



INSTITUTO TECNOLÓGICO SUPERIOR “CORDILLERA”

INSTITUTO TECNOLÓGICO SUPERIO “CORDILLERA”

CARRERA DE SISTEMAS



**LA DESERCIÓN ESTUDIANTIL Y LOS SISTEMAS DE INFORMACIÓN
GERENCIAL CON SOFTWARE DE MINERÍA DE DATOS Y
DATAWAREHOUSING PARA LA TOMA DE DECISIONES**

Proyecto de Trabajo de Graduación que se presenta como requisito para optar por
el título de Tecnólogo en Sistemas

Autor: BENITEZ Recalde, Daniel Alejandro

Director de Trabajo de Graduación: **Ing. Carlos Cadena**

Quito – Noviembre

2012



DECLARACIÓN DE AUTENTICIDAD

Declaro que los contenidos y los resultados obtenidos en el presente proyecto, como requerimiento previo para la obtención del Título de Tecnólogo Analista de Sistemas, son absolutamente originales, auténticos y personales y de exclusiva responsabilidad legal y académica del autor.

Daniel Alejandro Benitez Recalde

1717050171



AGRADECIMIENTO

A Dios, por iluminar mi camino cuando me sentía perdido, por haberme dado la sabiduría, conocimiento y madurez para cumplir mis objetivos y metas propuestas.

A mi familia, que han sido parte fundamental en mi formación como persona, hijo y padre de familia

A mis padres, me formaron como una persona responsable, solidaria, humilde y luchador.

A mi hijo, que fue, es y será siempre la motivación de luchar y seguir hacia adelante.

A mi esposa, apoyo incondicional donde puedo confiar en que siempre estará a mi lado,

Gracias



DEDICATORIA

A mi familia y Dios

Quiero dar las gracias a Dios porque cuando sentía que en mi camino no había luz, el me ilumino y me guio hacia mis metas.

Quiero dedicar este gran logro a toda mi familia, entre aciertos y tropiezos me han enseñado los valores de la vida, que las mejores cosas llegan con el esfuerzo constante, quiero dedicar unas palabras a mi hijo, quien es mi orgullo y motivación para siempre continuar hacia adelante, el es la razón de todos los días esforzarme y es mi satisfacción contar con su amor.

Daniel Benitez



INSTITUTO TECNOLÓGICO SUPERIOR “CORDILLERA”



INSTITUTO TECNOLÓGICO SUPERIOR “CORDILLERA”



INSTITUTO TECNOLÓGICO SUPERIOR “CORDILLERA”



ÍNDICE GENERAL

CARATULA	I
DECLARACIÓN DE AUTENTICIDAD	II
AGRADECIMIENTO	III
DEDICATORIA	IV
CERTIFICADO DE CONSEJO Y DIRECCIÓN DE CARRERA.....	V
APROBACIÓN DEL TRIBUNAL DE TRABAJO DE GRADUACIÓN	VI
CERTIFICADO DE ACEPTACIÓN DEL DIRECTOR DEL TRABAJO DE GRADUACIÓN	VII
ÍNDICE GENERAL	VIII
RESUMEN.....	XV
INTRODUCCION	1
CAPITULO I.....	3
1. EL PROBLEMA	3
1.1 PLANTEAMIENTO DEL PROBLEMA.....	3
1.1.1 CAUSAS Y CONSECUENCIAS	5
1.2 FORMULACION DEL PROBLEMA	5
1.3 OBJETIVOS	6
1.3.1 OBJETIVO GENERAL	6
1.3.2 OBJETIVOS ESPECIFICOS	6
1.4 JUSTIFICACION E IMPORTANCIA	7
CAPÍTULO II.....	8
2. MARCO TEORICO	8
2.1 ANTECEDENTE DEL ESTUDIO.....	8
2.2 FUNDAMENTACION TEORICA.....	9
2.2.1 BUSINESS INTELLIGENCE	9
2.2.2 PROCESO DE BUSINESS INTELLIGENCE.....	10
2.2.3 DATAWAREHOUSING	12
2.2.4 BASES DE DATOS MULTIDIMENSIONALES	12
2.2.4.1 CUBO MULTIDIMENSIONAL	13
2.2.4.2 TABLAS DE DIMENSIONES	16
2.2.4.3 TABLAS DE HECHOS	16
2.2.4.4 ESQUEMAS	17
2.2.4.4.1 ESQUEMA EN ESTRELLA.....	17
2.2.4.4.2 ESQUEMA EN COPO DE NIEVE.....	17
2.2.4.4.3 ESQUEMA EN CONSTELACION	17
2.2 GLOSARIO DE TERMINOS	18
2.3 FUNDAMENTACION LEGAL.....	20
2.4 CARACTERIZACION DE LAS VARIABLES.....	23
2.5 PREGUNTAS DIRECTRICES DE LA INVESTIGACION.....	24
CAPITULO III	25
3. METODOLOGIA	25



3.1.	TIPO DE INVESTIGACION.....	25
3.1.1.	BIBLIOGRAFICA DOCUMENTAL.....	25
3.1.2.	DESCRIPTIVA.....	25
3.1.3.	EXPLICATIVA.....	25
3.2.	METODOS DE INVESTIGACION.....	26
3.2.1.	HISTORICO LOGICO.....	26
3.2.2.	ANALITICO SINTETICO.....	26
3.2.3.	INDUCTIVO.....	26
3.2.4.	DEDUCTIVO.....	27
3.2.5.	HIPOTETICO DEDUCTIVO.....	27
3.3.	TECNICAS E INSTRUMENTOS.....	27
3.3.1.	OBSERVACION.....	27
3.3.2.	ENTREVISTA.....	28
3.3.3.	ENCUESTA.....	29
3.4.	POBLACION Y MUESTRA.....	30
3.5.	DISEÑO DE LA ENCUESTA.....	32
3.6.	CONCLUSION.....	34
3.7.	OPERACIONALIZACION DE LAS VARIABLES.....	34
	CAPITULO IV.....	35
4.	LA PROPUESTA.....	35
4.1	PROCESAMIENTO DE INFORMACION.....	35
4.1	RESPUESTAS A LAS INTERROGANTES DE LA INVESTIGACION.....	45
	CAPITULO V.....	47
5.	LA PROPUESTA.....	47
5.1.	CONCLUSIONES Y RECOMENDACIONES.....	47
	CAPITULO VI.....	50
6.	ASPECTOS ADMINISTRATIVOS.....	50
6.1.	ALCANCE.....	50
6.2.	ESTRUCTURA ORGANIZACIONAL DE LA EMPRESA.....	52
6.3.	ESTRUCTURA INFORMATICA.....	52
6.4.	ALTERNATIVAS.....	53
6.4.1.	ALTERNATIVAS.....	54
6.5.	SELECCIÓN DE ALTERNATIVAS.....	55
6.6.	PROCESOS.....	56
6.7.	DESCRIPCION DE LA METODOLOGIA DE DESARROLLO.....	67
6.7.1.	ANALISIS DE REQUERIMIENTOS.....	67
6.7.2.	ANALISIS DE SISTEMAS OLTP.....	70
6.7.3.	MODELO LOGICO DW.....	71
6.8.	MODELO LOGICO.....	75
6.9.	MODELO FISICO.....	75
6.10.	DICCIONARIO DE DATOS.....	76
6.11.	ESTANDARES.....	76
6.12.	PANTALLAS.....	76
6.13.	INSTALACION DEL SISTEMA.....	77
6.14.	RECOPIACION Y CARGA DE DATOS.....	80
6.15.	PRUEBAS Y DEPURACION.....	81



6.15.1.	PRUEBAS DE UNIDAD	81
6.15.2.	PRUEBAS DE INTEGRACION.....	81
6.15.3.	PRUEBAS DE VALIDACION.....	82
6.15.4.	PRUEBAS DEL SISTEMA	82
6.15.5.	PRUEBAS DE SEGURIDAD	82
6.15.6.	PRUEBAS DE INTERFAZ GRAFICA	82
6.16.	PUESTA EN MARCHA DEL SISTEMA.....	83
6.17.	CAPACITACION TECNICA.....	86
6.18.	CONCLUSIONES Y RECOMENDACIONES	87
6.19.	BIBLIOGRAFIA.....	88
ANEXOS.....		89



INDICE DE CUADROS

CUADRO N° 1	
Título: Causas y Consecuencias	16
CUADRO N° 2	
Título: Cubo Multidimensional	26
CUADRO N° 3	
Título: Variables e Indicadores	45
CUADRO N° 4	
Título: Tabulacion pregunta 1	46
CUADRO N° 5	
Título: Tabulacion pregunta 2	47
CUADRO N° 6	
Título: Tabulacion pregunta 3	48
CUADRO N° 7	
Título: Tabulacion pregunta 4	49
CUADRO N° 8	
Título: Tabulacion pregunta 5	50
CUADRO N° 9	
Título: Tabulacion pregunta 6	51
CUADRO N° 10	
Título: Tabulacion pregunta 7	52
CUADRO N° 11	
Título: Tabulacion pregunta 8	53
CUADRO N° 12	
Título: Tabulacion pregunta 9	54
CUADRO N° 13	
Título: Alternativas	63
CUADRO N° 14	
Título: Pregunta 6	26
CUADRO N° 15	



Título: Pregunta 7	26
CUADRO N° 16	
Título: Pregunta 4	25
CUADRO N° 5	
Título: Pregunta 5	25
CUADRO N° 6	
Título: Pregunta 6	26
CUADRO N° 7	
Título: Pregunta 7	26
CUADRO N° 6	
Título: Pregunta 6	26
CUADRO N° 7	
Título: Pregunta 7	26
CUADRO N° 4	
Título: Pregunta 4	25
CUADRO N° 6	
Título: Pregunta 6	26
CUADRO N° 7	
Título: Pregunta 7	26
CUADRO N° 4	
Título: Pregunta 4	25



INDICE DE GRAFICOS

GRÁFICO N° 1	
Título: Pregunta 1	35
GRÁFICO N° 2	
Título: Pregunta 2	37
GRÁFICO N° 3	
Título: Pregunta 3	38
GRÁFICO N° 4	
Título: Pregunta 4	39
GRÁFICO N° 5	
Título: Pregunta 5	40
GRÁFICO N° 6	
Título: Pregunta 6	41
GRÁFICO N° 7	
Título: Pregunta 7	42
GRÁFICO N° 8	
Título: Pregunta 8	43
GRÁFICO N° 9	
Título: Pregunta 9	44
GRÁFICO N° 10	
Título: Modelo del Sistema	77
GRÁFICO N°11	
Título: Ventana detalla la instalación de Base de Datos	78



GRÁFICO N°12

Título: Paquete de

Flujo de Datos

80



INSTITUTO TECNOLÓGICO SUPERIOR “CORDILLERA”

**“LA DESERCIÓN ESTUDIANTIL Y LOS SISTEMAS DE INFORMACIÓN
GERENCIAL CON SOFTWARE DE MINERÍA DE DATOS Y
DATAWAREHOUSING PARA LA TOMA DE DECISIONES”**

Autor/a: Daniel Alejandro Benítez Recalde

Director del Trabajo de Graduación: Ing. Hugo Heredia

RESUMEN

La deserción estudiantil se conoce al proceso no normal de un estudiante en su trayectoria estudiantil, esta deserción puede ser provocada por distintos motivos u factores personales, socio-económicos y académicos, es por ello que las técnicas de Datawarehouse y minería de datos son de fundamental importancia en el proceso de generar conocimiento para la toma de decisiones. El presente documento tiene un propósito estudiantil, pretende demostrar a través de un modelo de Minería de Datos (Data Mining) conocer los factores más importantes que determinan a un estudiante a abandonar sus estudios, tomando como base de conocimiento los patrones, características similares que existan entre los estudiantes que han desertado y poder predecir quienes se encuentran con alto riesgo y permitir tomar decisiones acertadas y preventivas al caso. Como input del modelo se propone desarrollar un sistema almacén de datos (Datawarehouse) que integrara información tomada de múltiples fuentes de datos consolidando información del estudiante limpia, confiable y centralizada. La construcción del Almacén de datos se desarrollo con la metodología de libre difusión “Hefesto”.

INTRODUCCION

El business intelligence (BI) se define como un conjunto de técnicas, metodologías, estrategias y sistemas de información que trabajan entre si, con la finalidad de descubrir conocimiento oculto entre grandes cantidades de datos que se generan en el proceso transaccional de las organizaciones. Tiene como objetivo transformar los datos en información, y la información en conocimiento, y en base al conocimiento generar estrategias para la toma de decisiones, aplicando el concepto podemos aplicarlo a la deserción estudiantil analizando la información que genera el estudiante en su trayectoria estudiantil dentro de la escuela, desarrollando un modelo de Datawarehouse medir el rendimiento en forma de indicadores, como porcentaje de deserción, rendimiento estudiantil, etc.

- **CAPITULO UNO**

El objetivo principal que el investigador debe cumplir con el desarrollo de un modelo de minería de datos, es determinar factores de deserción que inciden en los estudiantes para tomar la decisión de abandonar sus estudios.

- **CAPITULO DOS**

El Proyecto basa sus conceptos en la inteligencia empresarial, mas conocida como inteligencia de negocios, que es el conjunto de herramientas, tecnologías, sistemas, y estrategias enfocadas en un solo objetivo, que es transformar datos en información útil para la toma de sediciones.

- **CAPITULO TRES**

El proyecto se desarrollo en función a la metodología de distribución libre, denominada HEFESTO en la construcción del almacén de datos (Datawarehouse) y para la construcción del modelo de minería de datos se utilizo la metodología CRISP-DM



- **CAPITULO CUATRO**

Describe el procesamiento de la información esto tiene que ver con los cuadros estadísticos para poder contabilizar cada una de las preguntas hechas en las encuestas.

- **CAPITULO QUINTO**

En la finalización del estudio, en esta primera etapa se pudo analizar el rendimiento académico de los estudiantes, para determinar si es un factor importante, y el porcentaje de deserción estudiantil diferenciado por año o semestre escolar.

- **CAPITULO SEXTO**

Describe la propuesta de nuestro proyecto todo lo que tiene que ver con el desarrollo del software y sus diferentes metodologías utilizadas para la realización del mismo.



CAPITULO I

1. EL PROBLEMA

1.1 PLANTEAMIENTO DEL PROBLEMA

La deserción estudiantil se puede definir como aquel estudiante que inicia una trayectoria universitaria, y este pasa por uno o más de los siguientes estados:

- Problemas al terminar sus estudios (arrastrés, retiros temporales, rendimiento bajo)
- Cambio de carrera universitaria
- Abandono total de la carrera

Que pueden ser causados por distintos motivos o factores:

- Factores personales
- Factores Institucionales y pedagógicos
- Factores socioeconómicos

El problema de una deserción estudiantil sea parcial o total abandono, se vuelve un tema preocupante e importante para las instituciones de educación superior, gobierno y al país en general que están condicionando el desarrollo normal y la ampliación de la educación y profesionales especializados en el país. Todo esto ocurre debido a un desconocimiento sobre los estudiantes con características, factores y patrones de deserción, que motivan a los estudiantes a abandonar parcial o totalmente sus estudios

Actualmente en el país no existen cifras, estadísticas, ni estudios acerca de la deserción estudiantil por parte de ministerios u autoridades, existen estudios aislados que han sido desarrollados en algunas organizaciones de educación superior por parte de los mismos estudiantes que forman o fueron parte de la escuela, lo cual agrava mas la situación del país y su desarrollo como tal. Citamos a un nivel general de América Latina, un estudio realizado por la UNESCO entre el año 2000 y 2005 que señala que



pierde hasta \$420 millones por la deserción estudiantil, observando un desaprovechamiento de recursos para el país.

En el INEC (Instituto Nacional Estadísticas y Censos), Ecuador en cifra tampoco presentan datos estadísticos, ni estudios que pongan énfasis a esta problemática, presentando así un desconocimiento y una despreocupación, por instituciones que periódicamente deberían hacer estudios para tener un conocimiento complejo de que cantidad de estudiantes desertan de sus estudios, y que causas los llevan a tomar este tipo de decisiones.

Proyectos de investigación en cuanto a la deserción estudiantil, es un tema relevante que se debe analizar profundamente y críticamente para las instituciones de educación superior, el gobierno y al país en general, es un tema al que menos interés e importancia y recursos se los da, derivándose en un desconocimiento alarmante del estado de los estudiantes y que calidad de estudiantes están siendo graduados y titulados.

Por otro lado la gestión de la información es otra derivación de la poca importancia que se les da a este tipo de proyectos de investigación, generando un proceso complicado de medir, y difícil de obtener resultados verídicos de un negocio y más específicamente de los estudiantes, para la toma de decisiones, una cita importante nos dice: “Lo que no se mide no se puede controlar, y lo que no se controla no se puede gestionar”¹

Tomando en cuenta que cada día las organizaciones de educación superior, generan grandes cantidades de datos, entre notas, materias, niveles, etc. se vuelven datos inmedibles a través de técnicas antiguas que proporcionan muy poco conocimiento eficaz que la evolución de la educación superior lo demanda.

La toma de decisiones, es un proceso muy importante para cualquier negocio, considerando que cualquier desacierto, es decir una mala decisión conlleva a pérdidas incalculables y no renovables de recursos, entre los más importantes tiempo y dinero; No es suficiente la experiencia y basar decisiones de un negocio en la intuición y al no tener la información necesaria y las tecnologías adecuadas, se vuelve un proceso

¹ Frase: Ricardo Hirata



incierto, inseguro e inmedible para el usuario experto del negocio provocando vulnerabilidades en la organización.

“El objetivo principal de la educación es crear personas capaces de hacer cosas nuevas, y no simplemente repetir lo que otras generaciones hicieron”²

1.1.1 CAUSAS Y CONSECUENCIAS

Cuadro No.1 Causas y Consecuencias

Rendimiento Bajo	Produce desmotivación en el estudiante, y puede perder un interés total o parcial de la carrera.
Por falta de Recursos	Deserción de la carrera por un periodo o un abandono total.
Indisponibilidad de Tiempo	Poco tiempo para realizar investigaciones, consultas y tareas mal realizadas afectando su rendimiento.
Falta de Motivación	Producen un poco interés por las materias y poco interés en aprender

Autor: Daniel Benítez

1.2 FORMULACION DEL PROBLEMA

El indicador deserción estudiantil es uno de los más importantes de gestión para una organización estudiantil y es donde se centra el problema, en los estudiantes que no pueden culminar sus estudios y terminan abandonando la carrera universitaria parcial o totalmente, es por ello que se plantea la siguiente cuestión. ¿El sistema Datawarehouse (Almacén de Datos) podrá generar conocimientos referentes a los perfiles de los estudiantes con características de deserción y prevenir un abandono de una carrera universitaria para así culminar con la toma de decisiones?

Mediante la herramienta de gestión del sistema Datawarehouse (Almacén de Datos), el mismo que nos permitirá conocer los posibles perfiles de deserción estudiantil en la

² Frase: Jean Piaget



carrera de sistemas, y nos ayudara a prevenir el abandono de la misma y tomar medidas de prevención.

1.3 OBJETIVOS

1.3.1 OBJETIVO GENERAL

Demostrar posibles factores personales, socio-económicos, y académicos de deserción universitaria que motivan a los estudiantes a abandonar sus estudios superiores, y determinar perfiles de estudiantes con características de deserción con la finalidad de predecir y prevenir un retiro parcial o total, aplicando técnicas de un Datawarehouse.

1.3.2 OBJETIVOS ESPECIFICOS

1. Construir un datawarehouse que almacene información clara y transparente de los estudiantes y sus características más importantes para su análisis.
2. Desarrollar procesos de integración que transfiera los datos de un ambiente transaccional a un ambiente de analítico dentro del datawarehouse.
3. Desarrollar Cubos de información que permitan el análisis e interpretación de los indicadores
4. Identificar los indicadores más importantes que permitan cuantificar la deserción estudiantil de carreras universitarias y ayuden a identificar las variables de deserción académica
5. Brindar conocimiento que permita generar estrategias organizacionales de educación y reduzcan los porcentajes de deserción académica.



1.4 JUSTIFICACION E IMPORTANCIA

La información investigada hasta el momento, nos reflejan un desconocimiento y desinterés tanto de instituciones de educación superior como ministerios, SENESCYT (Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación), la misma que se lo se enfoca en estudios de infraestructura, pensum de estudio, docentes, pero no hace un estudio minucioso de cuantos estudiantes son posibles desertores de sus estudios y los factores que les llevaron a tomar una decisión tan apresurada y sin pensarla.

Toda organización indistintamente según la actividad económica a la que se dedique, siempre están produciendo grandes cantidades de datos, datos que contienen información que es de gran importancia para la toma de decisiones, es por ello que la inteligencia de negocios surge como un apoyo a la toma de decisiones y facilitar al usuario experto del negocio a concentrar sus esfuerzos en las mejores estrategias tomando como referencia la información proveniente de este conjunto de sistemas de inteligencia de negocios.

Esta solución de inteligencia de negocios, pretende mejorar la gestión del proceso de toma de decisiones en el ámbito estudiantil, como lo es la deserción estudiantil, pretende determinar los posibles factores que influyen en los estudiantes, para que ellos tengan problemas de terminar sus estudios, y decidan abandonar su carrera universitaria, es por ello que a través de inteligencia de negocios, se pretende demostrar paulatinamente a través del desarrollo de un sistema datawarehouse (Almacén de Datos) para la recolección de la información y que permitirá el análisis de la información mediante indicadores los mismos que determinara posibles patrones que existen dentro del conjunto de estudiantes y así poder prevenir una deserción estudiantil.



CAPÍTULO II

2. MARCO TEORICO

2.1 ANTECEDENTE DEL ESTUDIO

El presente proyecto tuvo su origen a partir de investigaciones en internet realizadas por parte del autor intelectual del proyecto, en países desarrollados en el ámbito de la informática, el estudio al que se relaciona es un proyecto desarrollado en la Universidad Nacional de Misiones

- Titulo -- Resultados Preliminares del Proceso de Minería de Datos Aplicado al Análisis de la Deserción Utilizando Herramientas Open Source³
- Autor -- Lic. J. Germán A. Pautsch

El tema de la investigación recae en el mismo ámbito estudiantil con la diferencia que se utilizaran diferentes criterios de análisis a la información con un diferente objetivo y enfoque, con diferentes resultados.

³ Software Libre que es una herramienta de gestión de información que no tiene costo.



2.2. FUNDAMENTACION TEORICA

2.2.1 BUSINESS INTELLIGENCE

Inteligencia de Negocios se puede definir como un conjunto de técnicas, estrategias, metodologías, y sistemas que interactúan entre si con la finalidad de descubrir conocimiento oculto entre grandes cantidades de datos que generan las organizaciones en sus procesos transaccionales que coadyuvan al proceso de la toma de decisiones

El concepto de la inteligencia de negocios es transformar los datos en información, y la información en conocimiento, y en base al conocimiento generar estrategias para la toma de decisiones.

La inteligencia de negocios fue creada para satisfacer preguntas claves de negocio, el comportamiento de un negocio y luego responder a esas grandes interrogantes, de porque esta sucediendo y cuales son los motivos del comportamiento.

Ventajas:

- Tener información clara, transparente, limpia, confiable y centralizada tomada de múltiples fuentes de datos
- Realizar un análisis y exploración de grandes cantidades de datos en tiempo real sobre el comportamiento del negocio
- Permite predecir comportamientos futuros, aprendiendo de hechos sucedidos del pasado
- Genera conocimiento útil para el proceso de toma de decisiones
- Entregar la Información necesaria a las personas indicadas en el momento indicado



2.2.2 PROCESO DE BUSINESS INTELLIGENCE

- Requerimientos o necesidades de información

En este primer paso, es donde se realiza un análisis de requerimientos de usuario a los interesados e involucrados del negocio sobre un área o proceso específico seleccionado, es decir el investigador define en conjunto con el usuario las preguntas clave de negocio que necesitan conocer y que información necesitan para la toma de decisiones, en este punto el investigador puede valerse de técnicas de recolección de información u aplicar la experiencia que le ha dado el tiempo y trabajo.

- Análisis y Recolección de los datos

Este ítem se puede entender como el análisis que realiza el investigador sobre las múltiples fuentes de datos que pueda tener la organización, sean internas, externas, textos planos, bases de datos, etc. Para satisfacer los requerimientos de información del usuario, es aquí donde se selecciona los datos necesarios para dar respuesta a las preguntas clave de negocio definidas con el usuario.



- Integración y carga de datos

En esta fase es donde los datos seleccionados son cargados a un único almacén de datos que integre información en un único formato, limpia de datos innecesarios, y confiable.

Generalmente este proceso es el mas importante, complejo y que mas tiempo consume, ya que las fuentes de datos pueden variar según el negocio y que tan limpia se encuentre la información, es decir que no existan datos nulos, vacíos, o corruptos.

- Análisis de la información

Aquí es donde se genera el conocimiento y se responden a las interrogantes planteadas, es aquí donde el usuario interpreta la información en forma de indicadores, tablas, gráficos, es decir información estadística que le sirvan de orientación sobre las interrogantes y coadyuven a generar estrategias y al proceso de toma de decisiones

- Difusión

En este punto es donde se distribuye la información y conocimiento a los respectivos usuarios que lo necesitan, proveyendo de las respectivas herramientas para el análisis.



2.2.3 DATAWAREHOUSING

Se conoce como una técnica de la inteligencia de negocios y forma parte de un conjunto de tecnologías y procesos que se encargan de recolectar, limpiar, integrar, y centralizar información histórica de una organización en una única fuente de datos y que sea almacenada en una estructura multidimensional que permita un análisis dinámico con la finalidad de ayudar al proceso de la toma de decisiones.

Permite extraer, recolectar los datos que sean necesarios para la investigación, sin importar que se encuentren en diversos sistemas operacionales, fuentes externas y puedan encontrarse con múltiples formatos, datos que genera la propia organización en su actividad diaria.

Los datos recolectados pasan a un proceso de limpieza, excluyen do datos que no se necesitan en el análisis o se encuentran con inconsistencias y transformarlos para consolidar la información en un solo formato y forma única para el almacén de datos

Teniendo información confiable, podemos integrar y centralizar la información en un único almacén de datos para su futura exploración y análisis de la información⁴

2.2.4 BASES DE DATOS MULTIDIMENSIONALES

Las bases de datos multidimensionales constituyen la forma de almacenamiento de los Datawarehouse (Almacén de Datos) facilitando una forma flexible de acceso a los datos almacenados, los cuales se van a localizar en diferentes dimensiones, visualizándose como un cubo multidimensional, en donde las variables asociadas existen a lo largo de varios ejes o dimensiones, y la intersección de las mismas representa la medida, indicador o el hecho que se está evaluando. Las bases de datos multidimensionales implican tres variantes posibles de modelamiento, que permiten realizar consultas de soporte de decisión:

⁴ Datawarehosing.- Nos ayuda a cumplir con sus objetivos, extrae, transforma y consolida, sea almacenada de manera centralizada en una base de datos. (HEFESTO, 2010).



- Esquema en estrella.
- Esquema copo de nieve.
- Esquema constelación.

La información es almacenada de forma multidimensional, conformada por tablas de hechos y tablas de dimensiones, y basan su modelamiento en tres tipos de esquemas: Estrella, Copo de nieve, Constelación que permiten estructurar la información de gestión con un acceso flexible y coherente, para ejecutar las complejas consultas de los usuarios y satisfagan sus necesidades de información.

2.2.4.1 CUBO MULTIDIMENSIONAL

Si bien existen diversas estructuras de datos, a través de las cuales se puede representar los datos del Datawarehouse (Almacén de Datos), solamente se entrará en detalle acerca de los cubos multidimensionales, por considerarse que esta estructura de datos es una de las más utilizadas y cuyo funcionamiento es el más complejo de entender.

Un cubo multidimensional o hipercubo, representa o convierte los datos planos que se encuentran en filas y columnas, en una matriz de N dimensiones. Los objetos más importantes que se pueden incluir en un cubo multidimensional, son los siguientes:

Indicadores: Sumarizaciones que se efectúan sobre algún hecho o expresiones basadas en sumarizaciones, pertenecientes a una tabla de hechos.

Los indicadores son sumarizaciones efectuadas sobre algún hecho o expresiones basadas en sumarizaciones, que serán incluidos en algún cubo multidimensional, con el fin de analizar los datos almacenados en el DW. El valor que estos adopten estará condicionado por los atributos/jerarquías que se utilicen para analizarlos.



Los indicadores, además de hechos, pueden estar compuestos por otros indicadores, pero no ambos simultáneamente. Pueden utilizarse para su creación funciones de sumarización (suma, conteo, promedio, etc), funciones matemáticas, estadísticas, operadores matemáticos y lógicos.

Atributos: campos o criterios de análisis, pertenecientes a tablas de dimensiones.

Los atributos constituyen los criterios de análisis que se utilizarán para analizar los indicadores dentro de un cubo multidimensional. Los mismos se basan, en su gran mayoría, en los campos de las tablas de dimensiones y/o expresiones.

Dentro de un cubo multidimensional, los atributos son los ejes del mismo.

Jerarquías: Una jerarquía representa una relación lógica entre dos o más atributos pertenecientes a un cubo multidimensional; siempre y cuando posean su correspondiente relación “padre-hijo”.

Las jerarquías poseen las siguientes características:

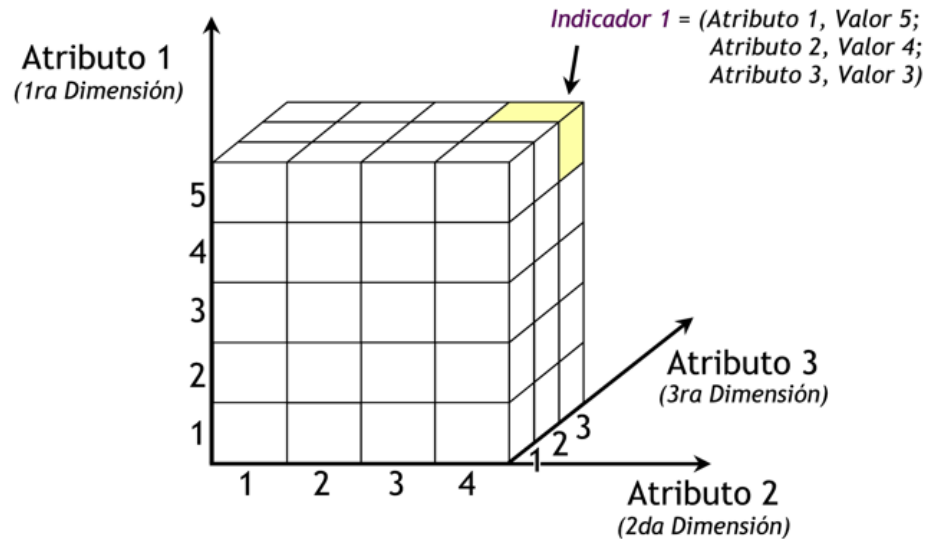
- Pueden existir varias en un mismo cubo.
- Están compuestas por dos o más niveles.
- Se tiene una relación “1-n” o “padre-hijo” entre atributos consecutivos de un nivel superior y uno inferior.

Por lo general, las jerarquías pueden identificarse fácilmente, debido a que existen relaciones “1-n” o “padre-hijo” entre los propios atributos de un cubo.

La principal ventaja de manejar jerarquías, reside en poder analizar los datos desde su nivel más general al más detallado y viceversa, al desplazarse por los diferentes niveles.

Representa una relación lógica entre dos o más atributos. De esta manera en un cubo multidimensional, los atributos existen a lo largo de varios ejes o dimensiones, y la intersección de las mismas representa el valor que tomará el indicador que se está evaluando.

En la siguiente representación matricial se puede ver más claramente lo que se acaba de decir:

Cuadro No.2 Cubo Multidimensional

Autor: Ing. RIZZI, Mauricio- Sistemas Avanzados de Base de Datos con Soporte para la Toma de Decisiones- Editorial Universidad Católica de Córdoba- 2006.

Para la creación del cubo de la figura anterior, se definieron tres Atributos (“Atributo 1”, “Atributo 2” y “Atributo 3”) y se definió un Indicador (“Indicador 1”). Entonces el cubo quedó compuesto por 3 dimensiones o ejes (una por cada Atributo), cada una con sus respectivos valores asociados. También, se ha seleccionado una intersección al azar para demostrar la correspondencia con los valores de las Atributos. En este caso, el indicador “Indicador 1”, representa el cruce del Valor “5” de “Atributo 1”, con el Valor “4” de “Atributo 2” y con el Valor “3” de “Atributo 3”.

gSe puede observar, que el resultado del análisis está dado por los cruces matriciales de acuerdo a los valores de las dimensiones seleccionadas.

Más específicamente, para acceder a los datos del DW, se pueden ejecutar consultas sobre algún cubo multidimensional previamente definido. Dicho cubo debe incluir entre otros objetos: indicadores, atributos, jerarquías, etc, basados en los campos de las tablas de dimensiones y de hechos, que se deseen analizar. De esta manera, las consultas son respondidas con gran performance, minimizando al máximo el tiempo



que se hubiese incurrido en realizar dicha consulta sobre una base de datos transaccional.

2.2.4.2 TABLAS DE DIMENSIONES

Las dimensiones representan los ejes del cubo multidimensional y son las perspectivas por las cuales se desea analizar un indicador, poseen características cualitativas, es decir describen cualidades, le dan criterio al objeto analizado o indicador.

2.2.4.3 TABLAS DE HECHOS

Las tablas de hechos contienen, precisamente, los hechos que serán utilizados por los analistas de negocio para apoyar el proceso de toma de decisiones. Contienen datos cuantitativos.

Los hechos son datos instantáneos en el tiempo, que son filtrados, agrupados y explorados a través de condiciones definidas en las tablas de dimensiones.

Los datos presentes en las tablas de hechos constituyen el volumen de la bodega, y pueden estar compuestos por millones de registros dependiendo de su granularidad y antigüedad de la organización. Los más importantes son los de tipo numérico.

El registro del hecho posee una clave primaria que está compuesta por las claves primarias de las tablas de dimensiones relacionadas a este.



2.2.4.4 ESQUEMAS

2.2.4.4.1 ESQUEMA EN ESTRELLA

El esquema en estrella, consta de una tabla de hechos central y de varias tablas de dimensiones relacionadas a esta, a través de sus respectivas claves.

Este modelo debe estar totalmente desnormalizado, es decir que no puede presentarse en “tercera forma normal”.

2.2.4.4.2 ESQUEMA EN COPO DE NIEVE

Este esquema representa una extensión del modelo en estrella cuando las tablas de dimensiones se organizan en jerarquías de dimensiones, existe una tabla de hechos central que está relacionada con una o más tablas de dimensiones, quienes a su vez pueden estar relacionadas o no con una o más tablas de dimensiones.

Este modelo es más cercano a un modelo de entidad relación, que al modelo en estrella, debido a que sus tablas de dimensiones están normalizadas.

2.2.4.4.3 ESQUEMA EN CONSTELACION

Este modelo está compuesto por una serie de esquemas en estrella, está formado por una tabla de hechos principal (“HECHOS_A”) y por una o más tablas de hechos auxiliares (“HECHOS_B”), las cuales pueden ser sumalizaciones de la principal. Dichas tablas yacen en el centro del modelo y están relacionadas con sus respectivas tablas de dimensiones.

No es necesario que las diferentes tablas de hechos compartan las mismas tablas de dimensiones, ya que, las tablas de hechos auxiliares pueden vincularse con solo algunas de las tablas de dimensiones asignadas a la tabla de hechos principal, y también pueden hacerlo con nuevas tablas de dimensiones.



2.2 GLOSARIO DE TERMINOS

Almacén de datos.- Representa el equivalente a un Datawarehouse, y es el repositorio o bases de datos que contiene la información a ser analizada.⁵

Drill down.- Permite partir de lo general a lo específico y apreciar los datos en un mayor detalle, bajando por una jerarquía definida en un cubo. Esto brinda la posibilidad de introducir un nuevo nivel o criterio de agregación en el análisis, disgregando los grupos actuales.⁶

Drill up.- Permite partir de lo particular a lo general y apreciar los datos en menor nivel de detalle, subiendo por una jerarquía definida en un cubo. Esto brinda la posibilidad de quitar un nivel o criterio de agregación en el análisis, agregando los grupos actuales.⁷

Drill Across.- Funciona de forma similar a drill-down, con la diferencia de que drill-across no se realiza sobre una jerarquía, sino que su forma de ir de lo general a lo específico es agregar un atributo a la consulta como nuevo criterio de análisis.⁸

Roll Across.- Funciona de forma similar a drill-up, con la diferencia de que roll-across no se hace sobre una jerarquía, sino que su forma de ir de lo específico a lo general es quitar un atributo de la consulta, eliminando de esta manera un criterio de análisis.⁹

⁵ RAMOS, Salvador - Vea el cubo medio lleno – Editorial SolidQ 2011

⁶ Ing. BERNABEU, Dario – Metodología para la construcción de un Datawarehouse – Editorial Instituto Universitario Aeronautico 2010

⁷ RAMOS, Salvador - Vea el cubo medio lleno – Editorial SolidQ 2011

⁸ <http://wikipedia.org>] Wikipedia.

⁹ <http://sistemasdecisionales.blogspot.com>] Sistemas Decisionales, algo más que Business Intelligence – Jorge Fernández González



Pivot.-Permite seleccionar el orden de visualización de los atributos e indicadores, con el objetivo de analizar la información desde diferentes perspectivas.¹⁰

Page.-Presenta el cubo dividido en secciones, a través de los valores de un atributo, como si se tratase de páginas de un libro¹¹

Drill Trought.- Permite apreciar los datos en su máximo nivel de detalle. Esto brinda la posibilidad de analizar cuáles son los datos relacionados al valor de un Indicador, que se ha sumariado dentro del cubo multidimensional.¹²

Hipercubo.- definición que representa el cubo que contiene la información¹³

¹⁰ [<http://todobi.blogspot.com>] Todo BI: Business Intelligence, Data Warehouse, CRM y mucho más....

¹¹ [<http://www.intelineg.com>] Inteligencia de Negocio – Rémi Grossat.

¹² KEPER, Charles – Análisis de Toma de Decisiones Datawarehouse – Editorial McGraw – Hill Años 2010.

¹³ [<http://todobi.blogspot.com>] Todo BI: Business Intelligence, Data Warehouse, CRM y mucho más....



2.3 FUNDAMENTACION LEGAL

La ley de Propiedad Intelectual

Sección II

Objeto del Derecho de Autor

Art. 8. La protección del derecho de autor recae sobre todas las obras del ingenio, en el ámbito literario o artístico, cualquiera que sea su género, forma de expresión, mérito o finalidad. Los derechos reconocidos por el presente Título son independientes de la propiedad del objeto material en el cual está incorporada la obra y su goce o ejercicio no están supeditados al requisito del registro o al cumplimiento de cualquier otra formalidad.

Las obras protegidas comprenden, entre otras, las siguientes:

Libros, folletos, impresos, epistolarios, artículos, novelas, cuentos, poemas, crónicas, críticas, ensayos, misivas, guiones para teatro, cinematografía, televisión, conferencias, discursos, lecciones, sermones, alegatos en derecho, memorias y otras obras de similar naturaleza, expresadas en cualquier forma;

Colecciones de obras, tales como antologías o compilaciones y bases de datos de toda clase, que por la selección o disposición de las materias constituyan creaciones intelectuales, sin perjuicio de los derechos de autor que subsistan sobre los materiales o datos;

Obras dramáticas y dramático musicales, las coreografías, las pantomimas y, en general las obras teatrales;



INSTITUTO TECNOLÓGICO SUPERIOR CORDILLERA

Composiciones musicales con o sin letra;

Obras cinematográficas y cualesquiera otras obras audiovisuales;

Las esculturas y las obras de pintura, dibujo, grabado, litografía y las historietas gráficas, tebeos, comics, así como sus ensayos o bocetos y las demás obras plásticas;

Proyectos, planos, maquetas y diseños de obras arquitectónicas y de ingeniería;

Ilustraciones, gráficos, mapas y diseños relativos a la geografía, la topografía, y en general a la ciencia;

Obras fotográficas y las expresadas por procedimientos análogos a la fotografía;

Obras de arte aplicada, aunque su valor artístico no pueda ser disociado del carácter industrial de los objetos a los cuales estén incorporadas;

Programas de ordenador; y,

Adaptaciones, traducciones, arreglos, revisiones, actualizaciones y anotaciones; compendios, resúmenes y extractos; y, otras transformaciones de una obra, realizadas con expresa autorización de los autores de las obras originales, y sin perjuicio de sus derechos.

Sin perjuicio de los derechos de propiedad industrial, los títulos de programas y noticieros radiales o televisados, de diarios, revistas y otras publicaciones periódicas, quedan protegidos durante un año después de la salida del último número o de la comunicación pública del último programa, salvo que se trate de publicaciones o



producciones anuales, en cuyo caso el plazo de protección se extenderá a tres años.

Sección V

Disposiciones Especiales sobre ciertas Obras

Parágrafo Primero

De los Programas de Ordenador

Art. 28. Los programas de ordenador se consideran obras literarias y se protegen como tales. Dicha protección se otorga independientemente de que hayan sido incorporados en un ordenador y cualquiera sea la forma en que estén expresados, ya sea en forma legible por el hombre (código fuente) o en forma legible por máquina (código objeto), ya sean programas operativos y programas aplicativos, incluyendo diagramas de flujo, planos, manuales de uso, y en general, aquellos elementos que conformen la estructura, secuencia y organización del programa.

Art. 30. La adquisición de un ejemplar de un programa de ordenador que haya circulado lícitamente, autoriza a su propietario a realizar exclusivamente:

Una copia de la versión del programa legible por máquina (código objeto) con fines de seguridad o resguardo;

Fijar el programa en la memoria interna del aparato, ya sea que dicha fijación desaparezca o no al apagarlo, con el único fin y en la medida necesaria para utilizar el programa; y,



Salvo prohibición expresa, adaptar el programa para su exclusivo uso personal, siempre que se limite al uso normal previsto en la licencia. El adquirente no podrá transferir a ningún título el soporte que contenga el programa así adaptado, ni podrá utilizarlo de ninguna otra forma sin autorización expresa, según las reglas generales.

Se requerirá de autorización del titular de los derechos para cualquier otra utilización, inclusive la reproducción para fines de uso personal o el aprovechamiento del programa por varias personas, a través de redes u otros sistemas análogos, conocidos o por conocerse.¹⁴

2.4 CARACTERIZACION DE LAS VARIABLES

Variable Independiente.- La deserción estudiantil

La deserción estudiantil es una variable incontrolable para el investigador, ya que los estudiantes, hablando en función al tiempo en pasado, y en el presente están abandonando sus estudios.

Variable Dependiente.- Software de Datawarehousing.

Es una variable que depende del investigador ya que el desarrollo del sistema es responsabilidad del desarrollador

¹⁴ Ley de Propiedad Intelectual Ecuatoriana



2.5 PREGUNTAS DIRECTRICES DE LA INVESTIGACION

Dentro de la investigación, han surgido varias interrogantes que en función al tiempo y desarrollo de la solución, se irán despejando, en las cuales tenemos las siguientes

1. Porque es importante el desarrollo de un datawarehouse?
2. Que tipo de estrategias tienen las organización de educación superior en cuanto a la deserción estudiantil
3. Es importante conocer el rendimiento de los estudiantes dentro de su carrera universitaria
4. Cuales son los indicadores mas importantes que se deben investigar para generar conocimiento
5. Es una limitante en la investigación no contar con una base de datos real de una institución de educación superior



CAPITULO III

3. METODOLOGIA

3.1. TIPO DE INVESTIGACION

3.1.1. BIBLIOGRAFICA DOCUMENTAL

Dentro del tema de proyecto, esta técnica permite recolectar, seleccionar y analizar la información que se encuentran en documentos de internet, consultando autores, foros, libros, que se relacionan con el objeto de análisis, permitiendo tener una visión y orientación al investigador sobre la temática a resolver.

3.1.2. DESCRIPTIVA

La técnica de investigación descriptiva, permite al investigador conocer características sobre el objeto de análisis, y definir similitudes que existen entre la muestra que se encuentra estudiando. Aplicando el concepto al objeto de estudio, nos permitió determinar y definir los factores que inciden en un estudiante para abandonar sus estudios

3.1.3. EXPLICATIVA

Nos dio una orientación y nos despejó dudas del porque los estudiantes abandonan sus estudios de carreras universitarias, permitiendo al investigador conocer los factores que están motivando a un estudiante a abandonar sus estudios y concentrar sus esfuerzos en resolución la problemática.



3.2. METODOS DE INVESTIGACION

3.2.1. HISTORICO LOGICO

El método histórico lógico es muy aplicable en el presente proyecto, ya que nos permite estudiar casos que se han dado en el pasado y tienen relación

con la problemática que se encuentra en estudio, conocer el proceso que se realiza para llevar acabo y como lo podemos mejorar en el presente, conociendo otros aspectos y como fueron llevados a cabo para su resolución, permitiendo al investigador tener mas alternativas.

3.2.2. ANALITICO SINTETICO

Este método permite tomar el objeto de investigación, descomponer la problemática en partes y realizar un análisis por parte del investigador, de cada uno de sus elementos, dando una solución eficaz a la problemática, sintetizando la información que es realmente relevante estudiar y explotando al máximo los recursos.

3.2.3. INDUCTIVO

Es un tipo de razonamiento lógico que nos permite partir de lo particular a lo general o partir de lo simple a lo complejo, es decir partir del proceso mas simple que es conocer el rendimiento de un estudiante hasta llegar a resolver el problemas macro que es la deserción de un estudiante de una carrera universitaria.



3.2.4. DEDUCTIVO

De la misma forma significa partir del macro del problema y derivarlo en partes, es decir partir de lo complejo a lo simple, aplicando este método al negocio podemos definir que parte de la deserción del estudiante hasta determinar las causas o cuales son los factores que lo motivan a abandonar sus estudios y conocer sus características para un futuro poder predecir estudiantes con características de deserción.

3.2.5. HIPOTETICO DEDUCTIVO

Este método permite al investigador realizar una observación del objeto de análisis, determinar un juicio de valor o hipótesis que posteriormente será comprobada su veracidad o falsedad, aplicando el concepto al negocio, podemos definir que se obtuvo los factores que inciden en un estudiante, y sometiéndolos a verificaciones podremos conocer si son los reales.

3.3. TECNICAS E INSTRUMENTOS

3.3.1. OBSERVACION

Gracias a las características de este tipo de técnica, podemos determinar las características de un estudiante desertor y obtener datos preliminares que servirán de orientación al investigador hacia donde profundizar, a través de una inspección detallada de lo que se vive día a día en las aulas de clases y experiencias propias por parte del investigador, podemos determinar factores que inciden en el abandono de la carrera universitaria.



3.3.2. ENTREVISTA

Podemos obtener características más acertadas a la realidad, a través de entrevistas a estudiantes, maestros, y expertos tanto del negocio como en la materia, podremos conocer las dificultades con las que se encuentran o han pasado para continuar y terminar sus estudios; en este punto el investigador puede conocer los puntos más críticos y determinar su foco de análisis.

PREGUNTA 1: Actualmente se esta dando gestión a los estudiantes que abandonan sus estudios?

RESPUESTA: No, no se esta dando ninguna gestión a los estudiantes desertores

ANALISIS: implementación de un proceso que mida, y analice dicha información

--

PREGUNTA 2: Conoce usted los factores que influyen en un estudiante para abandonar sus estudios??

RESPUESTA: No se conoce los factores, solo se asume el factor mas influyente es el dinero

ANALISIS: Es importante conocer con exactitud cuales son los factores que motivan a dejar sus estudios y así disminuir porcentajes de deserción y tener mas estudiantes graduados en las diferentes carreras.



INSTITUTO TECNOLÓGICO SUPERIOR CORDILLERA

PREGUNTA 3: Actualmente manejan algún tipo de software que analice el rendimiento de sus estudiantes??

RESPUESTA: No, ninguno

ANALISIS: Es un punto débil para la organización, no tener ningún software para analizar el rendimiento, es decir la toma de decisiones no es un proceso transparente.

PREGUNTA 4: Cree usted necesario la implementación de un proceso de análisis y gestión de la deserción de los estudiantes??

RESPUESTA: Si es importante conocer porque los estudiantes abandonan sus estudios.

ANALISIS: la toma de decisiones va a ser un proceso transparente y el sistema podrá proporcionar información segura y en tiempo real

3.3.3. ENCUESTA

En este punto el investigador ya tiene claro las preguntas clave que lo orientaran a determinar los factores y características que motivan a un estudiante a abandonar sus estudios, pudiendo apoyarse de técnicas estadísticas para determinar los mismos y obtener datos reales del objeto de análisis.



3.4. POBLACION Y MUESTRA

La población sobre la cual se va a trabajar, son los estudiantes los cuales serán tomados como el objeto de análisis, donde estudiaremos sus características e investigar los factores que inciden, y que lo motivan a abandonar sus estudios en carreras universitarias.

Factores de deserción:

- Rendimiento estudiantil (Promedio de Notas)
- Status social
- Sexo
- Localidad (Ubicación geográfica de domicilio)
- Promedio de egreso (colegio)
- Retiros de semestre
- Arrastres
- Arrastres con tercera matricula
- Cargas familiares
- Ocupación laboral
- Motivación

El subconjunto de la población será tomado de una muestra generada por el investigador que servirán como un análisis preliminar y obtener resultados tangibles en un tiempo o periodo corto, de lo que en posterior se pudiera aplicar a todos los estudiantes de todas las carreras de una Institución educativa, permitiendo llevar un control y hacer un seguimiento de los estudiantes que se encuentran con bajo rendimiento.

La formula para determinar el número de estudiantes, a los que aplicaremos criterios y técnicas de investigación, será la siguiente:



$$n = \frac{N}{(N-1)E^2 + 1}$$

De donde:

N = Tamaño de la población

n = Tamaño de la muestra

E² = Error máximo admisible

$$n = \frac{N}{(N-1)E^2 + 1}$$

$$n = \frac{921}{(921-1)((5)^2 + 1)}$$

$$n = \frac{921}{23001}$$

$$n = 0.04 * 100$$

$$n = 4\% \rightarrow 36 \text{ Encuestas}$$



3.5. DISEÑO DE LA ENCUESTA

Encuesta

En la ciudad de Quito en la provincia de Pichincha en el Instituto Tecnológico Superior Cordillera se realizara la siguiente encuesta, por favor se le pide total veracidad en sus respuestas:

1. Cual ha sido su rendimiento académico hasta la actualidad?

☐ Regular
☐ Bueno
☐ Sobresaliente

2. Que aspectos influyen en el bajo rendimiento?

☐ Profesores
☐ Falta de tiempo
☐ Descuido del estudiante

3. A pensado alguna vez abandonar sus estudios?

☐ Si
☐ No

4. Que lo motivo para tomar la decisión de abandonar sus estudios?

☐ Falta de dinero
☐ Falta de tiempo
☐ Rendimiento académico



INSTITUTO TECNOLÓGICO SUPERIOR CORDILLERA

☐ Materias con arrastre

☐ Otros

Especifique:_____

5. En que status social se encuentra actualmente?

☐ Bajo

☐ Medio

☐ Alto

6. Quien paga sus estudios actualmente?

☐ Padres

☐ Autogestión

☐ Otros

Especifique:_____

7. Cuantas materias ha tenido en arrastre?

☐ 1-3

☐ 4-8

☐ Ninguno

8. Cuantas materias con tercera matricula ha tenido?

☐ 1-3

☐ 4-8

☐ Ninguno



9. Le gusta la carrera que eligió?

☐ Si

☐ No

GRACIAS POR SU COLABORACION

3.6. CONCLUSION

Podemos concluir que la información recolectada tanto en las técnicas y métodos, servirán de apoyo para la tabulación e investigación de la problemática a tratar y dar la mejor alternativa de solución.

3.7. OPERACIONALIZACION DE LAS VARIABLES

Cuadro No. 3 Variables e Indicadores

Variable Independiente	Variable Dependiente	Indicadores
Analizar los factores de deserción	Sistemas de Datawarehouse y Dataminig	Porcentaje de estudiantes promovidos Porcentaje de estudiantes retirados Porcentaje de estudiantes retirados por tercera matricula Porcentaje de materias en arrastre Porcentaje de materias con tercera matricula

Autor: Daniel Benítez



CAPITULO IV

4. LA PROPUESTA

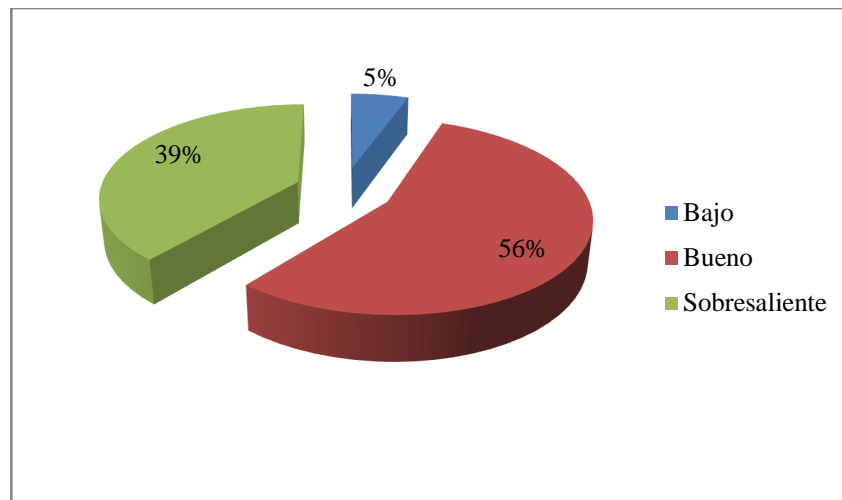
4.1 PROCESAMIENTO DE INFORMACION

Este ítem permite conocer mediante gráficos estadísticos el comportamiento y la realidad de los estudiantes, presentando lo siguiente:

Cuadro No. 4

1	Cual ha sido su rendimiento academico hasta la actualidad		
	Rendimiento Academico	Columna1	Porcentaje
	Bajo	2	5%
	Bueno	20	56%
	Sobresaliente	14	39%
	Total Encuestados	36	100%

Gráfico No. 1



Autor: Daniel Benítez



INSTITUTO TECNOLÓGICO SUPERIOR CORDILLERA

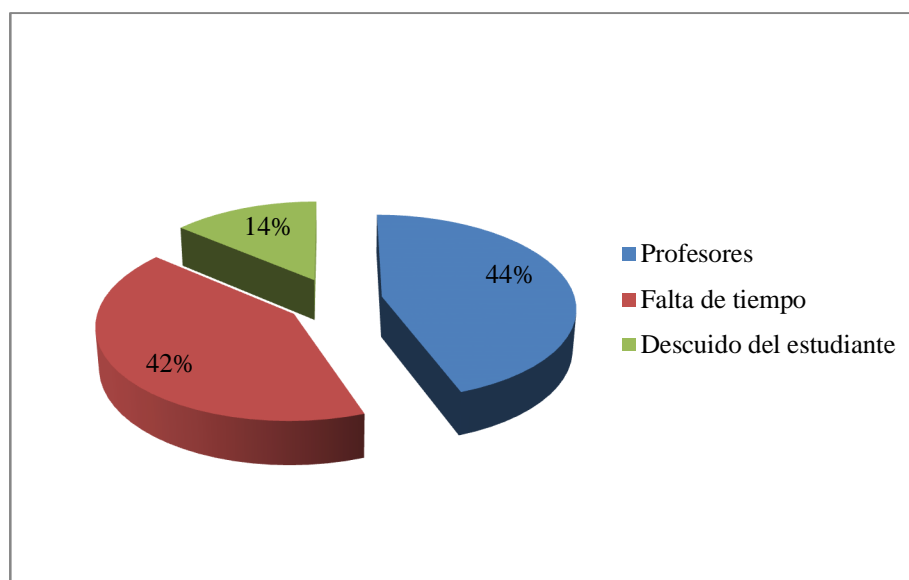
Conclusión: El estudio identifica en primer lugar que la variable fundamental y decisiva, es el rendimiento de los estudiantes, para la deserción, los mismos que toman esta decisión en base a su rendimiento el mismo que oscila siempre en un rendimiento Bueno 56%, que por falta de empeño, descuido, esfuerzo llega al grado de rendimiento Bajo con un 5%, de una totalidad de una muestra de 36 encuestados.



Cuadro No. 5

2	Que influye en el bajo rendimiento		
	Bajo Rendimiento	Columna1	Porcentaje
	Profesores	16	44%
	Falta de Tiempo	15	42%
	Descuido del Estudiantes	5	14%
	Total Encuestados	36	100%

Gráfico No. 2



Autor: Daniel Benítez

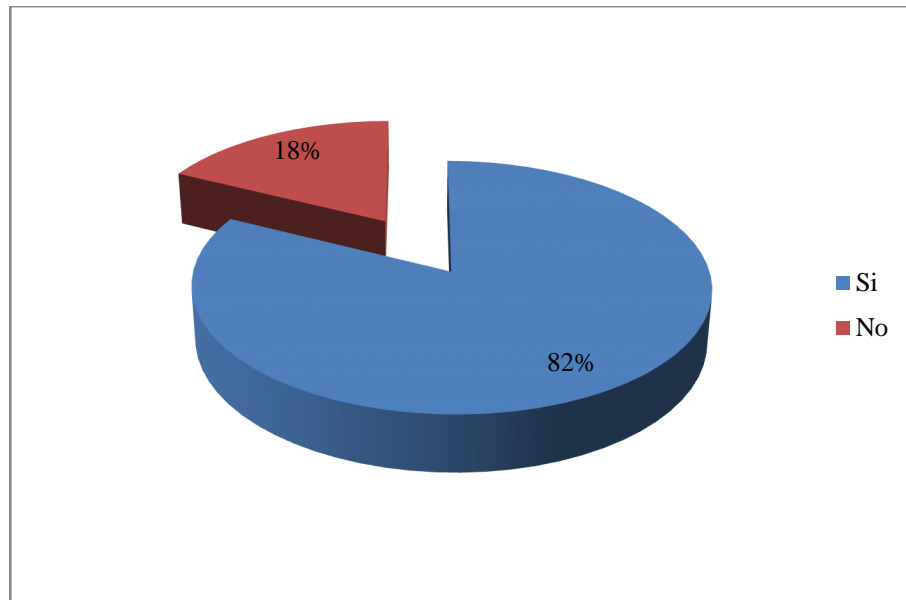
Conclusión: En base a resultados podemos afirmar que otra variable que tiene mas peso por decisión de los estudiantes, que su bajo rendimiento se fue afectado por los profesores con un 44% , seguido por un 42% por la falta de tiempo y solo un 14% por descuido del estudiante, los mismos factores que deberían ponerse en consideración para aclarar y no echar la culpa a profesores ya que el rendimiento depende del estudiante en si completamente.



Cuadro No. 6

3	Ha pensado alguna vez abandonar sus estudios		
	Abandonar Estudios	Columna1	Porcentae
	Profesores	28	82%
	Falta de Tiempo	6	18%
	Total Encuestados	36	100%

Gráfico No. 3



Autor: Daniel Benítez

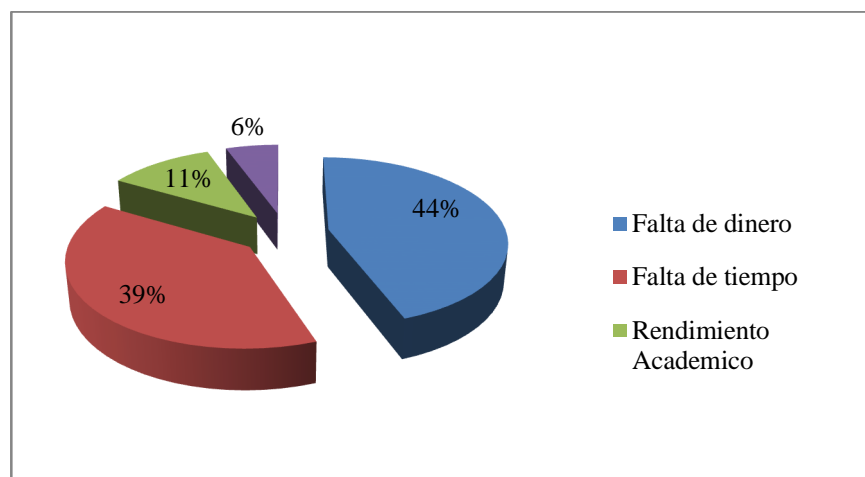
Conclusión: Entre las características que hay que destacar y son implícitas para la deserción de los estudiantes, y que hay que tomar riendas según la encuesta muchos estudiantes están pensando abandonar sus estudios que corresponde a un 82%, siendo superior esta decisión a la de continuar sus estudios ya que solo es el 18% el que desea continuar estudiando para mejorar su clase de vida y ser un ente productivo.



Cuadro No. 7

4	Que lo motivo para tomar la decisión de abandonar su estudios		
	Motivo de Abandono	Columna1	Porcentaje
	Falta de Dinero	16	44%
	Falta de Tiempo	14	39%
	Rendimiento Academico	4	11%
	Materias con Arrastre	2	6%
	Otros	0	0%
	Total Encuestados	36	100%

Gráfico No. 4



Autor: Daniel Benítez

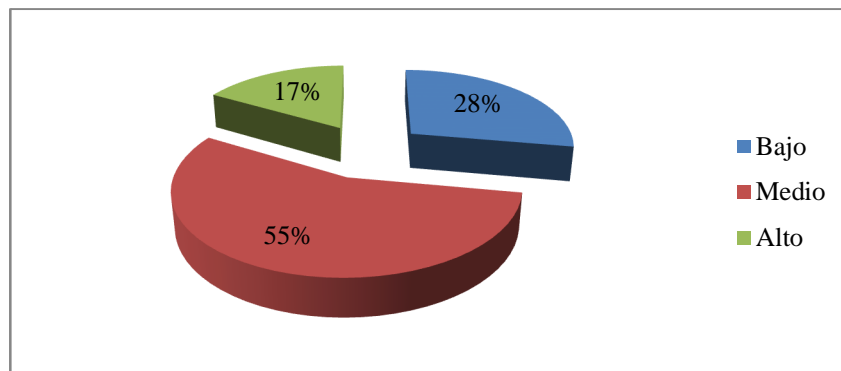
Conclusiones: Del mismo estudio se destaca que los motivos para la toma de decisión de abandonar sus estudios, depende muchísimo con la falta de dinero cubierto en este con el 44% de los encuestados, seguido por otro factor importante y decisivo que es la falta de tiempo por el tema de trabajo cubriendo el 39% el mismo que si no trabaja el estudiante no puede pagar sus estudios y abandonarlos, es muy inferior el tema de rendimiento académico con un 11% indicándonos así, que no toman decisiones los estudiantes por su bajo rendimiento motivo por el cual se da otro factor que es el de las materias con arrastre que cubre el 6 % de la muestra de encuestados.



Cuadro No. 8

5	En que status social se encuentra actualmente		
	Status	Columna1	Porcentaje
	Bajo	10	22%
	Medio	20	55%
	Alto	6	17%
	Total Encuestados	36	100%

Gráfico No. 5



Autor: Daniel Benítez

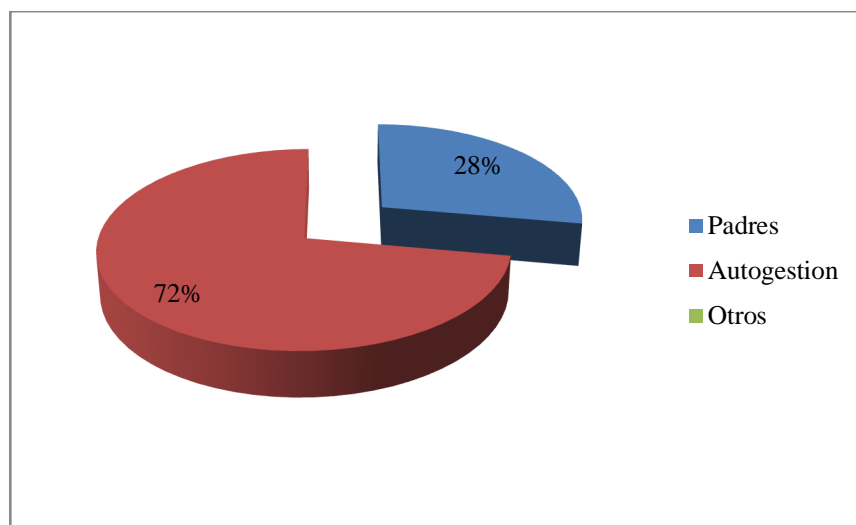
Conclusión: Algo que influye en el abandono de los estudiantes es el status social o capacidad de pago de los estudiantes que van de la mano del rendimiento por el tiempo que los mismo no poseen si trabajan para pagar sus estudios, nos da como resultado que el nivel Alto cubre el 17% teniendo posibilidades para cubrir todo para sus estudios, seguido por un Nivel Medio con el 55% dando como resultado que la mayoría de estudiantes lucha para poder cubrir pensiones, matriculas, optativas, y todo lo que implica seguir estudiando y cubrir todo, también hay que tomar en cuenta que un 28% que es considerable y llama la atención que pertenece al Nivel Bajo, dándonos que otro factor que sobresale es el nivel económico del estudiante, y toman la decisión de abandonar sus estudios siendo buenos y teniendo capacidad en lo que estudian.



Cuadro No. 9

6	Quien paga sus estudios actualmente		
	Estudios	Columna1	Porcentaje
	Padres	10	28%
	Autogestión	26	72%
	Total Encuestados	36	100%

Gráfico No. 6



Autor: Daniel Benítez

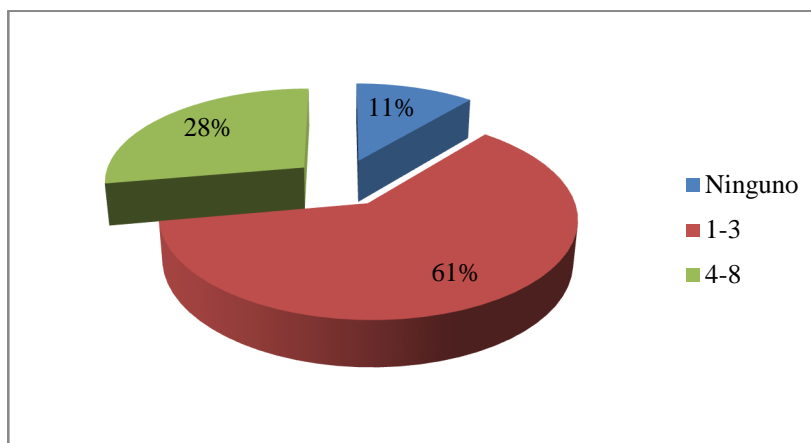
Conclusión: De los encuestados quien paga los estudio la mayoría con un 72% se autogestionan porque trabajan para pagarse sus estudios, dando también como resultado que eso les quita tiempo para cumplir con sus obligaciones estudiantiles, motivo por el cual sus estudios los pagan sus padres con un 28% para que sus hijos terminen sus estudios sin dificultad y lleguen a culminar sus estudios y no los abandonen por falta de tiempo.



Cuadro No. 10

7	Cuántas materias ha tenido en arrastre		
	Arrastres	Columna1	Porcentaje
	Ninguno	4	11%
	1-3	22	61%
	4-8	10	28%
	Total Encuestados	36	100%

Gráfico No. 7



Autor: Daniel Benítez

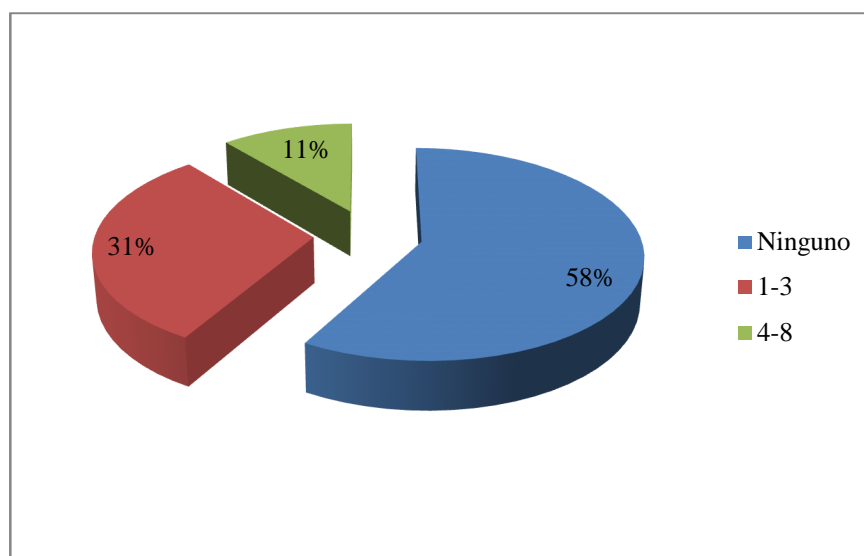
Conclusión: Los estudiantes enfrenta un gran numero de arrastres de materias por factores que influyen como por ejemplo dinero, tiempo, responsabilidad paternal, analizando los resultados esto cubre un 61% de 4-8 arrastres de los estudiantes encuestados, seguido por un 28% de 1-3 materias de arrastre las mismas que implican un ingreso adicional para poder culminar su estudios y es muy poco el numero de estudiantes que no tienen arrastres cubierto por un 11%.



Cuadro No. 11

8	Cuantas materias con tercera matricula ha tenido		
	Arrastre Tercera Matricula	Columna1	Porcentaje
	Ninguno	21	58%
	1-3	11	31%
	4-8	4	11%
	Total Encuestados	36	100%

Gráfico No. 8



Autor: Daniel Benítez

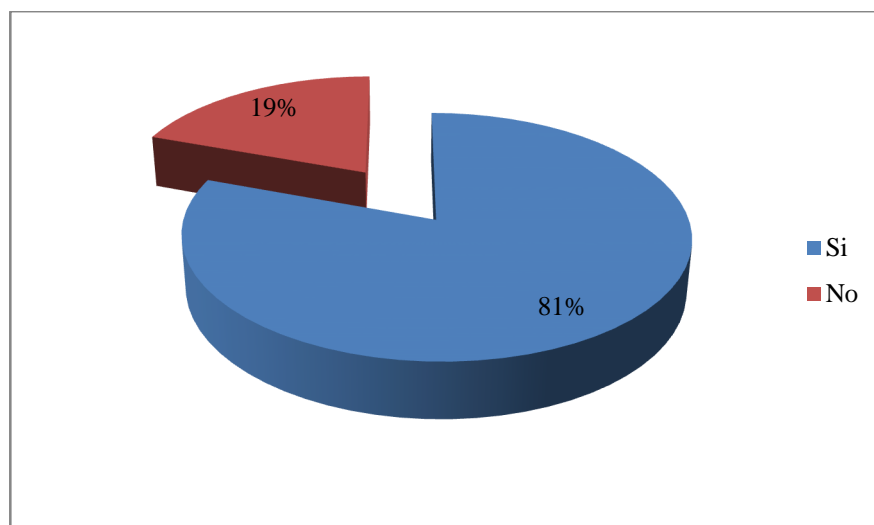
Conclusión: Para profundizar el diagnostico de la problemática de deserción de los estudiantes se debe enfocar en conocer las materias con tercera matricula, dándonos como resultado de 1-3 el 31% de los encuestados, de 4-8 el 11% y la mayoría de estudiantes con un 58% no posee ninguna materia con tercera matricula, es una variable hay que tomarla en cuenta ya que si tiene tercera matricula en varias materias los estudiantes en este escenario se encuentran en la cuerda floja que si no se toma cartas en el asunto son muy propensos para abandonar los estudios.



Cuadro No. 12

9	Le gusta la carrera que eligió		
	Carrera	Columna1	Porcentaje
	Si	29	81%
	No	7	19%
	Total Encuestados	36	100%

Gráfico No. 9



Autor: Daniel Benítez

Conclusión: Un factor predominante de abono de los estudios es no se eligió la carrera correcta y tuvieron mala orientación para decidir la carrera a seguir por eso los resultados son el 19% dicen que la carrera que eligieron no es la correcta, o no le gusta; y un 81% que les gusta la carrera que eligió, pero la decisión de abandonar sus estudios dependen de otros factores.



4.1 RESPUESTAS A LAS INTERROGANTES DE LA INVESTIGACION

1. Es importante debido a las características de un estudiante que son muy amplias de analizar y con el desarrollo de un DW podemos hacer un filtrado de datos que no son necesarios para un análisis y únicamente tener información de interés para generar conocimiento mediante indicadores de rendimiento clave y cuantitativo que faciliten estrategias y la toma de decisiones
2. La investigación demostró que no destinan esfuerzos en cuanto a los estudiantes que se retiran de las universidades, al igual que investigaciones, eso produce un total desconocimiento sobre esta gran problemática derivándose en que los estudiantes seguirán abandonando sus estudios, y que tienda a aumentar con el tiempo, en el transcurso de la investigación se consultaron varias fuentes de gran peso como ministerios de educación, gobierno, etc. En conclusión existe un desconocimiento de los estudiantes y sus características, eso imposibilita generar estrategias que minimicen la deserción y por ende tomar decisiones no adecuadas al problema.
3. Si es muy importante para toda organización de educación conocer el comportamiento de sus estudiantes, es decir conocer sus características mas importantes y caracterizarlos por grupos como desertores y no desertores, y definir perfiles de estudiantes desertores que me anticipen a que continúen su carrera universitaria, evitando la deserción y bajando índices, pudiendo alcanzar un mayor numero de estudiantes graduados y mejorar la calidad de los mismos
4. El investigador durante la fase de investigación, determino los siguientes indicadores, que permiten cuantificar y tener una visión general del problema y facilite la toma de decisiones
 - Porcentaje de estudiantes promovidos
 - Porcentaje de estudiantes retirados
 - Porcentaje de estudiantes retirados por tercera matricula
 - Porcentaje de materias en arrastre
 - Porcentaje de materias con tercera matricula



5. Tiene un impacto importante dentro de la solución contar con información real de una institución de educación, pero no significa una detención al proyecto, es por ello que el investigador propone que el desarrollo y mantenimiento se lo realice dentro de una organización y así cuantificar y apreciar datos reales que ayuden a la toma de decisiones.



CAPITULO V

5. LA PROPUESTA

El presente capítulo describe la puesta en marcha del sistema, y realiza una descripción detallada de los puntos mas importantes del desarrollo.

5.1. CONCLUSIONES Y RECOMENDACIONES

- CONCLUSIONES

Las herramientas, técnicas y análisis de la información, fueron de vital importancia dentro de la investigación, ya que permitió definir los indicadores más importantes y de más interés que pueden ayudar a interpretar la información de grandes cantidades de datos y extraer su conocimiento oculto.

Las técnicas de recolección de la información ayudaron en el aspecto investigativo, la entrevista fue de vital importancia, ya que permitió al investigador analizar con el director de escuela los diferentes criterios de análisis que motivan a los estudiantes a la deserción, con lo cual se concluye en los siguientes indicadores de rendimiento

- Porcentaje de estudiantes promovidos
- Porcentaje de estudiantes retirados
- Porcentaje de estudiantes retirados por tercera matricula
- Porcentaje de materias en arrastre
- Porcentaje de materias con tercera matricula



Estos indicadores brindan un conocimiento y van dirigidos a los directores de las diferentes escuelas, ya que ellos son los responsables de la toma de decisiones y los encargados de generar estrategias que bajen los índices de deserción.

El proceso de integración de datos es primer paso para una solución de business intelligence, en este punto se creo un paquete en la herramienta Business Intelligence Developement Studio, perteneciente a la Suite de desarrollo de Microsoft, el paquete de integración desarrollado brinda una solución de migración de grandes cantidades de datos, facilito el proceso de pasar los datos de un ambiente transaccional a un ambiente analítico, pudiendo hacerlo con poca cantidad de recursos.

Los datos antes de ser enviados al Datawarehouse, primero deben pasar por un proceso de limpieza, transformación, y estandarización, es decir no deben existir datos nulos, incompletos o corruptos ya que solo debe existir información dentro del Almacén de datos, también deben ser transformados y unificados en un solo formato, es decir no pueden haber datos con diferente concepto, si estamos cargando información del Tipo de sexo, no pueden existir varios datos, así como “masculino”, ”M” esto crearía una inconsistencia y no entregaría información real del criterio de análisis.

La carga de trabajo del servidor y el impacto es mínima, ya que el envío de los datos se lo puede hacer como un proceso batch que se ejecute cuando el servidor este sin carga de trabajo transaccional.

La construcción de del datawarehouse es la técnica de mayor peso dentro de la investigación, ya que es la encargada de recibir los datos y características de los estudiantes, procesar, y transformar los datos en



información a través de indicadores, que faciliten la interpretación a los usuarios.

En conclusión, Todo el proceso antes mencionado se deriva en la interpretación de los indicadores en forma estadística, para el ejemplo podemos citar el indicador de rendimiento “Porcentaje de deserción por tercera matricula” este criterio de análisis, nos permite observar la cantidad de estudiantes que son retirados obligatoriamente por la escuela, desde este punto ya podemos empezar a analizar estos estudiantes y sus características

- **RECOMENDACIONES**

El investigador luego de toda la investigación realizada propone que el proyecto tenga continuidad, y sea aplicado a una organización de educación para poder cuantificar datos reales de lo que esta sucediendo con los estudiantes, cual es su comportamiento, y que podría llegar a pasar en el futuro, es decir anticiparse a eventos futuros de deserción, y que permitan bajar los índices de deserción estudiantil, y mejore la calidad de profesionales que egresen.



CAPITULO VI

6. ASPECTOS ADMINISTRATIVOS

6.1. ALCANCE

El proyecto de inteligencia de negocios tiene como principio realizar una demostración de posibles factores de deserción, partiendo de indicadores claves de rendimiento que orienten al usuario final, partiendo de supuestos definidos por el investigador, para ello se realizara un desarrollo paulatino, por las características de una solución BI ya que el modelo de minería de datos necesita de una muestra extensa que solo el tiempo la puede proporcionar, para ello se iniciara con la construcción de un sistema datawarehouse, que es el almacén de datos que servirá de input para un modelo de minería de datos, que es el encargado de analizar el conjunto de información y determinar patrones de conducta o características similares, es decir determinar factores de deserción.

La propuesta por parte del desarrollador es la construcción de un sistema datawarehouse que contempla dentro de él, un conjunto de procesos como lo son:

Proceso ETL es el proceso de extracción de información de múltiples fuentes de datos, en este proceso el investigador definió un supuesto, es decir desarrollo un modelo de negocio con una base de datos operacional para extraer la información de dicha fuente a través de procedimientos, paquetes de integración de datos, Jobs, etc.

En la transformación, el investigador definió estándares dentro del almacén de datos, es decir que el sistema DW tenga información transparente y



unificada que coincida con los tipos de datos y longitudes definidos dentro del almacén

En la carga de datos, que es el proceso de integración de información dentro del sistema datawarehouse, se desarrollara un proceso de carga a través de paquetes SSIS con herramientas Sql Server Integration Services.

Finalmente el investigador, desarrollara el modelo del almacén de datos que es el contenedor de la información y será el input para el Datawarehouse que es el encargado de procesar los indicadores de rendimiento y entregar la información para su respectivo análisis, para ello se lo realizara dinámicamente con la herramienta Excel que permite realizar un análisis Olap de la información y estáticamente como reportes creados con la herramienta Reporting Services. Los indicadores de rendimiento que podremos visualizar y que fueron generados en la investigación por mayor peso e importancia, en esta fase inicial de desarrollo son los Factores Academicos, tales como:

- Rendimiento Academico (Promedio de Notas en la carrera)
- Porcentaje de Estudiantes promovidos
- Porcentaje de Estudiantes Retirados
- Porcentaje de Estudiantes Retirados por tercera matricula
- Porcentaje de Estudiantes Reprobados
- Porcentaje de Estudiantes Reprobados por tercera matricula



6.2. ESTRUCTURA ORGANIZACIONAL DE LA EMPRESA

Si bien es cierto que no esta orientado a ninguna organización de educación superior, pero si va dirigido u orientado a los directores de las escuelas. El sistema datawarehouse tendrá un tratamiento en un ambiente local, ya que es una demostración por parte del investigador, y por consiguiente este punto no es aplicable.

6.3. ESTRUCTURA INFORMATICA

El sistema datawarehouse tendrá un tratamiento en un ambiente local, ya que es una demostración por parte del investigador, y por consiguiente este punto no es aplicable.



6.4. ALTERNATIVAS

En este punto especificaremos las diferentes puntos que conlleva una propuesta, a fin de evaluar la mejor alternativa y tomar decisiones acertadas.

Cuadro No. 13 Alternativas

Actividades	Especificaciones	Peso	Cumple	No Cumple
Arquitectura	Tres capas	10	X	
Ambiente de Desarrollo	Web	5	X	
Plataforma	Windows	5	X	
Metodología	HEFESTO	5	X	
Diseño	Power Designer	10	X	
Estándares	Bases de datos multidimensionales	10	X	
Font_end	Analisis Services Integration Services	10	X	
Back-end	SQL Server	5	X	
Pruebas	Unidad, integración, validación, sistema,	10	X	
Manuales	Usuario , técnico	5	X	
Capacitación	Usuario y técnico	5	X	
Costo	Licencias, impresiones	5	X	
Tiempo	6 meses	5	X	
Garantía técnica	6 meses	5	X	
Soporte Técnico	6 mese	5	X	
Total		100		



6.4.1. ALTERNATIVAS

En este punto detallaremos la propuesta realizada por el investigador del proyecto versus empresas comerciales.

Cuadro No. 14 Investigador versus Empresas Comerciales

Actividades	DESARROLLADOR		EMP COMERCIALES	
	Cumple	No Cumple	Cumple	No Cumple
Arquitectura	X		X	
Ambiente de Desarrollo	X		X	
Plataforma	X		X	
Metodología	X			X
Diseño	X		X	
Estándares	X		X	
Font_end	X		X	
Back-end	X		X	
Pruebas	X		X	
Manuales	X		X	
Capacitación	X		X	
Costo	X		X	
Tiempo	X		X	
Garantía técnica	X		X	
Soporte Técnico	X		X	
Total				



6.5. SELECCIÓN DE ALTERNATIVAS

Punto Con la finalidad de poder evaluar las dos alternativas presentadas se determinaron algunos parámetros para la calificación de cada uno de los ítems presentados y de acuerdo a la importancia en la construcción de los mismos, por consiguiente se determinaron de la siguiente manera:

Cuadro No. 15 Selección de alternativas para evaluar

Actividades	Peso	Empresa	Tesis
Arquitectura	10	10	10
Ambiente de Desarrollo	5	5	5
Plataforma	5	5	5
Metodología	5	0	5
Diseño	10	10	10
Estándares	10	10	10
Font_end	10	10	10
Back-end	5	5	5
Pruebas	10	10	10
Manuales	5	5	5
Capacitación	5	5	5
Costo	5	5	5
Tiempo	5	5	5
Garantía técnica	5	5	5
Soporte Técnico	5	5	5
Total		95	100



6.6. PROCESOS

Estos indicadores fueron definidos y categorizados para realizar un análisis preliminar de hechos anteriores, presentes y futuros que pueden llegar a suceder, y que permiten entender el comportamiento de los estudiantes dentro de su carrera universitaria, y empezar a sacar conclusiones, motivos y factores que producen una deserción universitaria, predecir aspectos futuros que puedan llegar a suceder analizando con lógica la información clara, concisa que genera el Datawarehouse (Almacén de Datos).

FACTORES DE DESERCIÓN ACADÉMICA

Estos son los posibles motivos que afectan a los estudiantes, y que serán analizados para determinar su peso en las decisiones de los mismos, se han definido en forma de indicadores de rendimiento para su respectiva gestión y análisis que ayudan en la toma de decisiones

- INDICADOR NUMERO DE ESTUDIANTES PROMOVIDOS

Este indicador se creo bajo la perspectiva de entregar un conocimiento global y empezar a generar conocimiento y conocer en porcentajes si hay un buen numero de estudiantes que pasan de nivel, pudiendo diferenciar si hay un alto índice o un bajo, si existe un bajo porcentaje, podemos empezar a sacar conclusiones, motivos o factores sobre que esta pasando y empezar a deducir factores de análisis y tomar medidas de prevención.

Bajo este indicador podemos analizar la información sobre cuantos estudiantes están pasando de nivel sin materias con arrastre, la siguiente tabla representa el indicador y sus medidas de análisis



INSTITUTO TECNOLÓGICO SUPERIOR CORDILLERA

El indicador fue construido utilizando la función Count de los estudiantes que pasaron de nivel sin ningún arrastre, y para obtener su porcentaje se realizó una expresión matemática

Medida Matriculados = Count (Estudiantes Matriculados)

Medida Promovidos = Count (Estudiantes con pase de nivel normal)

Medida %Promovidos = (Medida Promovidos * 100) / Medida Matriculados

Cuadro No. 16 Número de Estudiantes Matriculados y Promovidos

AÑO	PERIODO	ESCUELA	NIVEL	MATERIA	MATRICULADOS	PROMOVIDOS	% PROMOVIDOS
2012	Abr - Sep	Sistemas	1er Nivel	Programacion I	1890	838	44.34
2012	Abr - Sep	Sistemas	2do Nivel	Programacion II	838	600	71.60
2012	Abr - Sep	Sistemas	3er Nivel	Programacion III	600	408	68
2012	Abr - Sep	Sistemas	4to Nivel	Programacion IV	408	260	63.73
2012	Abr - Sep	Sistemas	5to Nivel	Programacion V	260	200	76.92
2012	Abr - Sep	Sistemas	6to Nivel	Programacion VI	200	30	15

Autor: Daniel Benítez

- INDICADOR NUMERO DE ESTUDIANTES RETIRADOS

El concepto de este indicador es presentar información sobre estudiantes que se retiran voluntariamente sea de forma parcial o total de la carrera, por motivos que en un futuro serán demostrados con minería de datos.

Este indicador es importante para la toma de decisiones, permite conocer cuantos estudiantes se han retirado de la carrera y puede ayudar a empezar a sacar deducciones sobre factores de deserción, es decir si el



INSTITUTO TECNOLÓGICO SUPERIOR CORDILLERA

porcentaje de retirados de la escuela de sistemas es bastante alto, podemos deducir que los estudiantes están siendo motivados a dejar sus estudios, y es importante conocer cuales son esos motivos o factores de deserción.

Permite empezar a detectar patrones de conducta en nivel superior, para posterior minimizar el análisis, partiendo de lo general a lo particular. Permite además conocer en que situación actualmente se encuentra y analizar los factores que los llevaron a tomar esta decisión.

La construcción del indicador, se resuelve de la siguiente manera:

Medida Matriculados = Count (Estudiantes Matriculados)

Medida Retirados = Count (Estudiantes que han solicitado el retiro)

Medida %Retirados = (Medida Retirados * 100) / Medida Matriculados

Cuadro No. 17 Número de Estudiantes Matriculados y Retirados

AÑO	PERIODO	ESCUELA	NIVEL	MATERIA	MATRICULADOS	RETIRADOS	% RETIRADOS
2012	Abr - Sep	Sistemas	1er Nivel	Programacion I	1890	914	48.36
2012	Abr - Sep	Sistemas	2do Nivel	Programacion II	838	150	17.90
2012	Abr - Sep	Sistemas	3er Nivel	Programacion III	600	90	15
2012	Abr - Sep	Sistemas	4to Nivel	Programacion IV	408	14	3.43
2012	Abr - Sep	Sistemas	5to Nivel	Programacion V	260	15	5.77
2012	Abr - Sep	Sistemas	6to Nivel	Programacion VI	200	67	33.5

Autor: Daniel Benítez



Bajo los resultados que presenta el indicador en forma estadística, podemos observar donde se encuentra el mayor porcentaje de retirados y se puede empezar a sacar conclusiones de análisis de deserción, es decir observamos que en el 3er nivel y en la materia de Programación III existe mayor porcentaje y empezamos a investigar y determinar factores de deserción, y a hacer preguntas por que esa materia tiene mas deserción en comparación a las otras, de igual forma lo hacemos con el nivel, ahí es necesario aplicar técnicas de minería de datos que determinen patrones.

- **INDICADOR NUMERO DE ESTUDIANTES RETIRADOS TM**

Este indicador es el más importante de todos, ya que aquí el estudiante es removido o retirado de la escuela obligatoriamente, por perder una materia con tercera matricula. En este punto ya empezamos a recolectar la muestra de información que necesita el modelo de minería de datos, ya que estos estudiantes no volverán a continuar la carrera pero sus características nos pueden ayudar a definir perfiles de estudiantes con características de deserción.

La construcción del indicador, se resuelve de la siguiente manera:

Medida Matriculados = Count (Estudiantes Matriculados)

Medida RetiradosTM = Count (Estudiantes retirados obligatoriamente)

Medida %RetiradosTM = (Medida RetiradosTM * 100) / Medida Matriculados



INSTITUTO TECNOLÓGICO SUPERIOR CORDILLERA

Cuadro No. 18 Numero de Estudiantes Matriculados y Retirados por Tercera Matricula.

AÑO	PERIODO	ESCUELA	NIVEL	MATERIA	MATRICULADOS	RETIRADOS TM	% RETIRADOS TM
2012	Abr - Sep	Sistemas	1er Nivel	Programacion I	1890	29	1.53
2012	Abr - Sep	Sistemas	2do Nivel	Programacion II	838	14	1.67
2012	Abr - Sep	Sistemas	3er Nivel	Programacion III	600	36	6
2012	Abr - Sep	Sistemas	4to Nivel	Programacion IV	408	45	11.03
2012	Abr - Sep	Sistemas	5to Nivel	Programacion V	260	8	3.08
2012	Abr - Sep	Sistemas	6to Nivel	Programacion VI	200	20	10

Autor: Daniel Benítez

- INDICADOR NUMERO DE ESTUDIANTES REPROBADOS

Este indicador muestra información sobre el porcentaje de arrastres diferenciado la cantidad de estudiantes por periodo, nivel y fundamental por materia, dando así un claro énfasis en que materia hay un porcentaje elevado de estudiantes reprobados, los mismos que posiblemente se convertirán en estudiantes desertores, y también con este indicador se tomara en cuenta cuales son las posibles falencias de catedras o posibles dificultades en la materia dictada.

Este indicador es fundamental ya que el porcentaje nos indicara cuantos estudiantes se podrán convertir en posibles desertores de la carrera pudiendo así prevenir la deserción de los mismo poniendo en practica estrategias que los mantengan en la carrera.

La construcción del indicador, se resuelve de la siguiente manera:



INSTITUTO TECNOLÓGICO SUPERIOR CORDILLERA

Medida Matriculados = Count (Estudiantes Matriculados)

Medida Reprobados = Count (Estudiantes que han reprobado materias)

Medida %Reprobados = (Medida Reprobados * 100) / Medida Matriculados

Cuadro No. 19 Número de Estudiantes Matriculados y Reprobados

ANIO	PERIODO	ESCUELA	NIVEL	MATERIA	MATRICULADOS	REPROBADOS	%REPROBADOS
2012	Abr - Sep	Sistemas	1er Nivel	Programación I	1890	69	3.65
2012	Abr - Sep	Sistemas	2do Nivel	Programación II	838	38	4.53
2012	Abr - Sep	Sistemas	3er Nivel	Programación III	600	40	6.67
2012	Abr - Sep	Sistemas	4to Nivel	Programación IV	408	53	12.99
2012	Abr - Sep	Sistemas	5to Nivel	Programación V	260	26	10
2012	Abr - Sep	Sistemas	6to Nivel	Programación VI	200	35	17.5

- INDICADOR NUMERO DE ESTUDIANTES REPROBADOS TM

El concepto de este indicador es representar el porcentaje de estudiantes que están tomando una materia con tercera matricula, es decir que aun no se produce una deserción de carrera obligatoria pero están condicionados a que si ocurre una reprobación de la materia nuevamente, pues será removido de la carrera obligatoriamente.

Este es un indicador fundamental porque este nos permite visualizar con énfasis la realidad del país, y nos mostrara los posibles desertores y además que van a quedar imposibilitados para seguir estudiando e incluso desertar definitivamente de su carrera, y conformarse a conseguir un trabajo no tan



INSTITUTO TECNOLÓGICO SUPERIOR CORDILLERA

bien remunerado pero fijo, para cumplir con sus gastos mensuales, y estos estudiantes se convertirán en profesionales mediocres e inconclusos, dando una alerta de la cantidad de estudiantes que reprueban y dejan sus estudios obligatoriamente, y que no culminaran su carrera y no se convertirán en profesionales y frustraran su metas, sueños y objetivos. Además también se podrá hacer un seguimiento mediante el sistema de Datawarehouse, para analizar a los estudiantes que se encuentran en esta categoría para poderles ayudar a culminar sus estudios o mas claro prevenir su decisión de abandonar sus estudios.

La construcción del indicador, se resuelve de la siguiente manera:

Medida Matriculados = Count (Estudiantes Matriculados)

Medida ReprobadosTM = Count (Estudiantes)

Medida %ReprobadosTM = (Medida ReprobadosTM * 100) / Medida Matriculados

Cuadro No. 20 Número de Estudiantes Matriculados y Reprobados por Tercera Matricula

ANIO	PERIODO	ESCUELA	NIVEL	MATERIA	MATRICULADOS	REPROBADOS TM	%REPROBADOS TM
2012	Abr - Sep	Sistemas	1er Nivel	Programacion I	1890	40	2.12
2012	Abr - Sep	Sistemas	2do Nivel	Programacion II	838	36	4.30
2012	Abr - Sep	Sistemas	3er Nivel	Programacion III	600	26	4.33
2012	Abr - Sep	Sistemas	4to Nivel	Programacion IV	408	36	8.82
2012	Abr - Sep	Sistemas	5to Nivel	Programacion V	260	11	4.23
2012	Abr - Sep	Sistemas	6to Nivel	Programacion VI	200	48	24



- INDICADOR ESTUDIANTES CON BAJO RENDIMIENTO

El datawarehouse permite realizar un análisis del rendimiento de los estudiantes en su carrera universitaria, diferenciado por varias dimensiones, tales como su nivel, materia y año de matrícula.

Es un indicador el cual nos va ayudar a establecer los estudiantes que no están cumpliendo con el rendimiento académico establecido para pasar la materia, y tomar acciones de prevención mediante un análisis adelantado mediante la utilización del sistema Datawarehouse, pudiendo poner en marcha programas de seguimientos a los estudiantes con estos problemas académicos, para así ayudar a los estudiantes a culminen sus estudios y cumplan sus objetivos planteados para llegar hacer profesionales exitosos, además nos podremos dar cuenta que factores influyen para este tipo de toma de decisiones para su bajo rendimiento académico y por ejemplo si fuese su bajo rendimiento por cuestiones de trabajo, socioeconómicos, familiares, tiempo, etc., que repercuten en el rendimiento académico.



FACTORES DE DESERCIÓN PERSONAL

Se definieron indicadores que son de carácter personal, es decir criterios de análisis que afectan directamente las emociones de los estudiantes y los motivan a la deserción.

- **INDICADOR PORCENTAJE DE ESTUDIANTES CON VOCACION A LA CARRERA**

Este indicador permite conocer de manera global los estudiantes que tienen motivación para con la carrera diferenciado por motivados y desmotivados.

Cuadro No. 21 Porcentaje de Estudiantes Motivados y Desmotivados

Escuela	Vocación	Porcentaje
Sistemas	Motivados	80%
Sistemas	Desmotivados	20%

Autor: Daniel Benítez

Con este tipo de resultados el dueño del negocio, podría poner mayor preocupación e interés para analizar los estudiantes que se encuentran en este porcentaje, y poner estrategias preliminares para ayudar y enfocarse en este tipo de estudiantes que se podrían convertir en posibles estudiantes desertores de carrera.



INSTITUTO TECNOLÓGICO SUPERIOR CORDILLERA

- INDICADOR PORCENTAJE DE ESTUDIANTES INSATISFECHOS CON LA CARRERA

Con este tipo de resultados el dueño del negocio, podría poner mayor preocupación e interés para analizar los estudiantes que se encuentran en este porcentaje, y poner estrategias preliminares para ayudarles y enfocarse en este tipo de estudiantes que se podrían convertir en posibles estudiantes desertores de carrera. O tal vez ayudarlos para una mejor orientación, y ubicación de sus motivaciones y objetivos profesionales para así sacarlos de este porcentaje y convertirlos en estudiantes motivados para culminar su carrera y trabajar con éxito y conocimientos bien fundados.

FACTORES DE DESERCIÓN SOCIO – ECONÓMICAS

Se van a conceptualizar indicadores que tienen gran peso dentro de la investigación, y se trata de demostrar que son características de gran importancia y que definen a los estudiantes como desertores y definir perfiles de estudiantes con características de deserción.

- INDICADOR PORCENTAJE DE ESTUDIANTES CON CARGAS FAMILIARES

Este indicador permite tener en claro que estudiantes se encuentran en una situación complicada porque poseen cargas familiares o están a cargo de sus hermanos, ya que sus padres se encuentran fuera del país, siendo este un alto índice, ya que se encuentran obligados a dividir sus tareas u obligaciones, con su familia, trabajo y estudios, dejando a un lado o clasificando por prioridad sus tareas, dando como resultado de esto abandonar sus estudios y dejar de cumplir el sueño de llegar a ser un



profesional exitoso, trabajar en lo que le guste o tal vez ser emprendedor y ponerse su propio negocio.

- **INDICADOR PORCENTAJE DE ESTUDIANTES QUE TRABAJAN**

Este indicador demuestra puntualmente las cantidad de estudiantes que se encuentran laborando, mientras culminan sus estudios, pero también se encuentran motivados para trabajar ya que actualmente en el mercado laboral exigen experiencia laboral mayor a un año, a veces también porque tienen que solventarse sus estudios (son independientes), pero esto arrastra consigo una causa que es fundamental que por tiempo y cumplir actividades laborales dejan de asistir a las clases, faltan o llegan atrasados y los docentes no les dejan ingresar a la hora de clases, siendo así otro factor fundamental para la toma de decisión de abandonar sus estudios por cumplir con su trabajo.

- **INDICADOR PORCENTAJE DE ESTUDIANTES DEPENDIENTES E INDEPENDIENTES**

Este índice nos permitirá tener una clara visión de que cantidad de estudiantes trabajan y se solventa sus estudios y están en riesgo de faltar a clases por su trabajo, llegar atrasados a clases por cuestiones de trabajo, no tener tiempo de presentar trabajos en grupo, reunirse con sus compañeros y que todo esto sea causa de un bajo rendimiento, pero sino se pagan ellos mismo sus estudios quien?, en el caso de estudiantes que dependen de sus padres para continuar estudiando, las altas pensiones o no poder cancelar a tiempo las cuotas, eso también es causa de abandono de estudios, de un bajo rendimiento, y no culminar sus estudios por este tipos de factores, si cumplir sus objetivos estudiantiles.



- INDICADOR STATUS SOCIAL DE LOS ESTUDIANTES

Este indicador representa algo fundamental ya que permite que el dueño del negocio, tenga en claro que estudiantes tienen posibilidades económicas para cumplir con los pagos a tiempo, no retrasarse, ni quedar impago por varios semestres, y así tiene la cantidad exacta de estudiantes que pueden acceder a una beca estudiantil por falta de solvencia económica, y obviamente cumplan académicamente con las metas establecidas para acceder a la misma.

6.7. DESCRIPCION DE LA METODOLOGIA DE DESARROLLO

El proyecto basa su desarrollo bajo la metodología HEFESTO con licencia de libre distribución, de construcción de un Datawarehouse en su versión 2.0 del Autor Ing. Bernabeu Ricardo Dario.

6.7.1. ANALISIS DE REQUERIMIENTOS

A) Identificar Preguntas

El primer paso de la metodología es identificar las preguntas clave de negocio (KPI) a las que necesitamos dar respuesta, esto lo podemos conseguir en función a técnicas de investigación como entrevistas, cuestionarios, observación dependiendo de las habilidades del investigador utilizara la técnica más apropiada que tenga las características necesarias, para conocer y obtener todos los requerimientos del usuario experto del negocio. El Objetivo de esta fase es obtener las necesidades de información y las respuestas que el usuario espera del Sistema Almacén de Datos, al ser un proyecto con fines de demostración, se parte de



un supuesto y gracias a la técnica de investigación “La Observación” se concluyo en las siguientes:

- Conocer el porcentaje de estudiantes retirados de la Escuela de sistemas en función al semestre de la carrera
- Conocer el porcentaje de estudiantes que fueron promovidos al siguiente nivel de la carrera de sistemas en función al semestre de la carrera
- ¿Conocer los estudiantes que tienen un rendimiento bajo del promedio, pudiendo conocer las materias en las que están fallando y saber a qué nivel corresponden y a qué escuela pertenecen de la sucursal matriz, del semestre actual o en curso?
- ¿Conocer los estudiantes que tienen un rendimiento alto, pudiendo conocer las materias en las que son sobresalientes y saber a qué nivel corresponden y a qué escuela pertenecen de la sucursal matriz, del año 2011?

Es de vital importancia poder tener dicha información para la toma de decisiones, es decir anticiparse a eventos que pueden llegar a suceder con alumnos y así evitar su deserción o pérdida de niveles.



B) Identificar Indicadores y perspectivas

La siguiente fase es descomponer las preguntas o necesidades de información supuestas y determinar los respectivos Indicadores y Perspectivas, para ello debemos diferenciar que los indicadores nos permiten medir un proceso y conocer su variación en función al tiempo, y las perspectivas son las vistas por las cuales vamos a observar o examinar los indicadores con el fin de responder a las preguntas planteadas en la fase anterior.

A continuación analizaremos las preguntas e identificaremos sus respectivos indicadores y perspectivas.

- Indicadores:
 - Promedio de notas
- Perspectivas:
 - Unidad de Negocio
 - Escuelas
 - Niveles
 - Materias
 - Estudiantes
 - Tiempo

C) Modelo Conceptual

En esta etapa se desarrollará el modelo conceptual del Datawarehouse, partiendo de los indicadores y perspectivas. En la construcción del modelo conceptual de un DW se deben colocar todas las perspectivas en la parte izquierda y los indicadores en la parte derecha, relacionadas por un hecho en



común, aplicando el concepto al proyecto la relación que une las perspectivas e indicadores, es “RENDIMIENTO”.

6.7.2. ANALISIS DE SISTEMAS OLTP

En esta fase analizaremos los sistemas transaccionales que serán los proveedores de la información que necesitamos para dar respuestas y definiremos los campos y cálculos necesarios para los indicadores y perspectivas. En este punto el investigador desarrollo un sistema de bases de datos simulando un negocio

A) Conformar Indicadores

En esta paso se debe definir como se realizara el cálculo de los indicadores, aplicando el concepto al proyecto tendríamos lo siguiente:

- Promedio de notas
 - o Hechos: Promedio de notas
 - o Función de sumarización: AVG

B) Establecer Correspondencias

El objetivo de este paso es examinar puntualmente sus bases de datos de todos los sistemas transaccionales que proveerán de la información necesaria para el cálculo de los indicadores y los campos necesarios para las perspectivas de análisis e identificar las correspondencias con el Datawarehouse.

Esta fase no es aplicable al proyecto ya que los datos residen en documentos de texto plano, y no son generados por ningún sistema transaccional.



C) Nivel de Granularidad

En este punto se establecen o se seleccionan los campos de la base de datos transaccional que corresponderán a las perspectivas

Esta fase de la metodología no es aplicable al proyecto ya que los datos residen en documentos de texto plano, y no son generados por ningún sistema transaccional.

6.7.3. MODELO LOGICO DW

En esta fase se realiza la construcción del modelo lógico de la estructura del Datawarehouse, partiendo con el resultado del modelo conceptual. Para ello debemos seleccionar el tipo de esquema (Estrella, Copo de nieve y Constelación) y se confeccionara la tabla de hechos que contiene los indicadores con sus respectivas dimensiones que contienen las perspectivas de análisis.

A) Tipo de modelo Lógico del DW

El tipo de esquema con el que se construirá el modelo lógico del Datawarehouse, ya que por sus características y ventajas es el que mejor se ajusta a las necesidades del negocio.

B) Tablas de Dimensiones

En este punto se deben definir las tablas de dimensiones que formaran parte del esquema seleccionado, estas dimensiones se construirán a partir de las perspectivas definidas en el modelo conceptual, es decir habrá tantas dimensiones como perspectivas



INSTITUTO TECNOLÓGICO SUPERIOR CORDILLERA

existentes. Siguiendo los lineamientos que define la metodología, deberemos realizar el siguiente proceso:

- Se elegirá un nombre que represente la tabla dimensión
- Se añadirá un campo que identifique su clave principal
- Se redefinirán los nombres de los campos, si no son lo suficientemente intuitivos

Aplicando el proceso a nuestra investigación, obtendremos el siguiente resultado:

- Perspectiva “UNIDAD NEGOCIO”

DIM_Unidad_Negocio	
<u>ID Negocio</u>	<pi>
Descripcion	

- Perspectiva “ESCUELA”

DIM_Escuela	
<u>ID Escuela</u>	<pi>
Descripcion	

- Perspectiva “NIVEL”

DIM_Nivel	
<u>ID Nivel</u>	<pi>
Descripcion	



- Perspectiva “MATERIAS”

DIM_Materias	
<u>ID Materia</u>	<pi>
Descripcion	

- Perspectiva “ESTUDIANTES”

DIM_Estudiente	
<u>ID_Estudiente</u>	<pi>
Nombres	

- Perspectiva “TIEMPO”

DIM_Tiempo	
<u>ID Fecha</u>	<pi>
Año	
Semestre	
Bimestre	

C) Tablas de Hechos

En este punto se deben definir las tablas de hechos que formaran parte del esquema seleccionado, estos hechos se construirán a partir de los indicadores definidos en el modelo conceptual, es decir habrá tantas hechos como indicadores exista. Siguiendo los lineamientos que define la metodología, deberemos realizar el siguiente proceso:

- Se elegirá un nombre que represente la tabla hechos
- Se definirán las claves primarias que se componen de las claves primarias de las tablas de dimensiones relacionadas



INSTITUTO TECNOLÓGICO SUPERIOR CORDILLERA

- Se crearan tantos campos de hechos como indicadores se hayan definido en el modelo conceptual

Aplicando el proceso a nuestra investigación, obtendremos el siguiente resultado:

- Indicador “Rendimiento”

Rendimiento
ID_Negocio
ID_Estudiante
ID_Materia
ID_Escuela
ID_Nivel
ID_Fecha
Rendimiento

D) Uniones

En este ítem se debe realizar las uniones de las dimensiones a las tablas de hechos para el tipo de esquema seleccionado, Esquema en estrella se obtiene el modelo lógico del Datawarehouse, que se puede visualizar en el punto **6.8 Modelo Lógico**.



6.8. MODELO LOGICO

La construcción del modelo lógico fue desarrollada con la herramienta de modelación Power Designer en su versión 16.1.0.3637

Cuadro No. 22 Modelo Lógico del Sistema Datawarehouse

Modelacion Power Designer

AutoDaniel Benítez

Anexo No.3

Modelo Logico Almacen de datos

6.9. MODELO FISICO

La construcción del modelo Físico se basa en la construcción del modelo lógico y fue desarrollada con la herramienta de modelación Power Designer en su versión 16.1.0.3637

Cuadro No. 4 Modelo Físico

Modelación Power Designer

Autor: Daniel Benítez

Anexo No.3

Modelo Fisico Almacen de datos



6.10. DICCIONARIO DE DATOS

Para el desarrollo del diccionario de datos, se utilizó la herramienta de modelación Power Designer en su versión 16.1.0.3637 herramientas que permite generar el diccionario de datos utilizando el modelo Físico.

6.11. ESTANDARES

Las bases de datos multidimensionales por sus características no existen patrones que guíen al diseño y construcción del modelo

6.12. PANTALLAS

Se describe gráficamente el sistema en marcha y la información como es representada, para ello se elaboraron mediante informes con la herramienta de Reporting Services perteneciente a la suite de desarrollo de Microsoft



ID NIVEL	ID MATERIA	ID PERIODO	MATRICULADOS	PROMOVIDOS	Porcentaje Promovidos Value
Iro					
	ALG TRI				
		1S	224	139	62.0555714285714
		2S			
	ELECT				
		1S			
		2S	188	101	53.7234042553191
		2S			
			199	137	68.8442211055276
	HTML				
		1S			
			228	126	55.2631578947368
		2S			
			193	117	60.6217616580311
	ENGI				
		1S			
			188	105	55.8510638297872
		2S			
			181	131	72.3756906077348
	INT SIS				
		1S			
			214	131	61.214953271028
		2S			
			201	106	52.736318407960

Gráfico No.10 Modelo del Sistema

6.13. INSTALACION DEL SISTEMA

La instalación del sistema corresponde a la instalación del motor de base de datos SQL Server 2008 y los respectivos paquetes necesarios para desarrollar la solución, tales como de Integration Services que será el encargado de generar un paquete de integración de información, que alimentara al Datawarehouse y Analisis Services que tendrá la responsabilidad de realizar los cálculos y generar los indicadores de rendimiento, a continuación se describe gráficamente los pasos de su instalación.



Instalación de una nueva Base de Datos

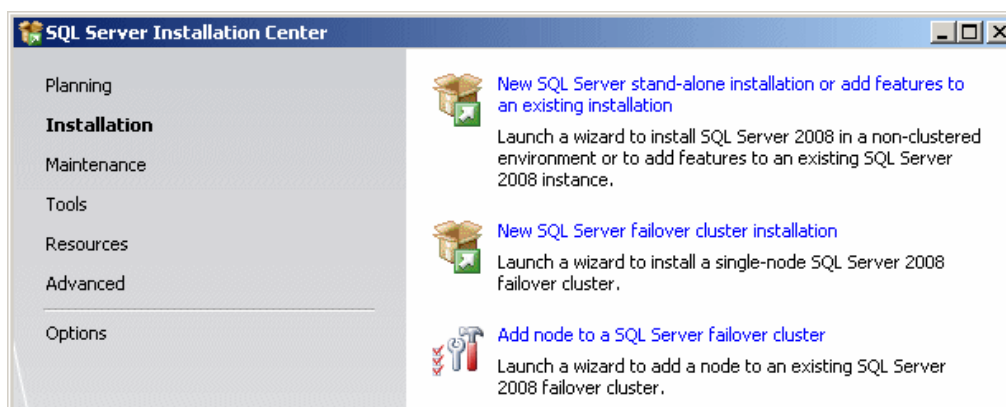
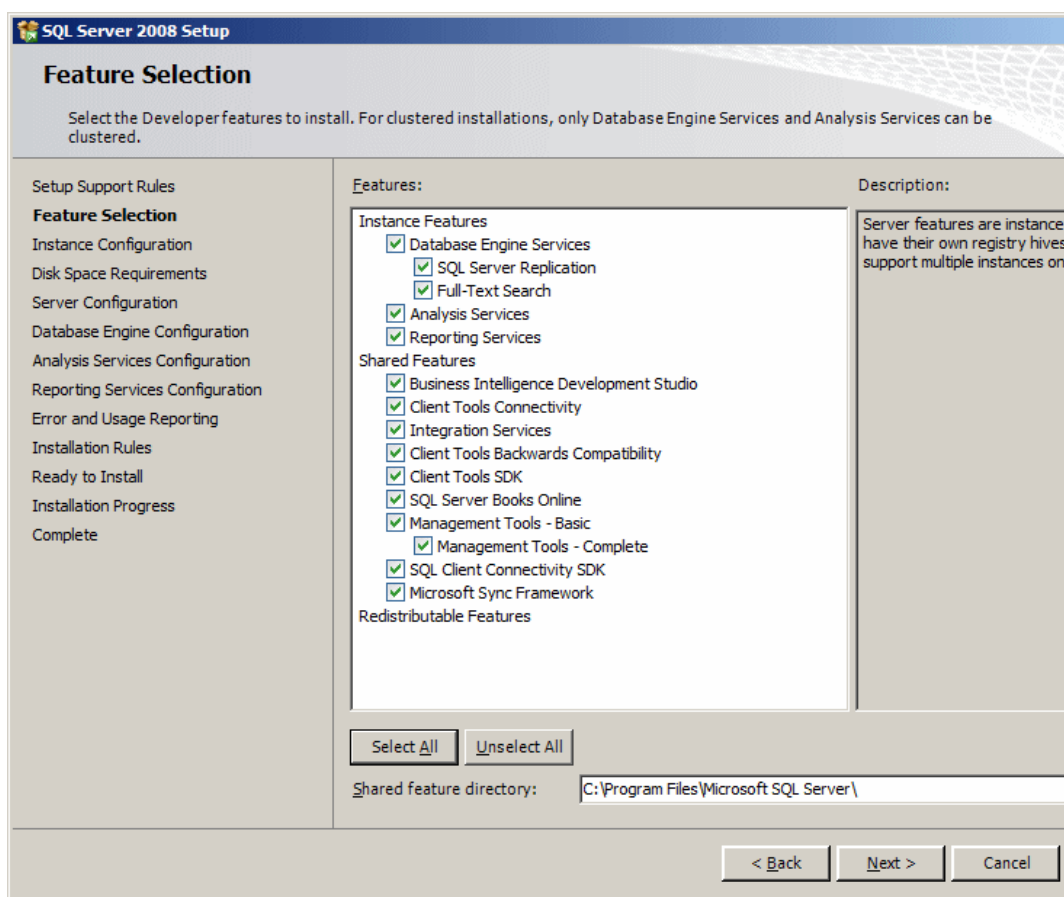


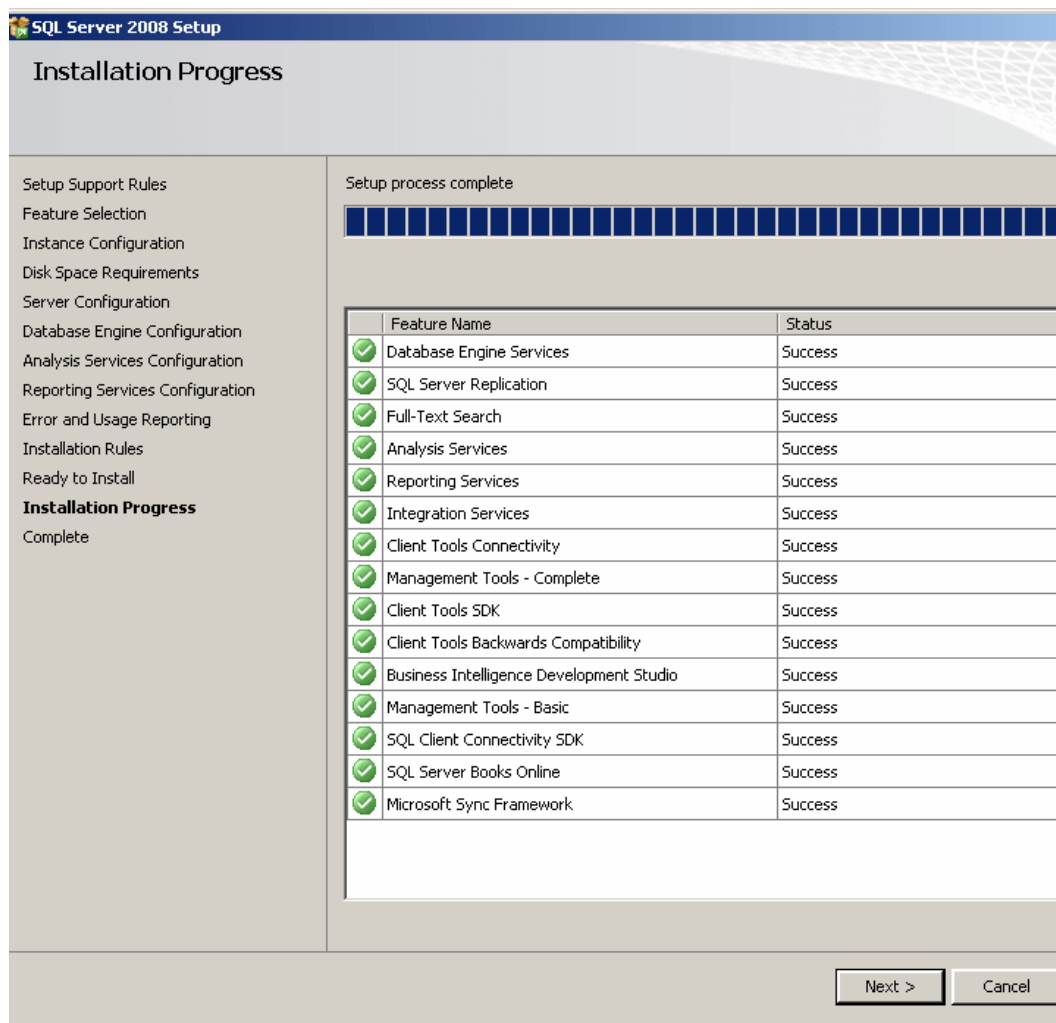
Gráfico No.11 Ventana detalle de la instalación de Base de Datos

Instalación paquetes de Integration Services y Analisis Services y Motor de BDD





Instalación completa del motor de Base de Datos y sus respectivos paquetes





6.14. RECOPIACION Y CARGA DE DATOS

Por parte del investigador, se desarrollo un paquete de integración de datos para la carga de información necesaria para el Datawarehouse, los datos fueron dentro y la carga de información se la realizo mediante el desarrollo de paquetes de integración de datos, con la herramienta Integration Services de SQL Server 2008 describiendo gráficamente el paquete

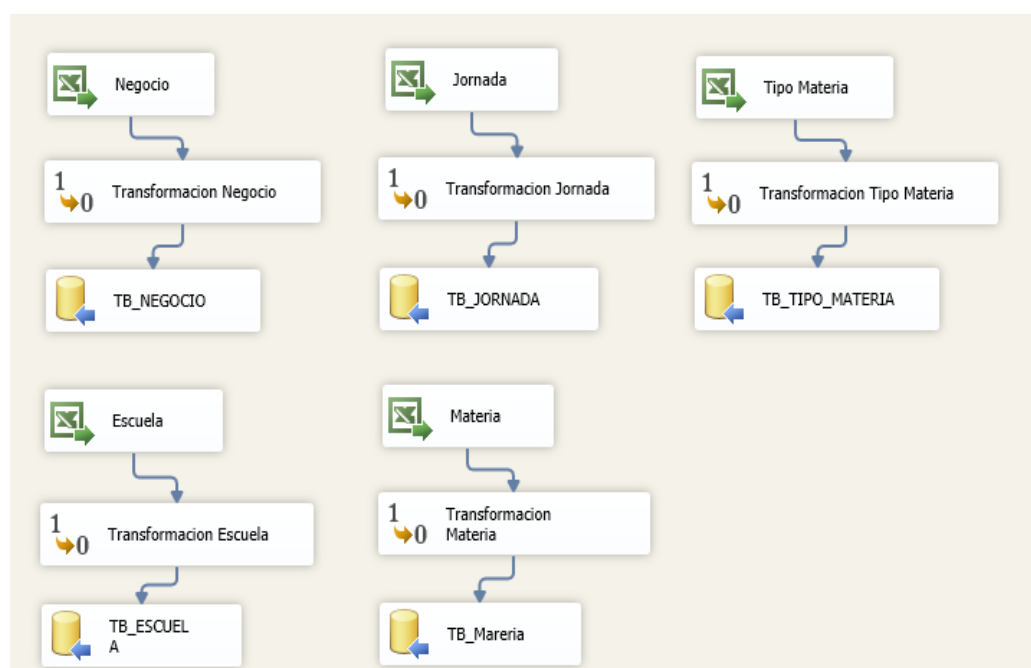


Gráfico No.12 Paquete de flujo de datos

El Paquete de flujo de datos describe las siguientes tareas:

- Origen de datos: esta tarea realiza una conexión al archivo Excel y recupera los datos que contenga, permitiendo seleccionar solo los campos necesarios
- Transformación: La siguiente tarea ejecuta una conversión de datos de tipo Unicode ya que no son compatibles con el tipo de dato del modelo DW



- Carga de datos: Esta tarea ejecuta una conexión hacia el destino que es la base de datos multidimensional, antes de realizar la carga se debe mapear los campos del origen con el destino.

6.15. PRUEBAS Y DEPURACION

6.15.1. PRUEBAS DE UNIDAD

Se concentra en la verificación de la unidad más pequeña del diseño del software o componente, aplicando el concepto al proyecto de inteligencia de negocios, definimos a las tablas de hechos que contienen los indicadores, por el motivo de que el modelo multidimensional es la base de las complejas consultas e históricas que debe ejecutar, por lo que se realizaron las respectivas pruebas, comprobando que los indicadores contengan las cantidades y sumalizaciones proporcionen la información correcta al usuario final.

6.15.2. PRUEBAS DE INTEGRACION

En este punto se realizaron las pruebas a los proceso de extracción, transformación y carga que se encuentren ejecutando correctamente para la población de los datos al Datawarehouse, comprobando que los paquetes desarrollados en SQL Server en la extensión de Integration Seviles no generen inconsistencias en la población al DW, concluyendo en que los paquetes tienen la robustez suficiente para ejecutar la tarea.



6.15.3. PRUEBAS DE VALIDACION

Las pruebas de validación se ejecutan cuando terminan las pruebas de integración de información, comprobando que la consistencia de la información entregada al usuario final.

6.15.4. PRUEBAS DEL SISTEMA

Las pruebas del sistema se realizaron en conjunto con el usuario final, es decir comprobar que el sistema sea lo suficientemente intuitivo para un usuario no experto con una interfaz grafica amigable y fácil de entender utilizando gráficos para la presentación de indicadores, utilizando semáforos que son utilizados por expertos del negocio que miden la varianza de los indicadores.

6.15.5. PRUEBAS DE SEGURIDAD

Las pruebas que se realizaron fueron para comprobar que el acceso a la información del negocio, no pueda ser vista por usuarios no autorizados, ya que se utilizara herramientas que son de uso comercial, existe un modulo de seguridad definido y las pruebas no son aplicables al proyecto ya que los esfuerzos se concentran en la creación del Datawarehouse

6.15.6. PRUEBAS DE INTERFAZ GRAFICA

Las pruebas de interfaz grafica fueron realizadas en conjunto con el usuario final, determinado que la interfaz debe ser intuitiva ya que la concentración que debe tener el usuario final será en los indicadores y la manera en que se los representen, debe ser amigable y de fácil interpretación.



6.16. PUESTA EN MARCHA DEL SISTEMA

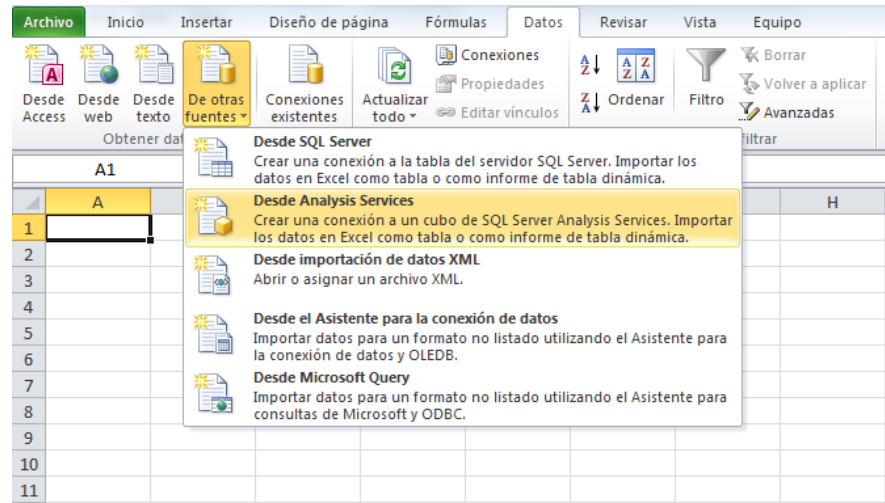
Para que el sistema funcione dentro de una estacion de trabajo cliente, se debe tomar en cuenta que debe tener una herramienta de visualizacion de datos Olap. Microsoft como parte de su suite, tiene a disposicion la herramienta Excel que permite analizar los datos de una manera sencilla para el usuario y representar graficamente y en forma de tabla sus indicadores de rendimiento. Desde un punto de vista viable, se selecciono la herramienta bajo el punto de vista del usuario, ya que la finalidad es la interpretacion de la informacion y la toma de desiciones, Microsoft en su trayectoria a generado una cultura con la herramienta Excel se ha convertido en el mejor interpretador y gestor de informacion, es por ello que se va a incluir dentro de la herramienta la informacion para su respectivo analisis.

Para ello es necesario realizar una conexión al Cubo Rendimiento Academico que contiene la informacion y los indicadores de rendimiento

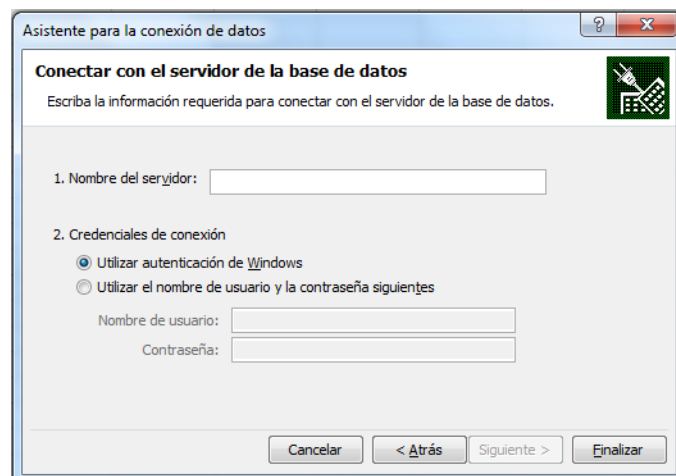


INSTITUTO TECNOLÓGICO SUPERIOR CORDILLERA

Los pasos para generar un conexión al servidor de Sql Server, al servicio de Analysis Services es necesario realizar una configuración que realizar un string de conexión con el origen de datos.



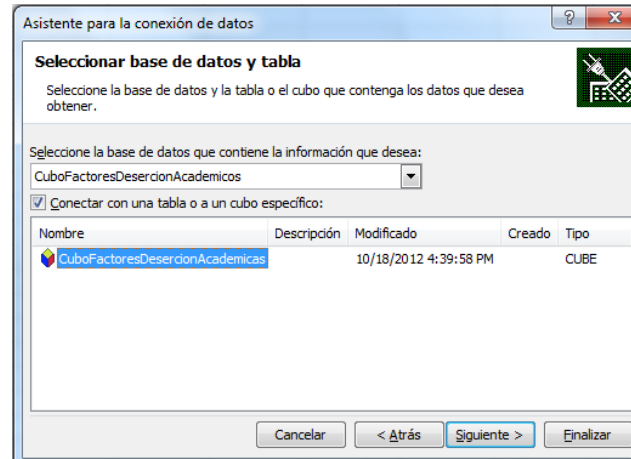
Posteriormente direccionamos el servidor, para el caso del proyecto se encuentra en un ambiente local, las credenciales pueden ser configuradas en función a la administración de usuarios y logins que se hayan hecho en el Management Studio de Sql Server



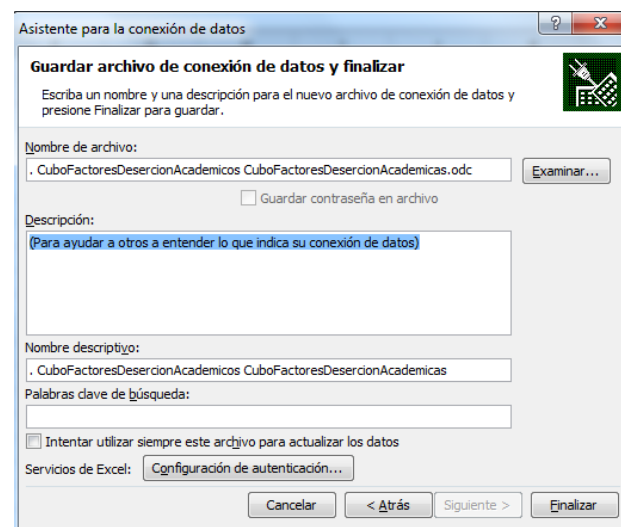


INSTITUTO TECNOLÓGICO SUPERIOR CORDILLERA

El siguiente paso es seleccionar el cubo de informacion o la base de datos multidimensional que contiene la informacion de los estudiantes



Finalmente configuramos si deseamos que la conexión este presente siempre





6.17. CAPACITACION TECNICA

La capacitación técnica es bastante clara y comprensible tanto para usuarios técnicos como para usuarios expertos en el negocio, ya que se manejan términos y se utiliza un lenguaje utilizado para administradores de negocios, además el modelo de Datawarehouse es sencillo de interpretar

La cual debe contener y realizar un recorrido de los siguientes puntos:

- Descripción del modelo DW
- Descripción de indicadores de negocio
 - o Calculo de indicadores
- Descripción de perspectivas de análisis
 - o Agregación de perspectivas
 - o Niveles de detalle de información



6.18. CONCLUSIONES Y RECOMENDACIONES

El sistema determina un factor importante dentro de una organización y genera una competencia ante el resto, ya que permite tener información de lo que está sucediendo, y de lo que puede llegar a pasar, determinar tendencias entre otras competencias, y permite que el proceso de toma de decisiones se convierta en un proceso claro y transparente, dejando a un lado la incertidumbre que existía al no poseer la información necesaria, ya que la toma de decisiones no puede solo basarse en la experiencia que tenga el usuario experto del negocio, en síntesis el sistema se convierte en un apoyo a la toma de decisiones para el usuario experto del negocio.

Es importante que se de el respectivo seguimiento al proyecto para que se pueda dar continuidad, y determinar factores de decisión y perfiles de estudiantes con la finalidad de poder tomar decisiones acertadas y proponer estrategias que bajen el porcentaje de estudiantes con perfiles de deserción y bajo rendimiento.



6.19. BIBLIOGRAFIA

1. Frase: Ricardo Hirata
2. Frase: Jean Piaget
3. Datawarehosing.- Nos ayuda a cumplir con sus objetivos, extrae, transforma y consolida, sea almacenada de manera centralizada en una base de datos. (HEFESTO, 2010).
4. RAMOS, Salvador - Ve a el cubo medio lleno – Editorial SolidQ 2011
5. Ing. BERNABEU, Dario – Metodologia para la construcción de un Datawarehouse – Editorial Instituto Universitario Aeronautico 2010
6. RAMOS, Salvador - Ve a el cubo medio lleno – Editorial SolidQ 2011
7. <http://wikipedia.org>] Wikipedia.
8. <http://sistemasdecisionales.blogspot.com>] Sistemas Decisionales, algo más que Business
9. Intelligence – Jorge Fernández González
10. [<http://todobi.blogspot.com>] Todo BI: Business Intelligence, Data Warehouse, CRM y mucho más....
11. [<http://www.intelineg.com>] Inteligencia de Negocio – Rémi Grossat.
12. KEPER, Charles – Análisis de Toma de Decisiones Datawarehouse – Editorial McGraw – Hill Años 2010.
13. [<http://todobi.blogspot.com>] Todo BI: Business Intelligence, Data Warehouse, CRM y mucho más....
14. Ley de Propiedad Intelectual Ecuatoriana



ANEXOS



ANEXO No. 1: Encuesta

En la ciudad de Quito en la provincia de Pichincha en el Instituto Tecnológico Superior Cordillera se realizara la siguiente encuesta, por favor se le pide total veracidad en sus respuestas:

10. Cual ha sido su rendimiento académico hasta la actualidad?

- ☐ Regular
- ☐ Bueno
- ☐ Sobresaliente

11. Que aspectos influyen en el bajo rendimiento?

- ☐ Profesores
- ☐ Falta de tiempo
- ☐ Descuido del estudiante

12. A pensado alguna vez abandonar sus estudios?

- ☐ Si
- ☐ No

13. Que lo motivo para tomar la decisión de abandonar sus estudios?

- ☐ Falta de dinero
- ☐ Falta de tiempo
- ☐ Rendimiento académico
- ☐ Materias con arrastre
- ☐ Otros



Especifique: _____

—

14. En que status social se encuentra actualmente?

- ☐ Bajo
- ☐ Medio
- ☐ Alto

15. Quien paga sus estudios actualmente?

- ☐ Padres
- ☐ Autogestión
- ☐ Otros

Especifique:

16. Cuantas materias ha tenido en arrastre?

- ☐ 1-3
- ☐ 4-8
- ☐ Ninguno

17. Cuantas materias con tercera matricula ha tenido?

- ☐ 1-3
- ☐ 4-8
- ☐ Ninguno

18. Le gusta la carrera que eligió?



INSTITUTO TECNOLÓGICO SUPERIOR CORDILLERA

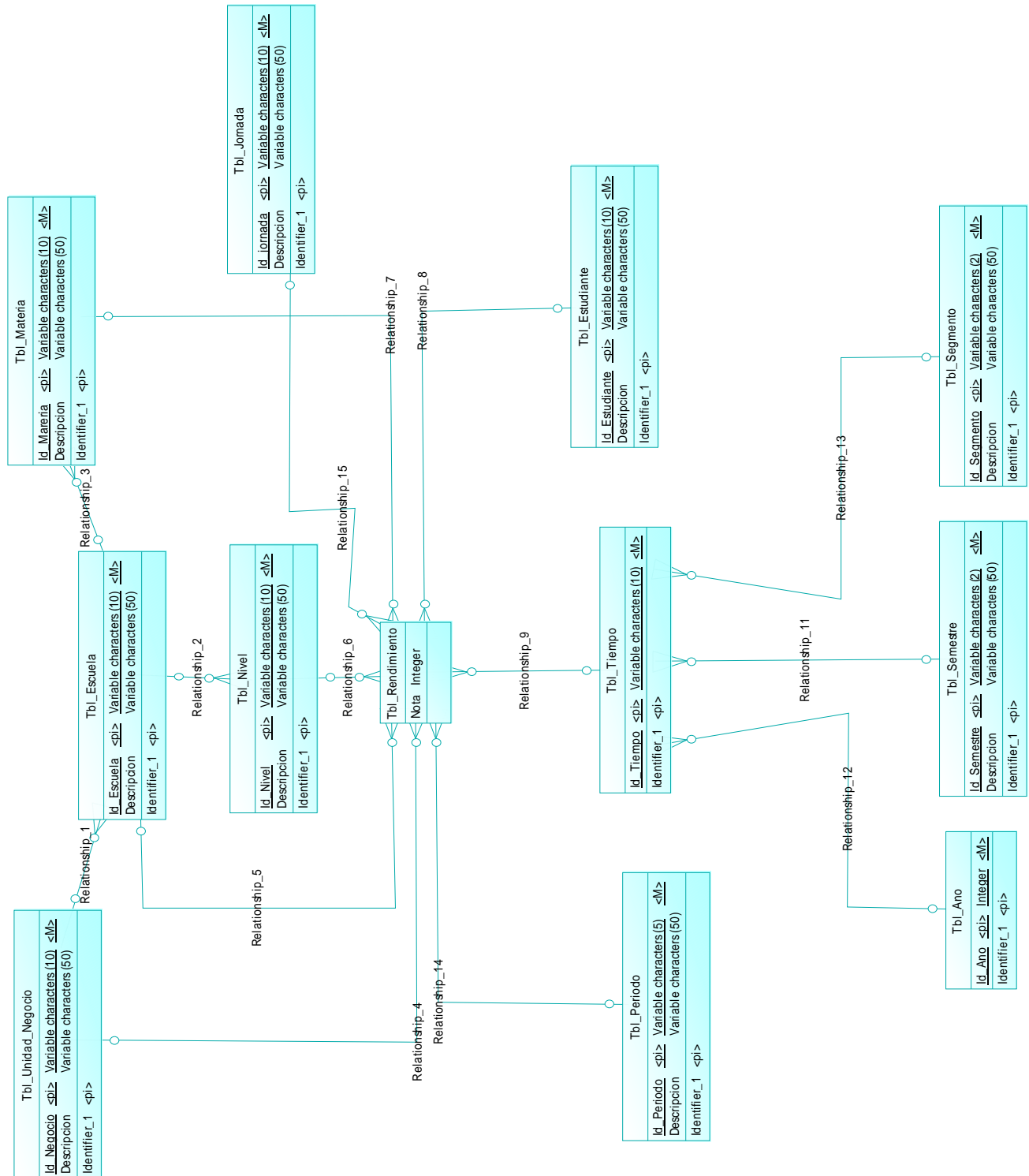
☐ Si

☐ No

GRACIAS POR SU COLABORACION

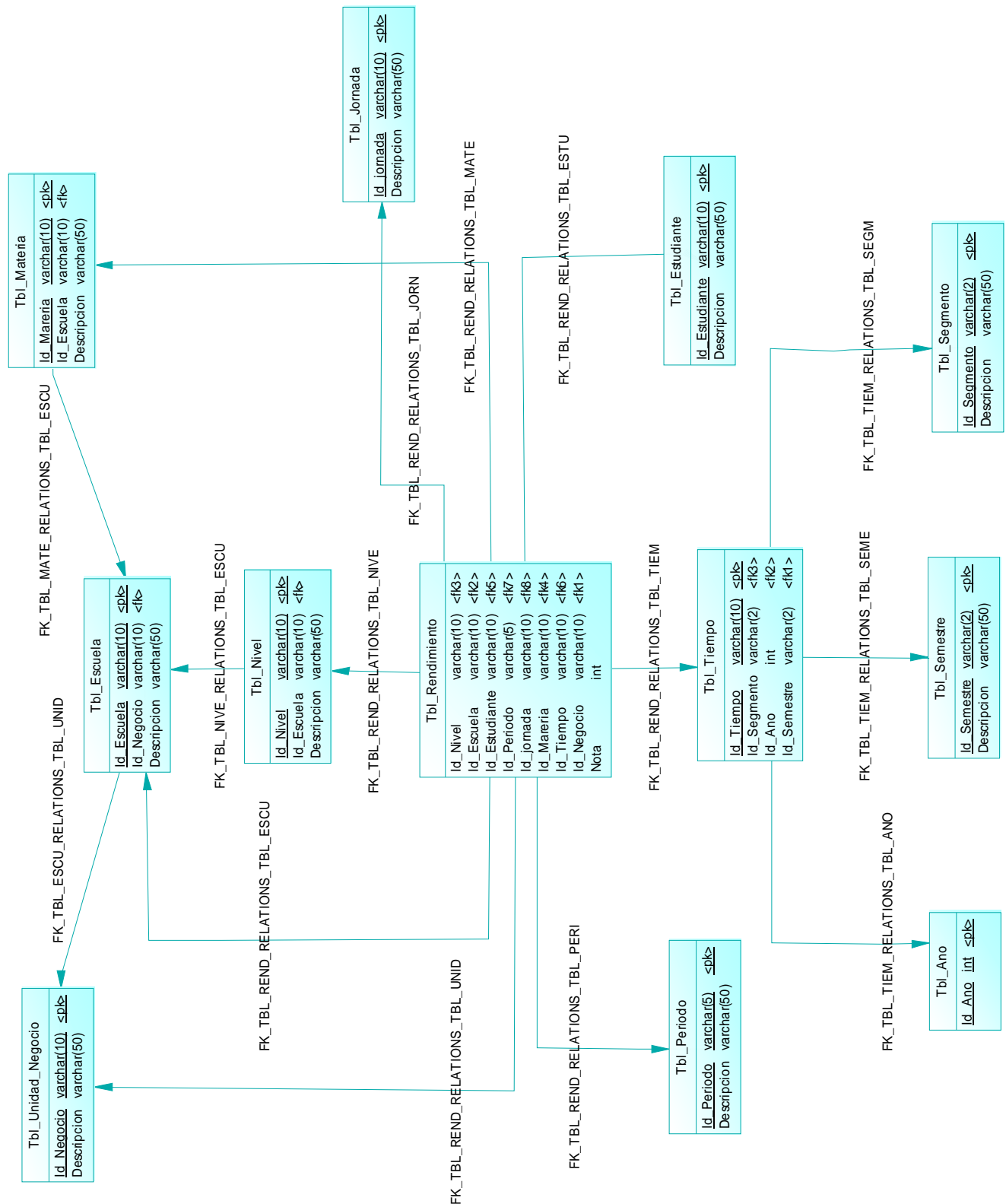


ANEXO No. 2: Modelo Logico Datawarehouse Rendimiento Notas





ANEXO No. 3: Modelo Físico Datawarehouse Rendimiento Notas





ANEXO No. 4: Diccionario de Datos Datawarehouse Rendimiento Notas

Card of model Fisico_DW_Rendimiento

Name	Fisico_DW_Rendimiento
Code	FISICO_DW_RENDIMIENTO
DBMS	Microsoft SQL Server 2005
Comment	Modelo fisico que se encarga de recibir los datos de multiples fuentes e integrarlos en un solo almacen y su objetivo es alimentar de informacion al datawarehouse
Creator	USER
Modification Date	Monday, September 10, 2012 10:47:33 AM
Author	Daniel Benitez

List of tables

Name	Code	Comment
Tbl_Ano	TBL_ANO	tabla que permite analizar la informacion en el tiempo diferenciado por anios, su sintaxis es un numero entero que contiene un anio determiando
Tbl_Escuela	TBL_ESCUELA	tabla que permite diferenciar la escuela a la que pertenece un estudiante, y obtener un rendimiento estudiantil por escuelas
Tbl_Estudiante	TBL_ESTUDIANTE	tabla que pemite analizar la informacion por estudiante y conocer su rendimiento, sumado otras dimensiones, en esta tabla contiene criterios de analisis y factores influyentes en la desercion estudiantil
Tbl_Jornada	TBL_JORNADA	tabla que permite diferenciar a los estudiantes por jornada
Tbl_Materia	TBL_MATERIA	tabla que permite analizar el rendimiento de los estudiantes por materias
Tbl_Nivel	TBL_NIVEL	tabla que permite diferenciar los niveles que ha cursado un



Tbl_Periodo	TBL_PERIOD O	estudiante dimension que permite analizar las notas por periodos de tiempo, la sintaxis que presenta: P1B --> periodo del primer biemestre P2B --> periodo del segundo biemestre P3B --> periodo del tercer biemestre P1T --> periodo del primer trimestre P2T --> periodo del segundo trimestre
Tbl_Rendimiento	TBL_RENDIMIENTO	tabla que contiene el indicador rendimiento de los estudiantes
Tbl_Segmento	TBL_SEGMENTS	tabla que permite analizar la informacion en el tiempo diferenciado por bimestres o trimestres segun el criterio de analisis, su sintaxis esta compuesta por: 1b --> corresponde al primer biemestre de un anio y semestre determinado 1t --> corresponde al primer trimestre de un anio y semestre determinado
Tbl_Semestre	TBL_SEMESTRE	tabla que permite analizar la informacion en el tiempo diferenciado por semestres, su sintaxis esta compuesta por: 1s --> primer semestre de un determinado anio y segmento 2s --> segundo semestre de un determiando anio y segmento
Tbl_Tiempo	TBL_TIEMPO	Tabla master que contiene la dimension de tiempo y es alimentada por sus tablas hijas
Tbl_Unidad_Negocio	TBL_UNIDAD_NEGOCIO	tabla que contiene y hace diferencia entre la matriz y sucursales, esta tabla puede ir creciendo segun el negocio avance en el tiempo y permite agregar nuevas sucursales



INSTITUTO TECNOLÓGICO SUPERIOR CORDILLERA

Table Tbl_Ano

List of columns of the table Tbl_Ano

Name	Code	Data Type	Length	Primary	Foreign Key
Id_Ano	ID_ANO	Int		X	

Table Tbl_Escuela

List of columns of the table Tbl_Escuela

Name	Code	Data Type	Length	Primary	Foreign Key
Id_Escuela	ID_ESCUELA	varchar(10)	10	X	
Id_Negocio	ID_NEGOCIO	varchar(10)	10		X
Descripcion	DESCRIPCION	varchar(50)	50		

Table Tbl_Estudiante

List of columns of the table Tbl_Estudiante

Name	Code	Data Type	Length	Primary	Foreign Key
Id_Estudiante	ID_ESTUDIANT E	varchar(10)	10	X	
Descripcion	DESCRIPCION	varchar(50)	50		



INSTITUTO TECNOLÓGICO SUPERIOR CORDILLERA

Table Tbl_Jornada

List of columns of the table Tbl_Jornada

Name	Code	Data Type	Length	Primary	Foreign Key
Id_jornada	ID_JORNAD	varchar(10)	10	X	
Descripcion	DESCRIPCION	varchar(50)	50		

Table Tbl_Materia

List of columns of the table Tbl_Materia

Name	Code	Data Type	Length	Primary	Foreign Key
Id_Materia	ID_MARERI	varchar(10)	10	X	
Id_Escuela	ID_ESCUEL	varchar(10)	10		X
Descripcion	DESCRIPCION	varchar(50)	50		

Table Tbl_Nivel

List of columns of the table Tbl_Nivel

Name	Code	Data Type	Length	Primary	Foreign Key
Id_Nivel	ID_NIVEL	varchar(10)	10	X	
Id_Escuela	ID_ESCUEL	varchar(10)	10		X
Descripcion	DESCRIPCION	varchar(50)	50		



INSTITUTO TECNOLÓGICO SUPERIOR CORDILLERA

Table Tbl_Periodo

List of columns of the table Tbl_Periodo

Name	Code	Data Type	Length	Primary	Foreign Key
Id_Periodo	ID_PERIODO	varchar(5)	5	X	
Descripcion	DESCRIPCION	varchar(50)	50		

Table Tbl_Rendimiento

List of columns of the table Tbl_Rendimiento

Name	Code	Data Type	Length	Primary	Foreign Key
Id_Nivel	ID_NIVEL	varchar(10)	10		X
Id_Escuela	ID_ESCUELA	varchar(10)	10		X
Id_Estudiante	ID_ESTUDIANTE	varchar(10)	10		X
Id_Periodo	ID_PERIODO	varchar(5)	5		X
Id_jornada	ID_JORNADA	varchar(10)	10		X
Id_Mareria	ID_MARERIA	varchar(10)	10		X
Id_Tiempo	ID_TIEMPO	varchar(10)	10		X
Id_Negocio	ID_NEGOCIO	varchar(10)	10		X
Nota	NOTA	Int			



Table Tbl_Segmento

List of columns of the table Tbl_Segmento

Name	Code	Data Type	Length	Primary	Foreign Key
Id_Segmento	ID_SEGMENTO	varchar(2)	2	X	
Descripcion	DESCRIPCION	varchar(50)	50		

Table Tbl_Semestre

List of columns of the table Tbl_Semestre

Name	Code	Data Type	Length	Primary	Foreign Key
Id_Semestre	ID_SEMESTRE	varchar(2)	2	X	
Descripcion	DESCRIPCION	varchar(50)	50		

Table Tbl_Tiempo

List of columns of the table Tbl_Tiempo

Name	Code	Data Type	Length	Primary	Foreign Key
Id_Tiempo	ID_TIEMPO	varchar(10)	10	X	
Id_Segmento	ID_SEGMENTO	varchar(2)	2		X
Id_Ano	ID_ANO	Int			X
Id_Semestre	ID_SEMESTRE	varchar(2)	2		X



Table Tbl_Unidad_Negocio

Name	Code	Data Type	Length	Primary	Foreign Key
Id_Negocio	ID_NEGOCIO	varchar(10)	10	X	
Descripcion	DESCRIPCION	varchar(50)	50		