

MONITOR DE RENDIMIENTO ACADÉMICO

Abraham Granados Herrera

Instituto Tecnológico Superior del Sur de Guanajuato
abran.killer@gmail.com

José Roberto Custodio Zamudio

Instituto Tecnológico Superior del Sur de Guanajuato
robcz10@gmail.com

Azucena Rico Torres

Instituto Tecnológico Superior del Sur de Guanajuato
azu.ne12@gmail.com

Resumen

El estudio de los Data Warehouse nos ha llevado a pensar en implantar uno dentro del Instituto Tecnológico del Sur de Guanajuato, esto surgió debido al gran número de alumnos que desertaron de la carrera de Ing. Informática llevándola al punto que cerrarla y no promoverla más, para evitar este suceso nuevamente, se pensó realizar un monitor académico implantando esta tecnología, que a dado muy buenos resultados en los campos empresariales, ya que ha mejorado los procesos en la toma de decisiones y generado información la cual no se tiene contemplada en ese momento, también es flexible en cuanto a la integración de nuevas necesidades que dese la institución. Para la planeación del proyecto se pensó en los requerimientos principales de las áreas las cuales estarán involucradas, también se consultaron las metodologías y estructuras que ayudarían a la creación de Data Warehouse, llegando a conclusiones de porque este sistema se implantaría perfectamente dentro de la institución y los posibles resultados que podría demostrar para los encargados de las áreas involucradas

Palabra(s) Clave(s): 3 a 5 palabras claves en español, ordenadas alfabéticamente.
Data Warehouse (DWH), Data Mart (DM) y Instituto Tecnológico del Sur de Guanajuato (ITSUR)

Abstract

The study of the Data Warehouse us led to think implement one within the Technological Institute of Southern Guanajuato, this came about because of the large number of students who dropped out of the race of Ing. Informática taking her to the point that close and not promote it more, to avoid this happening again, he thought to do an academic monitor implementing this technology, which has given very good results in business fields, since improved the process in decision-making and generated information which was not contemplated at the time it is also flexible in integrating new needs that dese the institution. For project planning was thought the main requirements of the areas which will be involved, methodologies and structures that help create Data Warehouse, reaching conclusiones of also consulted because this system perfectly implanted within the institution and the possible outcomes that could prove to managers of the areas involved.

Keywords: *3 a 5 palabras claves en inglés, ordenadas alfabéticamente en letra cursiva: Data Warehouse (DWH), Data Mart (DM) y South Instituto Tecnológico de Guanajuato (ITSUR).*

1. Introducción

En México la educación superior es de gran importancia, ya que es el nivel en el cual se les prepara a los jóvenes para salir al mundo laboral y que estos cuenten con las capacidades necesarias para enfrentar los retos, y circunstancias que nuestro país llegue a presentar. Un estudio de la ANUIES en los dos últimos años se registró un aumento en cuanto a los alumnos titulados en el estado de Guanajuato, pero cabe mencionar que el número de alumnos no corresponde al total de alumnos que ingresaron a las instituciones públicas, privadas y demás sectores de estudio al igual con el análisis se ha observado el aumentado los índices de deserción escolar. Se pensó en realizar un sistema de monitor académico que pudiera generar información sobre la vida del alumno en cuanto a sus estudios se refiera como calificaciones, créditos, materias, etc.

Para ello se pensó en la implantación de un Data Warehouse ya que los resultados observados en función a nivel empresarial y con gran cantidad de datos, esto se

podría implementar perfectamente, para hacer llegar la información necesaria a las áreas las cuales lo necesiten, sin problemas en cuanto a redundancia de datos y a la información innecesaria que se muestre. El sistema sería ágil, rápido y lo principal manejaría un sistema histórico, el cual ayudaría a la institución en cuanto a su desempeño. Por último es importante destacar que dentro de los objetivos del proyecto no se contempla la realización de la parte visual y de interacción de los usuarios con el sistema que corresponde con las aplicaciones de usuario final. De esta manera los puntos del proyecto que se refieren al desarrollo de éstas y a la implantación real del sistema se limitan a ser ligeras recomendaciones propias de la metodología seleccionada, ya que estos apartados exceden los límites del trabajo planteados al comienzo del mismo.

2. Métodos

El interés de este proyecto se va a centrar en analizar las etapas por las que puede pasar el alumno en cuanto a su rendimiento académico: desde su comienzo en los estudios hasta llegar a terminar la carrera, pasando por la superación de cada uno de los semestres. También será objeto de análisis de las razones por las que el alumno tiene dificultades a la hora de superar algún semestre en concreto dentro de la carrera y detectar focos de deserción por parte de los alumnos y en casos muy extraños la producción de ciertos alumnos que ni abandonan ni terminan la carrera pero su expediente se mantiene. Un tecnológico universitario se compone de un gran número de carreras en las que podemos realizar el estudio y además, las distintas carreras se encuentran en diferentes áreas en las que se desarrollan, con estos datos, se podrán sacar patrones en la toma de decisiones del alumnado dentro de cada área y los porcentajes de éxito, abandono, cambios de carrera o cambios de institución. Por otro lado las carreras incluyen un gran número de asignaturas a lo largo de la misma, lo cual nos da un lado que se encuentran orientadas dependiendo de la carrera. Esto hace el hecho de que se impartan diferentes tipos de asignaturas de una carrera puede hacer que los alumnos tengan mayor o menor número de dificultades para superarlas. En cuanto al tipo de alumno, está claro que no todos tienen las ideas claras sobre la carrera en que se matriculan,

lo que hace que en un cierto número de casos, la tasa de abandono del alumno en ciertas carreras universitarias debido a esta razón. En el caso de los alumnos que terminan sus estudios es también interesante saber la tasa de aprobados y las calificaciones dependiendo de la carrera, el tipo de alumno que se quiera observar, el número de matriculados en la asignatura, el semestre en el que se imparten las asignaturas y la convocatoria del examen, todo ello puede influir en las tasas de aprobados, ya que puede haber alumnos que saquen más rendimiento al estudio. Como conclusión, el proyecto debería ofrecer información útil sobre el ciclo académico de los alumno para así prevenir la deserción y tomar decisiones para la mejora de los alumnos, además el sistema permitirá demostrar información sobre el ciclo académico para detectar cuellos de botella y porcentajes de éxito entre otras cosas, y poder brindar un mejor grado de satisfacción para los alumnos y la calidad de estudio.

Centrándose en el data Warehouse se pensó en la revisión de los conceptos fundamentales asociados con la tecnología, con la finalidad de recopilar los conocimientos básicos que nos permitan desarrollarlo, analizando propuestas de metodología más habituales para el desarrollo, y tener el mayor aprovechamiento e implementación [5] un DWH es un almacén de datos nuevos y antiguos, con ciertas propiedades con el que la empresa puede realizar un profundo análisis de su organización y apoyar sus decisiones de negocio sobre las consultas realizadas a partir de dichos datos. Este almacén de datos contiene toda la información relativa a una misma temática de la empresa con el fin de agrupar los datos y unirlos entre sí. Un DWH es un repositorio de datos, Orientado a temas, Variante en el tiempo, No volátil e Integrado y no con lleva exclusivamente el hecho de que se realice una copia masiva de datos, sino que esa copia tiene un determinado fin y que su propia existencia involucra una dinámica de trabajo diferenciada. Tomando en cuenta que uno de los procesos característicos de un entorno de DWH sea el proceso de "llenado", ya que la información del DWH proviene de los datos existentes en las bases de datos operacionales de la organización. Asiendo así otra característica de un DWH resida en que el diseño debe orientarse a las nuevas necesidades de tiempo de vida útil mínimo de la base de datos y a que los datos son no volátiles, lo

que hace de la fecha un atributo importante a la hora de convertir en "históricos" a dichos datos. El usuario del DWH no efectúa operaciones de escritura sobre él, ya que solamente se carga con el proceso de llenado en los sistemas de datos operacionales.

Por último, hay que decir que el acceso al DWH es diferente al tradicional, y que, otra vez debido a ese gran volumen, se debe dedicar un gran esfuerzo a optimizar el acceso a los datos por parte de las consultas. Pequeños fallos de diseño en las consultas adquieren ahora gran relevancia dado que se notará mucho la falta de optimización. Todo esto nos lleva a poder representar el flujo de información existente en un DWH como un flujo lineal que comienza en las bases de datos tradicionales y que tiene como característica que la escritura o actualización se produce en un punto muy concreto y que el otro extremo solamente efectúa operaciones de lectura.

Sin embargo, los DWH se diferencian en muchos factores con respecto a los entornos operacionales, lo que les hace tener una mayor potencia a la hora de realizar búsquedas sobre los datos en entornos orientados a la toma de decisión.

Por todos los aspectos descritos, se puede decir que las propiedades de los DWH los hacen especialmente idóneos para estudiar el comportamiento del alumnado universitario a lo largo del proceso educativo ya que permiten su análisis desde diferentes perspectivas, así como valorar los resultados académicos del mismo. Este análisis permitirá obtener información acerca de qué carreras son más complejas mediante porcentajes de aprobados, tasas de abandono universitario, fechas de finalización de las carreras por parte de los alumnos y multitud de datos relevantes acerca del proceso educativo en estudio.

Así que para la construcción e implantación de un DWH involucra el proceso evolutivo que se tiene que apoyar en una metodología específica para este tipo de procesos, si bien es más importante que la elección de la mejor de las metodologías, el realizar un control para asegurar el seguimiento de la misma e identificar elementos desconocidos del entorno de negocio y transformar la información en conocimiento a partir del análisis detallado de los datos recopilados. Es importante no confundir una gran base de datos estructurada con un auténtico DWH ya que

éste trata de crear un método, un sistema para explotar auténticamente la información en beneficio del negocio, donde lo importante es la aplicación, el proceso, la manera en que se aplica el conocimiento del negocio. Esto lleva a incluir aspectos tanto del hardware como de las aplicaciones y herramientas haciendo especial énfasis en los Sistemas de Soporte de Decisiones (DSS), los cuales son herramientas de Business Intelligence (BI) enfocadas al análisis de los datos de una organización. En principio, puede parecer que el análisis de datos es un proceso sencillo y fácil de conseguir mediante una aplicación hecha a medida o uno sofisticado. Sin embargo, no es así, estas aplicaciones suelen disponer de una serie de informes predefinidos en los que presentan la información de manera estática, pero no permiten profundizar en los datos o navegar entre ellos, así el DSS una de las herramientas más emblemáticas del BI ya que, entre otras propiedades, permiten resolver gran parte de las limitaciones de los programas de gestión. Así que un proyecto de DWH no es únicamente un proyecto tecnológico, es una forma de vida de las organizaciones y así como tal, tiene que contar con el apoyo de todo el personal, tanto de los usuarios como de la alta dirección de la empresa. Dando la importancia de compartir la información, de tal manera que no solo una persona conozca el funcionamiento del negocio.

Teniendo en cuenta la información de los apartados anteriores y el interés de este proyecto en centrar el estudio en ciertos aspectos dentro del entorno universitario, la metodología de Ralph Kimball se ajusta más a lo que se quiere desarrollar al permitir la creación del DWH partiendo de los DM. En particular, debido a que el interés de este proyecto se enfoca en los aspectos académicos de los alumnos, se puede considerar que este planteamiento es útil desde el punto de vista de los departamentos universitarios que tratan con alumnos y sus progresos académicos, considerándose este grupo un subconjunto del total de departamentos y organismos de la universidad. De esta forma, el proyecto puede considerarse un primer acercamiento al BI académico a partir del estudio de un proceso de negocio concreto. La metodología de Kimball nos permite, por tanto, ofrecer soluciones en un plazo inferior al que resultaría de abordar un proyecto global destinado a toda la institución universitaria. Por otro lado, también ofrece una clara exposición de las

fases y actividades propias de cada fase, así como un buen número de ejemplos documentados en los cuales apoyarse cuando no se dispone de gran experiencia en el desarrollo de DM y DWH [7], considera que el almacén de datos tiene que ser entendido fácilmente por los usuarios y ofrecer respuestas correctas a la mayor brevedad posible. Este enfoque parte de los requisitos de negocio, mientras que el enfoque top-down propone la validación de los requisitos una vez que se tiene el sistema.

Esta metodología está definida en el libro “Building the Data Warehouse” [5]. En él proponía los mecanismos necesarios para llevar a cabo la correcta realización de un DWH. Para Bill Inmon, el diseño de un DWH comienza ya con la mera introducción de datos en el mismo, debido a las grandes cargas de datos que deben hacerse antes de su introducción en el DWH, dependiendo de ello la eficiencia de estos sistemas para acceder a los datos. El desarrollo de la metodología propuesta [5] se aprecia en la figura 1.

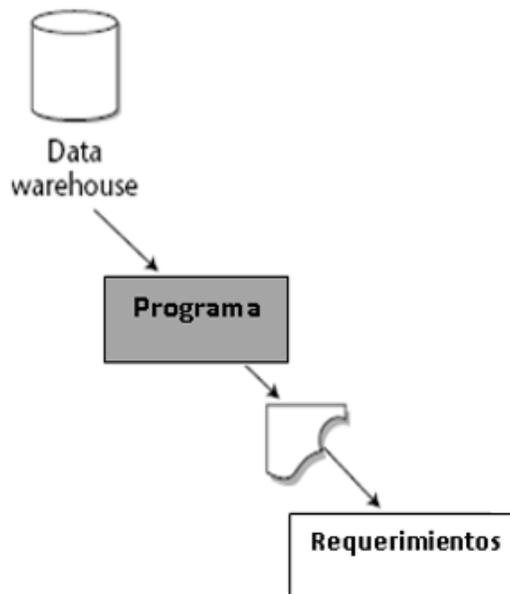


Figura 1 Desarrollo del DWH según la metodología de Bill Hinmon.

La implantación de un DWH lleva ciertos pasos:

- Extracción de los datos del sistema operacional y transformación de los mismos.

- Carga de los datos validados en el DWH. Esta carga deberá ser planificada con una periodicidad que se adaptará a las necesidades de refresco detectadas durante la fase de diseño del nuevo sistema.
- Explotación del DWH mediante diversas técnicas dependiendo del tipo de aplicación que se dé a los datos.

También existe una metodología utilizada en el modelo botton-up que es muy popular llamada “Rapid Warehousing Methology (RWM)”.dicha metodología es iterativa y está basada en el desarrollo incremental del proyecto de DWH dividido en cinco fases:

- Para la construcción del DWH no finalizara con la implantación del mismo, sino que es una tarea iterativa en la que se trata de incrementar su alcance aprendido de las experiencias anteriores.
- Después de implantarlo, se debería realizar una revisión del DWH, planteando preguntas que permitan, después de tiempo a su propuesta en marcha, definir cuáles serían los aspectos a mejorar o potenciar en función de la utilización que se haga del nuevo sistema.
- Un factor determinante en el éxito de un proceso de DWH es la interpretación correcta de los diferentes niveles de requerimientos expresados por los distintos grupos de usuarios. La técnica utilizada para revelar los requerimientos de los analistas del negocio difiere de los enfoques tradicionales guiados por los datos. Los diseñadores de los DWH deben entender los factores claves que guían el negocio para determinar efectivamente los requerimientos y traducirlos en consideraciones de diseño apropiadas.
- Los usuarios finales y sus requerimientos impactan siempre en la implementación de un DWH. Según la perspectiva de Kimball, los requerimientos del negocio se posicionan en el centro del “universo del Data Warehouse”. Como destaca siempre el autor, los requerimientos del negocio deben determinar el alcance del DWH (qué datos debe contener, cómo deben estar organizados, cada cuánto tiempo debe actualizarse, quiénes y desde

dónde accederán, etc.). La definición de los requerimientos del negocio determina los datos necesarios para cumplir los requerimientos analíticos de los usuarios.

- Diseñar los modelos de datos para soportar estos análisis requiere un enfoque diferente al usado en los sistemas operacionales. Básicamente, se comienza con una matriz donde se determina la dimensionalidad de cada indicador y luego se especifican los diferentes grados de detalle dentro de cada concepto del negocio, así como la granularidad de cada indicador y las diferentes jerarquías que dan forma al modelo dimensional del negocio (MDN) o mapa dimensional. El diseño físico de la base de datos se centra sobre la selección de las estructuras necesarias para soportar el diseño lógico. Un elemento principal de este proceso es la definición de estándares del entorno de la base de datos. La indexación y las estrategias de particionamiento se determinan también en esta etapa. Otra sería la de modelado ya que un modelo es un conjunto de conceptos, reglas y convenciones que permitirán describir y manipular los datos que queremos almacenar en una base de datos. Pero para esto se pueden utilizar varios modelos distintos para los datos y así poder describir como se usarían. Así un modelo de datos permite descubrir:
 - ✓ La estructura de datos de la base: el tipo de los datos que hay en la base y la forma en que se relacionan.
 - ✓ Las restricciones de integridad: Un conjunto de condiciones que deben cumplir los datos para reflejar correctamente la realidad deseada.

En estos modelos existen unos considerados más extendidos para la representación de los datos que se basan en el hecho de que no todas las entidades tienen igual número de ocurrencias y no presentan igual curva de crecimiento en el volumen de dichos datos.

El esquema en estrella es el más sencillo de los esquemas de almacenamiento de datos. Se llama así porque el diagrama se asemeja a una estrella, con los puntos que irradian desde un centro. El centro de la estrella consta de una o más tablas de

hechos y los puntos de la estrella son las tablas de dimensiones, como se muestra en la figura 2. En concreto este esquema en estrella es ideal por su simplicidad y velocidad para ser usado en análisis multidimensionales como los DM, ya que permite acceder tanto a datos agregados como de detalle. Además, ofrece la posibilidad de implementar la funcionalidad de una base de datos multidimensional utilizando una clásica base de datos relacional. El esquema en estrella consiste en estructurar la información en procesos, vistas y métricas a modo de estrella. En la tabla de hechos encontramos los atributos destinados al hecho que constituye el proceso de negocio a medir, es decir, sus métricas. La figura 6 muestra un ejemplo de diagrama utilizando el Modelo en Estrella [3].

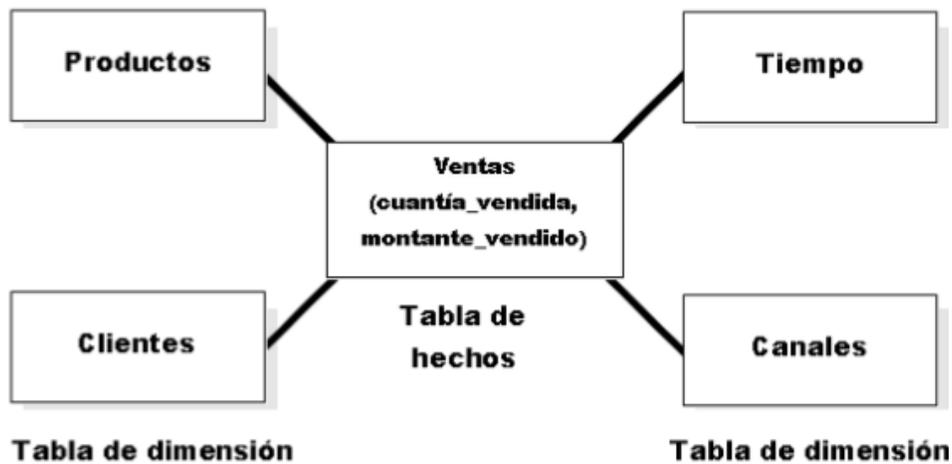


Figura 2 Ejemplo de un esquema de estrella.

También se considera el modelo de copo de nieve o Snowflake (figura 3), que es un esquema que representa la derivación del esquema en estrella, en el que las tablas de dimensión se normalizan en múltiples tablas. Por esta razón, la tabla de hechos deja de ser la única tabla del esquema que se relaciona con otras tablas, y aparecen nuevas uniones entre tablas gracias a que las dimensiones de análisis se representan ahora en tablas de dimensión normalizadas. En la estructura dimensional normalizada, la tabla que representa el nivel base de la dimensión es la que hace join directamente con la tabla de hechos. La diferencia entre ambos esquemas reside entonces en la estructura de las tablas de dimensión. Para

conseguir un esquema en copo de nieve se ha de tomar un esquema en estrella y conservar la tabla de hechos, centrándose únicamente en el modelado de las tablas de dimensión, que si bien en el esquema en estrella se encontraban totalmente desnormalizadas, ahora se dividen en subtablas tras un proceso de normalización.

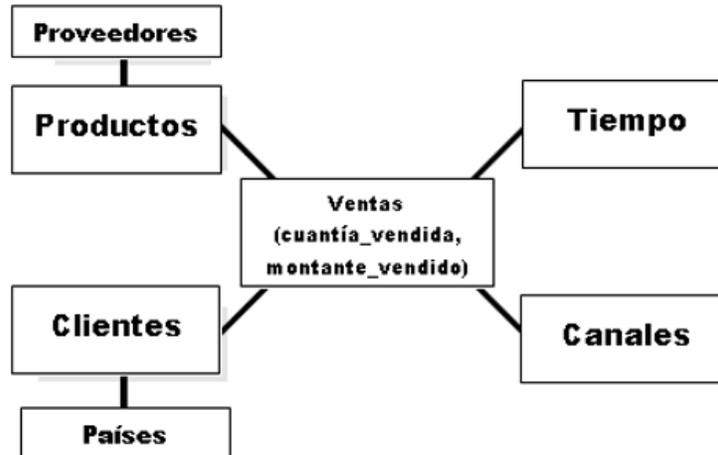


Figura 3 Ejemplo de un esquema en copo de nieve.

Es posible distinguir dos tipos de esquemas en copo de nieve, un snowflake completo (en el que todas las tablas de dimensión en el esquema en estrella aparecen normalizadas) o un snowflake parcial (sólo se lleva a cabo la normalización de algunas de ellas). En la siguiente figura se observa el modelo en copo de nieve descrito obtenido de “Data Warehousing Guide” [3].

Tomando en cuenta el modelo dimensional enfocado dentro del proceso de BI los modelos de estrella son los más empleados, también se utiliza la tercer forma normal para utilización en sistemas de almacenamiento de datos. El modelado de la Tercera Forma Normal (3FN) es una técnica clásica de las bases de datos relacionales que minimizan la redundancia de datos a través de la normalización de los datos. Cuando se compara con un esquema en estrella, un esquema 3FN tiene normalmente un mayor número de tablas debido al proceso de normalización, utilizado en los almacenes de datos grandes, especialmente en entornos con importantes requisitos de carga de datos que se utilizan para alimentar DM y ejecutar consultas de larga ejecución, figura 4.



Figura 4 Ejemplo de un esquema en 3FN.

Con la recopilación de información para la arquitectura del DWH se pudo interpretar como la organización de datos involucrados, pero cada vez con mayor frecuencia se están creando herramientas de DWH para ayudar en este proceso. En la figura 5, se pueden observar las tres principales capas de la arquitectura de un DWH, es decir, el nivel de organización, el nivel de directorio y por último el nivel de gestión de procesos desde abajo hacia arriba en la imagen.

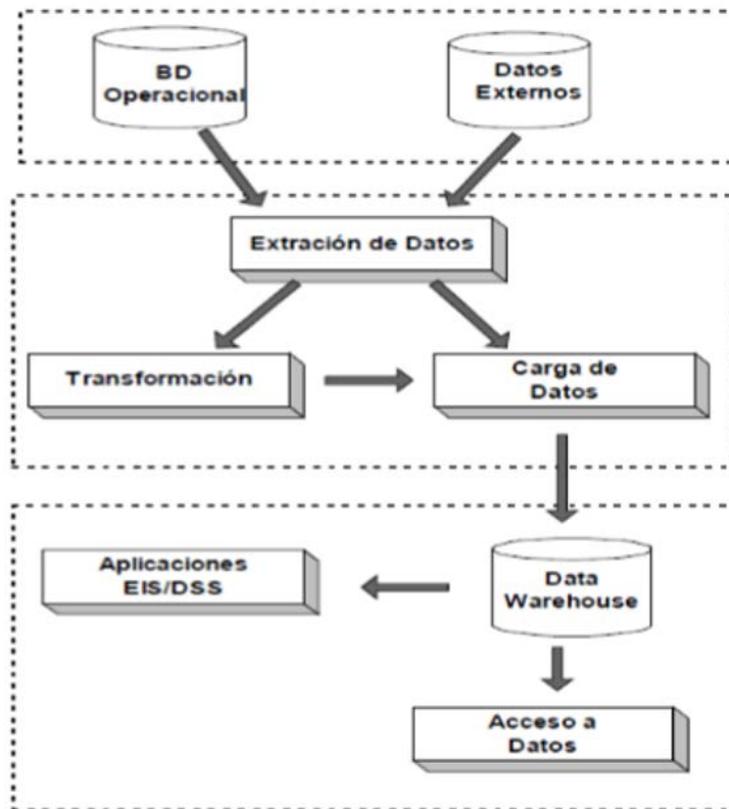


Figura 5 Estructura básica de un Data Warehouse.

Con lo analizado en todos los apartados anteriores se define perfectamente la aplicación del DWH dentro de la institución, ya que esta presentara muchas ventajas en cuanto a la toma de decisiones de las futuras carreras, alumnos y demás partes que se involucren, disminuyendo así la deserción de los alumnos y la detección de posibles casos de reprobación de los mismos.

3. Resultados

El interés de este proyecto se va a centrar en analizar las etapas por las que puede pasar el alumno en cuanto a su rendimiento académico: desde su comienzo en los estudios hasta llegar a terminar la carrera, pasando por la superación de cada uno de los semestres. También será objeto de análisis de las razones por las que el alumno tiene dificultades a la hora de superar un semestre en concreto dentro de la carrera y detectar focos de deserción por parte de los alumnos y en casos muy extraños la producción de ciertos alumnos que ni abandonan ni terminan la carrera pero su expediente se mantiene. Un tecnológico universitario se compone de un gran número de carreras en las que podemos realizar el estudio y además, las distintas carreras se encuentran en diferentes áreas en las que se desarrollan, con estos datos, se podrán sacar patrones en la toma de decisiones del alumnado dentro de cada área y los porcentajes de éxito, abandono, cambios de carrera o cambios de institución. Por otro lado las carreras incluyen un gran número de asignaturas a lo largo de la misma, lo cual nos da un lado que se encuentran orientadas dependiendo de la carrera. Esto hace el hecho de que se impartan diferentes tipos de asignaturas de una carrera puede hacer que los alumnos tengan mayor o menor número de dificultades para superarlas. En cuanto al tipo de alumno, está claro que no todos tienen las ideas claras sobre la carrera en que se matriculan, lo que hace que en un cierto número de casos, la tasa de abandono del alumno en ciertas carreras universitarias debido a esta razón. En el caso de los alumnos que terminan sus estudios es también interesante saber la tasa de aprobados y las calificaciones dependiendo de la carrera, el tipo de alumno que se quiera observar, el número de matriculados en la asignatura, el semestre en el que se imparten las

asignaturas y la convocatoria del examen, todo ello puede influir en las tasas de aprobados, ya que puede haber alumnos que saquen más rendimiento al estudio. Como conclusión, el proyecto debería ofrecer información útil sobre el ciclo académico de los alumno para así prevenir la deserción y tomar decisiones para la mejora de los alumnos, además el sistema permitirá demostrar información sobre el ciclo académico para detectar cuellos de botella y porcentajes de éxito entre otras cosas, y poder brindar un mejor grado de satisfacción para los alumnos y la calidad de estudio.

Una vez descritas las metodologías más habituales que se emplean para el desarrollo de DWH, pudimos seleccionar cuál de ellas proporciona mayores ventajas frente a las demandas en el estudio académico de alumnos universitarios. Teniendo en cuenta la información de los apartados anteriores y el interés de este proyecto en centrar el estudio en ciertos aspectos dentro del entorno universitario, la metodología de Ralph Kimball se ajusta más a lo que se quiere desarrollar al permitir la creación del DWH partiendo de los DM. En particular, debido a que el interés de este proyecto se enfoca en los aspectos académicos de los alumnos, se puede considerar que este planteamiento es útil desde el punto de vista de los departamentos universitarios que tratan con alumnos y sus progresos académicos, considerándose este grupo un subconjunto del total de departamentos y organismos de la universidad. De esta forma, el proyecto puede considerarse un primer acercamiento al BI académico a partir del estudio de un proceso de negocio concreto. La metodología de Kimball nos permite, por tanto, ofrecer soluciones en un plazo inferior al que resultaría de abordar un proyecto global destinado a toda la institución universitaria. Por otro lado, también ofrece una clara exposición de las fases y actividades propias de cada fase, así como un buen número de ejemplos documentados en los cuales apoyarse cuando no se dispone de gran experiencia en el desarrollo de DM y DWH. Especialmente importante a este respecto son sus recomendaciones sobre el modelado dimensional

4. Discusión

El proyecto de implantar un data Warehouse dentro de la institución, pero esto no disminuirá el índice de reprobación ya que este solo dará a conocer los posibles focos de los alumnos que pueden estar en riesgo en materias o pensando en abandonar sus estudios, así que dependerá de la institución para poder resolver ese problema, pero al igual podrá dar resultado históricos en cuanto a las calificación de los alumnos, índices de aprobación y reprobación, y brindar datos a las áreas que lo ocupan ahorrando así tiempo y esfuerzo.

5. Bibliografía

- [1] Casares, Claudio. Tutorial Data Warehousing. 1999.
- [2] Corporation, IBM. IBM Inoformix MetaCube ROLAP Option. 2002.
- [3] Corporation, Oracle. Data Warehousing Guide. 2005.
- [4] Dumer, Michelle. Microsoft SQL Server2005 Product Guide. 2005.
- [5] Inmon, Willian H. Building the Data Warehouse. Technical Publishing Group, 1992.
- [6] Kimball, R. The Data Warehouse Lifecycle toolkit. Ed. John Wiley, 1998.
- [7] The Data Warehouse Toolkit. Ed. Jhon Wiley , 1996.
- [8] Royo, José A. Data Warehouse and DataMining. Departamento de Informatica e ingenieria de Sistemas, 2003.