER TRIMESTRE	1	2	1	2	5

Ilustración 56 Número de docentes que cumplen un indicador por periodo y grado en SAIKU.

Fuente: (Elaboración propia)

CAPÍTULO VI. RESULTADOS

En la siguiente sección se comentarán los resultados obtenidos de esta investigación.

6.1 Resultados

Para realizar las pruebas de la solución planteada se coordinó con la IE N° 170 Santa Rosa del Sauce – San Juan de Lurigancho, se utilizó datos de los estudiantes del nivel secundaria por tres años 2013, 2014 y 2015, de todos sus grados y secciones. Los datos de los estudiantes fueron anonimizados para los propósitos de este trabajo y procesados para poder ser cargados en el datamart. Finalmente se obtuvo un datamart con información de tres años y se diseñaron consultas predefinidas en SAIKU BI, proporcionando de esta manera una plataforma de consultas para satisfacer las necesidades de los usuarios.

El sistema demostró permitir a sus usuarios realizar menos tareas de recopilación de información lo que les significó una disminución de tiempos muertos, además gracias al registro de los datos del monitoreo, se pudo automatizar este proceso y brindar datos históricos que mediante el método tradicional anterior eran difíciles de obtener. A continuación, se muestra un cuadro comparativo en el que se detallan los tiempos que requerían los docentes para la obtención y análisis de los resultados de manera tradicional.

Tabla 18 Comparativa proceso actual vs proceso propuesto.

Fuente: (Elaboración propia)

Criterio	Actual (días)	Propuesto (días)
Tiempo de recopilación de monitoreos.	2 – 10	1 – 2
Tiempo de procesamiento de los monitoreos.	3 – 10	1
Tiempo de recopilación de datos de estudiantes.	10 – 15	10 – 15
Tiempo de procesamiento de datos de estudiantes	10 – 15	1 – 2

Los resultados mostrados en la tabla anterior pueden ser variables de acuerdo a los parámetros establecidos por la administración, dándose a notar que el proceso propuesto es muy prometedor debido a que el procesamiento automático eliminaría

la necesidad de tener a alguna persona teniendo este trabajo. Además el acceso a la información histórica es una ventaja de la propuesta ya que muchas veces mantener los historiales de docentes o estudiantes es un trabajo pesado.

CAPÍTULO VII. CONCLUSIONES Y TRABAJOS FUTUROS

7.1 Conclusiones

- El caso de estudio demuestra que la aplicación propuesta cumple con el objetivo de este trabajo de investigación, se logró implementar el datamart con la información histórica de tres años y definir un proceso de carga periódico para la actualización de datos en ejecuciones futuras, facilitando de esta manera el acceso a datos históricos y disminuyendo los tiempos de recopilación de los mismos.
- Se realizó una comparación de las metodologías para la implementación de un datamart que incluye las metodologías de Raph Kimball, Bill Inmon, HEFESTO y DWEP. Se seleccionó la metodología HEFESTO ya que no requiere un nivel alto de documentación y permite realizar la implementación de un data warehouse incremental permitiendo que se desarrolle este trabajo con una sola persona.
- Se realizó una comparación de las herramientas existentes para la implementación del datamart, donde se analizaron tanto herramientas licenciadas como MS SQL Server Integration Services & Analysis además de herramientas open source como Pentaho y Jaspersoft. Se seleccionó Pentaho y Saiku BI ya que no tiene un costo de licenciamiento lo que hace este proyecto viable para las instituciones educativas públicas.
- El modelo seleccionado durante la implementación de la metodología para el diseño del datamart fue determinado en su mayor parte por la relación entre las perspectivas que fueron analizadas para los indicadores solicitados, así por ejemplo tanto los indicadores necesarios para las calificaciones y monitoreo requieren las perspectivas periodo y grado por lo que en este caso se seleccionó un modelo constelación, este modelo facilita reutilizar las tablas de dimensiones lo que a su largo plazo significa un ahorro en espacio de almacenamiento.

7.2 Recomendaciones y trabajos futuros

- El proyecto no realiza un trabajo de minería de datos, lo que podría ser importante para la búsqueda de patrones en las bases de datos de estudiantes, teniendo un enfoque más amplio, como utilizar información de estudiantes de diferentes instituciones educativas, para así llevar la búsqueda de buenas prácticas en educación a un universo con más opciones.
- El proyecto puede ser ampliado tomando en cuenta otros factores que influyen en el aprendizaje de los estudiantes, como información de las familias de los estudiantes.
- El proyecto puede ser expandido tomando en cuenta un alcance mayor gracias a que parte de procesos estándar en las instituciones públicas.

REFERENCIAS BIBLIOGRÁFICAS

- Basurto Figueroa, C. (2011). *ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN DATAMART PARA EL ÁREA DE MANTENIMIENTO Y LOGÍSTICA DE UNA EMPRESA DE TRANSPORTE PÚBLICO DE PASAJEROS*. Lima: PUCP.
- Benavides, M., Espinosa, G., & Montané, A. (2002). *Evaluación de sistema y evaluación de aula*. Lima: UMC - Ministerio de Educación.
- Bernabeu, R. D. (2010). *Metodología para la construcción de un data warehouse*. Córdoba.
- Chen, H., Chiang, R. H., & Storey, V. C. (2012). BUSINESS INTELLIGENCE AND ANALYTICS: FROM BIG DATA TO BIG IMPACT. *MIS Quarterly*, 1165-1188.
- Dario, B. R. (2009). *DATA WAREHOUSING: Investigación y sistematización de conceptos*. Argentina.
- Faker. (12 de Junio de 2016). *Faker*. Obtenido de <https://github.com/joke2k/faker>
- Falcón Cisneros, S. D., Palma Stanciuc, R. I., & Celi Peña, R. A. (2007). *CONSTRUCCIÓN DE UNA HERRAMIENTA DE DESARROLLO DE SOLUCIONES PARA INTELIGENCIA DE NEGOCIOS - EXPLOTACIÓN*. Lima.
- Fernandez Ochoa, E. (2009). *ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN DATAMART DE CLIENTES PARA EL AREA DE MARKETING DE UNA ENTIDAD ASEGURADORA*. Lima.
- Galvis Lista, E., Gonzales Zabala, M., García Avendaño, A., & Benavides Fawcett, L. (2011). Soluciones de Inteligencia de negocios en la práctica: apoyo a la toma de decisiones en proyectos educativos para población infantil vulnerable en el caribe Colombiano. *UIS Ingenierías*, 123-135.
- INEI. (Septiembre de 1997). *ONGEI*. Recuperado el Junio de 2013, de ONGEI: <http://www.ongei.gob.pe/publica/metodologias/Lib5084/libro.htm>
- Inmon, W. H. (2002). *Building the Data Warehouse*. John Wiley and Sons Inc.
- Jaspersoft. (12 de Junio de 2016). *JasperReports Server Architecture*. Obtenido de <http://community.jaspersoft.com/wiki/jasperreports-server-architecture>
- Jaspersoft. (12 de Junio de 2016). *Jaspersoft*. Obtenido de <http://www.jaspersoft.com/es>
- Kimball, R., & Caserta, J. (2004). *The Data Warehouse ETL Toolkit*. Indianapolis: Wiley Publishing, Inc.

- Kimball, R., & Ross, M. (2002). *The Data Warehouse Toolkit 2ed.* John Wiley and Sons.
- Laursen, G. H., & Thorlund, J. (2010). *Business Analytics for Managers.* Hoboken, New Jersey: SAS Institute Inc.
- Luján Mora, S., & Trujillo, J. (2004). A Data Warehouse Engineering Process. *3rd International Conference in Advances in Information Systems (ADVIS 2004)*, (págs. 14-23). Alicante.
- Meteorite BI. (29 de Abril de 2016). *Saiku*. Obtenido de Productos: <http://www.meteorite.bi/products/saiku>
- Microsoft. (2011). *Education System Deploys Business Intelligence Solution to Improve Performance.* Washington: Microsoft. Obtenido de https://www.microsoft.com/danmark/cases/Case_Study_Detail.aspx?CaseStudyID=4000010377
- Microsoft. (12 de Junio de 2016). *An Introduction to Fast Track Data Warehouse Architecture.* Obtenido de [https://technet.microsoft.com/en-us/library/dd459146\(v=sql.100\).aspx](https://technet.microsoft.com/en-us/library/dd459146(v=sql.100).aspx)
- Microsoft. (12 de Junio de 2016). *BI and Analytics.* Obtenido de <https://www.microsoft.com/en-us/cloud-platform/bi-analytics>
- Microsoft. (12 de Junio de 2016). *SQL Server.* Obtenido de <http://www.microsoft.com/en-us/sqlserver/solutions-technologies/enterprise-information-management/integration-services.aspx>
- Microstrategy. (2010). *Architecture for enterprise business intelligence - an overview of the microstrategy platform architecture for big data, cloud bi, and mobile applications.* Microstrategy.
- MINEDU. (15 de 08 de 2016). *ESCALE - Estadística de la calidad educativa.* Obtenido de Tendencias: <http://escale.minedu.gob.pe/tendencias>
- MINEDU. (13 de Junio de 2016). *PETI – Ministerio de Educación.* Obtenido de http://www.minedu.gob.pe/normatividad/resoluciones/rm_0191-2011-ed.pdf
- MINEDU. (13 de Junio de 2016). *SIAGIE.* Obtenido de http://siagie.minedu.gob.pe/archivos/brochure_pbp.pdf
- Ministerio de Economía y Finanzas. (2007). *Programa Estratégico: Logros de aprendizaje al finalizar el III ciclo de educación básica regular.* Lima. Obtenido de https://www.mef.gob.pe/contenidos/presu_public/documentacion/programa_est

art/Programas_Estrategicos_Logros_de_aprendizaje_logros_aprendizaje_ciclo_III.pdf

- Nagabhushana, S. (2006). *Data Warehousing: OLAP and Data Mining*. New Delhi: New Age International.
- Núñez Soto, G. (2010). *Análisis, Diseño e Implementación de una solución de Inteligencia de Negocios para el área de finanzas de la Municipalidad Metropolitana de Lima*. Lima: PUCP.
- Oracle. (12 de Junio de 2016). *Business Analytics*. Obtenido de <https://www.oracle.com/solutions/business-analytics/index.html>
- PDFMiner. (13 de Junio de 2016). *PDFMiner*. Obtenido de <https://github.com/euske/pdfminer>
- Pentaho. (12 de Junio de 2016). *Pentaho Community*. Obtenido de <http://community.pentaho.com/>
- Prabhu, S., & Venkatesan, N. (2007). *Data Mining and Warehousing*. New Delhi: New age International (P) Limited.
- Qlik. (12 de Junio de 2016). *Data visualization, embedded analytics and reporting*. Obtenido de <http://www.qlik.com/products>
- Rational - the software development company. (s.f.). *Rational Unified Process - Best practices for software development teams*.
- Villanueva Ojeda, Á. (2008). *Análisis, Diseño e Implementación de un DataWarehouse de Soporte de Decisiones para un Hospital del Sistema de Salud Público*. Lima.

ANEXOS

ANEXO 01: Evaluación de la organización

Introducción

Este documento tiene como finalidad el desarrollo de un software que permita brindar información de la institución educativa

Propósito

El presente desarrollo se ejecuta con la finalidad de apoyar la tarea de compilación y procesamiento de información manual que se realiza actualmente en la institución educativa Santa Rosa del Sauce.

Referencias

1. Proyecto Educativo Institucional 2011 – 2015 (PEI) (http://www.micentroeducativo.pe/2013/ce/fileproject/file_colegios/pei_49c00d.pdf)

Contexto del Negocio

La Institución Educativa Santa Rosa del Sauce se desenvuelve en el ámbito educativo en el nivel inicial, primario y secundario. Ubicada en el distrito de San Juan de Lurigancho, esta institución educativa pública tiene como objetivo brindar la mejor formación a sus estudiantes.

Visión

“La I.E. N° 170 “Santa Rosa del Sauce” al 2015 es un centro de desarrollo humano; donde se incentiva, forma y potencia las habilidades y capacidades de sus estudiantes; capaces de integrarse a la sociedad y convivir en armonía consigo mismos, con los demás y su ambiente; aportando en la resolución de problemas de su comunidad; generando su propio trabajo y la de los demás, camino al desarrollo sostenible y sustentable de nuestro distrito y país.” – (Fuente: PEI SRS)

Misión

“Somos una Institución Educativa integral que contribuye a que los niños y jóvenes del AA.HH. Santa Rosa del Sauce desarrollen competencias básicas que les garanticen el cultivo del pensamiento crítico y creativo para la solución de problemas la toma de decisiones en su vida y la de su comunidad; basándonos en el

desarrollo de una cultura de paz, ética, intelectual, artística-cultural, afectiva, física y espiritual.” – (Fuente: PEI SRS)

Factores Externos

La comunidad

La comunidad del AA.HH. es el factor externo más importante de la institución educativa ya que depende de esta el comportamiento de los estudiantes, sus costumbres y el desarrollo de los estudiantes. La comunidad espera que la institución educativa logre brindar calidad educativa y que sus estudiantes salgan adelante.

Entidades Externas

Dentro de estos factores tenemos dos factores importantes, las Unidades de Gestión Educativa (UGEL) y otras Instituciones Educativas públicas. Las UGEL, tienen como finalidad la regularización de las Instituciones Educativas, controlar que la normativa vigente se cumpla en las instituciones educativas y que no se encuentren irregularidades. Otras Instituciones Educativas públicas tienen un rol importante en la compartición de información y de experiencias en solución de problemas.

Factores Internos

Procesos

A continuación, veremos los dos procesos que tiene un rol importante en este proyecto.

Proceso de monitoreo

EL proceso de monitoreo es un proceso de control pedagógico que tiene como finalidad mejorar los métodos que usan los docentes durante el desarrollo de clases. El proceso comienza con la definición de una asamblea de docentes en la cual se realiza la selección de los docentes quienes realizarán el monitoreo y además se asignarán los monitoreos que se deben realizar. Durante esta asamblea, también se elabora una ficha de monitoreo derivada de una plantilla proporcionada por la Dirección Regional de Educación (DRE). Una vez organizado los monitoreos, se procede a realizar la evaluación en el aula por parte del docente monitor asignado

y además por el director de la institución educativa, de acuerdo con la asignación que se realizó en la asamblea. Al terminar cada monitoreo, el docente se compromete con ciertas opciones de mejoras que debe implementar. Seguidamente, se procede en una fase de Evaluación, donde los docentes recopilan los datos que tienen de los distintos monitoreos que realizaron y se prepara un documento de resultados de monitoreo. Finalmente se organiza una reunión en la que se evalúan los resultados y plantean mejorar a seguir.

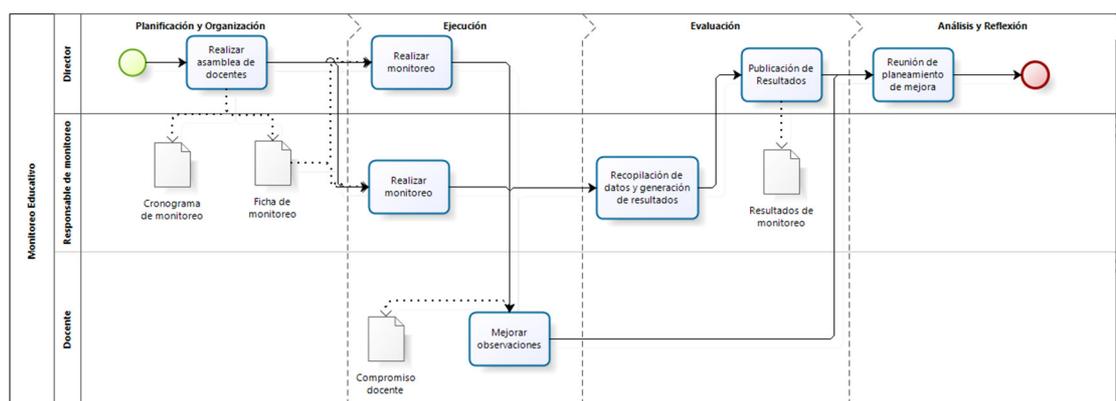


Ilustración A1.1 Monitoreo Educativo

Fuente: Elaboración propia.

Proceso Desarrollo de Año Escolar

Este proceso es realizado cada año y tiene la finalidad de realizar la organización y ejecución del año escolar. El proceso comienza con la planificación del siguiente año escolar, en esta planificación se define el número aproximado de estudiantes por grado y sección, el número de vacantes que se tendrá y la cantidad de docentes necesarios para cubrir la demanda, esto es conocido como metas de atención y son procesadas por la UGEL para que pueda gestionar el techo presupuestal de la institución educativa. Durante la matrícula, se realiza la inscripción de los estudiantes, tanto automática (estudiantes que han pasan de año), de nuevos estudiantes (inicial, primero de primaria) o traslado, además se realiza el planeamiento de la infraestructura necesaria para cubrir la demanda (asignación de salones de clase y mobiliario). Finalmente, el desarrollo de clases y todos los ajustes que se ven durante esta actividad.

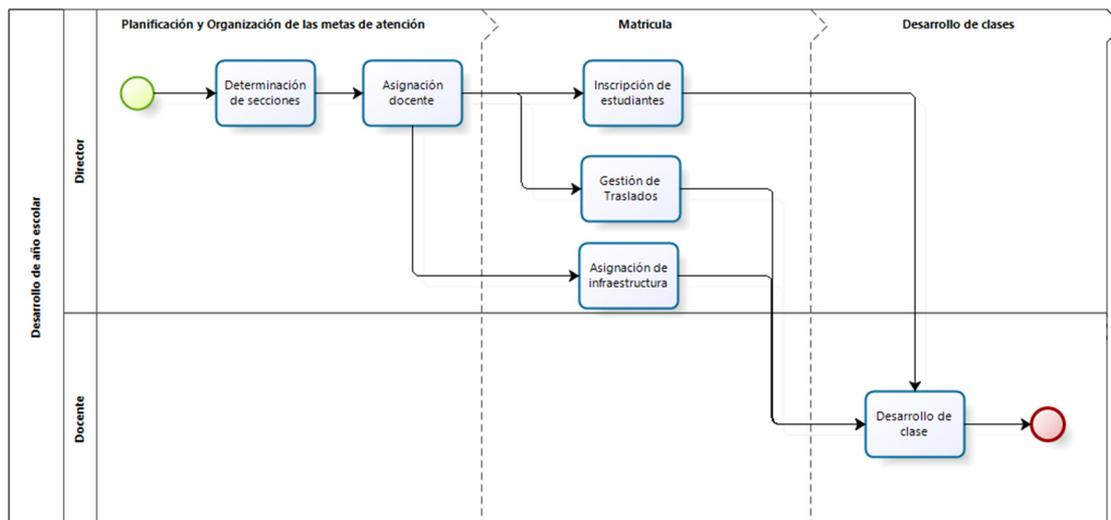


Ilustración A1.2 Desarrollo de año escolar.

Fuente: Elaboración propia

Herramientas de soporte

1. Sistema de Información de Apoyo a la Gestión de Instituciones Educativas (SIAGIE)

Actualmente todas las instituciones educativas cuentan con un sistema de información llamado SIAGIE, que tiene como finalidad apoyar a las instituciones educativas en manejo de la información de sus estudiantes. Esta aplicación cuenta la funcionalidad de registro de notas, asistencia, gestión de traslados, y otra información de los estudiantes.

2. Herramientas ofimáticas

A pesar del gran apoyo que brinda el SIAGIE a las Instituciones Educativas, la institución educativa hace uso de la herramienta Excel para generar sus informes internos y realizar un análisis de sus estudiantes.

Organización Interna

A continuación, se presenta la organización interna de la institución educativa.

1. Dirección IE N° 170 “SRS”.- Encargado de la gestión correcto funcionamiento de la IE.
2. CONEI.- El consejo educativo institucional es un grupo representativo de la institución educativa que tiene como finalidad velar por la buena gestión de la institución educativa.
3. Auxiliar de educación.- Personal encargado de apoyar en las actividades que se desarrollan durante las clases, especialmente en el nivel primario.
4. Coordinación Tutorial.- Personal encargado de la coordinación de tutorías en la institución educativa, asignación de docentes a tutorías.
5. Personal de Servicio.- Personal de apoyo en la institución educativa.
6. Consejo Académico
7. Coordinación Académica Inicial
8. Coordinación Académica Primaria
9. Coordinación Académica Secundaria
10. Coordinación por Ciclo
11. Coordinación por Área
12. Docente.- Encargado de desarrollar las clases en distintos grados/secciones.

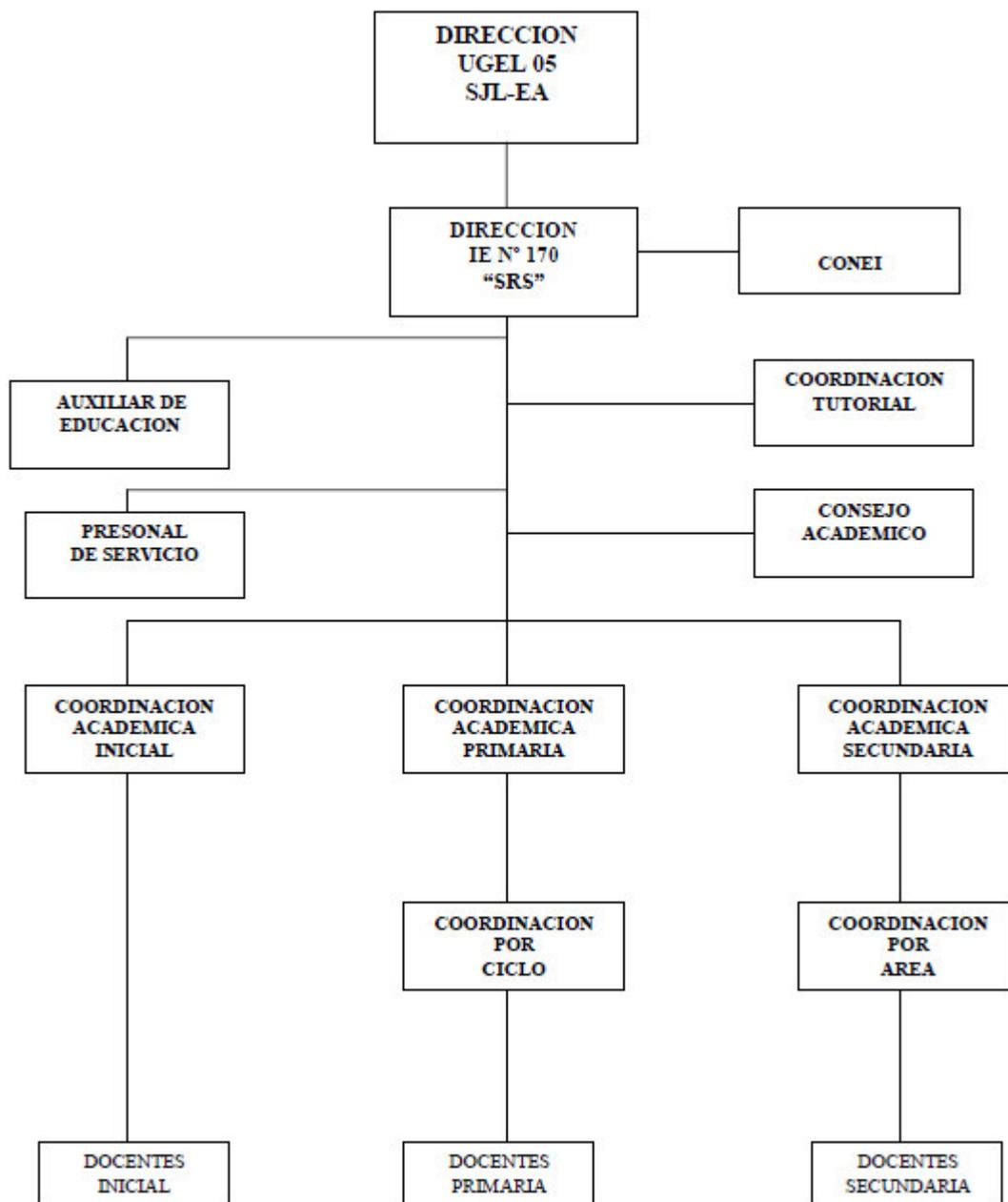


Ilustración A1.3 Organigrama Estructural.

Fuente: PEI - SRS

Características del producto

Esfuerzo de desarrollo

El presente proyecto es un proyecto medianamente complejo ya que busca generar información útil a sus usuarios que les permita operar más efectivamente.

Grado de novedad

El proyecto es novedoso ya que cambiará la forma de operar actual al momento de generar indicadores en las instituciones educativas. Además, incluirá el uso de nuevas tecnologías para la recopilación de información.

Tipo de aplicación

Esta es una aplicación web que se encuentra a la vez dentro de las aplicaciones de gestión y de apoyo a la toma de decisiones.

Complejidad técnica

El producto tiene una complejidad mediana ya que soporta dos procesos básicos en las instituciones educativas.

Conclusiones de la evaluación

De acuerdo con la evaluación realizada a la organización, la institución educativa tiene soporte a sus procesos administrativos mediante el SIAGIE, pero sin embargo, al momento de presentarse oportunidades de mejora, la institución educativa tiene dificultades para identificarlas y trabajarlas en conjunto con la finalidad de obtener mejores propuestas de educación.

ANEXO 02: FICHA DE MONITOREO 2014

FICHA DE ACOMPAÑAMIENTO PEDAGOGICO AL DOCENTE

IE N° 170 “*Santa Rosa del Sauce*”

Observador:

I. DATOS GENERALES

A. Datos de la sesión observada (Datos a ser registrados mediante la OBSERVACION)						
1. Área o áreas desarrolladas						
2. Denominación de la sesión y/o proyecto de aprendizaje						
3. Nivel educativo:	Inicial ()	Primaria ()	Secundaria ()	4. Grado(s) o año(s) observados(s)		5. Sección
6. Turno	Mañana ()	Tarde ()	7. Tiempo de duración de la sesión observada		_____ hrs., _____ min.	
8. Cantidad de estudiantes matriculados en la sección:	a. Mujeres		9. Cantidad de estudiantes asistentes a la sesión observada:		a. Mujeres	
	b. Hombres				b. Hombres	
	c. Total				c. Total	
10. Cantidad de estudiantes inclusivos matriculados en la sección:	a. Mujeres		11. Cantidad de estudiantes inclusivos asistentes a la sesión observada:		a. Mujeres	
	b. Hombres				b. Hombres	
	c. Total				c. Total	

II. DESARROLLO DE LA SESION DE APRENDIZAJE

A. Estrategias metodológicas empleadas por el docente → Datos a ser registrados mediante la OBSERVACION				
			Si	No
1	Al iniciar la sesión el o la docente motiva y genera el interés y la atención de los estudiantes?			
2	A continuación, el o la docente comunica a los y las estudiantes los aprendizajes que se esperan lograr en la sesión?			
3	¿El/la docente utiliza diversas formas de organización de los estudiantes: trabajos individuales, trabajos en grupos, pequeños, grandes, con toda el aula, etc.?			
4	¿El/la docente observado brinda atención simultánea y diferenciada a sus estudiantes?			
5	El/la docente recoge e incorpora permanente los conocimientos previos de los/las estudiantes para el desarrollo de los aprendizajes?			
6	El/la docente promueve los conocimientos y practicas interculturales para el desarrollo de los aprendizajes?			

7	El/la docente hace uso de algunas estrategias planteadas en las rutas del aprendizaje EBR y/o EIB (uso del juego, del error, la resolución de problemas, entre otras)?		
8	El/la docente ha seguido una secuencia didáctica que corresponde al desarrollo de los aprendizajes propuestos?		

B. Uso de materiales y recursos didácticos por el docente → Registrar estos datos a partir de la OBSERVACION

		Si	No
1	El/la docente utiliza materiales y/o recursos didácticos que ayudan al desarrollo de las actividades de aprendizaje propuestas para la sesión?		
2	El/la docente acompaña y orienta a los y las estudiantes durante el uso de los materiales en función del aprendizaje a lograr?		
3	El aula cuenta con la cantidad suficiente de materiales educativos para el grupo de estudiantes observados?		
4	El/la docente organiza y facilita materiales y recursos didácticos a los y las estudiantes en el momento oportuno?		
5	El/la docente adecua el uso del material educativo considerando la realidad de los estudiantes?		

C. Gestión del tiempo para los aprendizajes → Registrar estos datos a partir de la OBSERVACION

		Si	No
1	→ Responder a la siguiente pregunta solo si se logra estar en el aula en el momento en el que se inició la sesión. En caso contrario, pasar a pregunta 2 El/la docente empieza la sesión a la hora prevista?		
2	El/la docente culmina la sesión a la hora prevista?		
3	El/la docente permanece durante toda la sesión en el espacio de aprendizaje? (sea el aula u otro espacio en el que se esté desarrollando la sesión)		
4	Los /las estudiantes permanecen durante toda la sesión en el espacio de aprendizaje? (sea el aula u otro espacio en el que se esté desarrollando la sesión)		
5	Anotar si se dieron interrupciones al desarrollo de la sesión → En el caso de respuesta afirmativa, continuar. En caso contrario pasar a pregunta 7		
6	Anotar la cantidad de interrupciones		
7	El/la docente ha optimizado el desarrollo para actividades significativas de aprendizaje		

D. Clima de aula → Registrar estos datos a partir de la OBSERVACION

		Si	No
1	El/la docente dialoga y escucha con atención a los y las estudiantes?		
2	El/la docente se dirige a sus estudiantes por sus nombres sin utilizar apodosos o números?		
3	El/la docente trata a sus estudiantes con respeto, sin ridiculizarlos?		
4	El/la docente emplea palabras positivas para reafirmar el esfuerzo individual o grupal de los y las estudiantes?		
5	El/la docente recurre a normas y acuerdos que ayuden a mejorar la convivencia en el aula?		

F. Planificación → Datos a ser registrados CONSULTANDO al docente

		Si	No
1	<p>¿Las actividades planificadas para la sesión, incorporan los contenidos de las rutas de aprendizaje? = En caso de respuesta afirmativa, continuar. En caso contrario pasar la pregunta 2</p> <p>¿De qué manera?</p>		
2	<p>¿Las actividades desarrolladas en la sesión, están incluidas en algún documento de planificación? = En caso la respuesta afirmativa, continuar</p>		
3	<p>¿Cuál es? = N o leer las opciones</p>		
	a. Sesión de aprendizaje		
	b. Carpeta pedagógica		
	c. Unidades de aprendizaje		
	d. Proyecto		
	e. Diario de clase		
	f. Otros= Anotar en el siguiente espacio		
TOTAL			

III. CONCLUSIONES GENERALES PRODUCTO DE ASESORIA PERSONALIZADA

Fortalezas	Situaciones a mejorar

COMPROMISO

Firma del Docente

Firma del Observador

V°B° Director