

10 Errores mas Comunes en la Construcción de un Datawarehouse Corporativo

Autor: Jorge Valenzuela Posadas

Mail: jmvalenzuelap@thematika.com

Ingeniero de Sistemas de la Universidad del Callao.

Escuela de Postgrado de la Facultad de Ingeniería Industrial y de Sistemas -

Universidad Nacional de Ingeniería – Lima - Perú.

Resumen

Muchas cosas en la vida se afrontan acertadamente, sin embargo de las cosas que se precia tener mas experiencia, quedan aquellas que surgen de errores cometidos, debido a que se sabe el resultado negativo obtenido, las causas de su contexto que influenciaron al error y el tratar de no-repetirlo en la medida de lo posible.

Este listado de errores en la construcción de un datawarehouse corporativo, es uno de los primeros artículos, que buscan generar una base de conocimiento y experiencia de lecciones aprendidas en proyectos de tecnologías de información, en este caso especifico en proyectos de desarrollo de datawarehouse corporativos.

Este listado fue creado para servir de utilidad a todas las personas que tienen alguna participación en el desarrollo de un datawarehouse y permitirá evaluar cual es el camino al que se dirigen y las oportunidades de éxito final, pero el consejo mas sintetizado que se puede extraer de él, es mantener el trabajo en conjunto en los niveles y áreas organizacionales participantes así como trabajar sobre bases sólidas teórico-practicas de la tecnología involucrada.

El objetivo del presente listado es aportar una fuente de consulta de lecciones aprendidas, de la aplicación de esta tecnología en diversas empresas, como también proporcionar basamento teórico-practico, el cual descansa sobre el dominio de las ciencias de computación y mas propiamente, de la tecnología de datawarehouse, y que debe permitir finalmente, conllevar al éxito futuro, el proyecto de datawarehouse corporativo de su empresa.

Palabras Clave: datawarehouse, datamart, datamining, construcción de un datawarehouse corporativo, errores comunes, gestionando un proyecto de datawarehouse, riesgos de construir un datawarehouse, perspectivas del datawarehouse, métrica de rentabilidad en la inversión de un datawarehouse

Índice

1. Introducción
2. Listado de 10 Errores mas Comunes en la Construcción de un Datawarehouse Corporativo
3. Conclusiones
4. Bibliografía
5. Acerca del autor

1. Introducción

Hoy en día, áreas de tecnología de la mayor parte de empresas financieras y comerciales, que hasta hace poco dedicaban la mayor parte de sus esfuerzos a proveer sistemas de información transaccionales - que soportaban la carga de la mayoría de actividades de sus cadenas de valor - están concentrados en lograr la explotación coherente de su data almacenada: histórica y transaccional, la realidad en torno a ello, es el ingente tamaño de data obtenido de las operaciones diarias de sus sistemas transaccionales, y el problema de analizar y extraer conocimiento de toda ésta información que permanece sepultada dentro de si misma.

Existen una serie de técnicas de computación que permiten la transformación de esta información transaccional, operativa y cotidiana en información con diferentes niveles de agregación, resumida, exacta y especializada por temas y que permite el análisis de las mismas facilitando la toma de decisiones gerenciales, a esta tecnología se le denomina Repositorio de Datos o Datawarehouse.

Sin embargo, y en base a la practica profesional, la construcción de un datawarehouse no esta ajena a problemas que impiden su éxito y el logro de su objetivo final: el apoyo a la toma de decisiones gerenciales.

En este articulo se intenta ir mas allá de lo teórico, presentar asuntos problemáticos que impiden el desarrollo con éxito de un datawarehouse corporativo.

Por ello, se debe especificar que la siguiente lista - de errores mas comunes en el desarrollo de un datawarehouse corporativo - emerge del desarrollo de esta tecnología en representativas empresas pertenecientes al rubro financiero y comercial peruano.

Partiendo que un proyecto debe satisfacer al cliente representando ganancias sobre su inversión - sabemos que no siempre es fácil estimar ello en proyectos de tecnología informática sin embargo – el éxito de un proyecto de desarrollo de datawarehouse es directamente proporcional a la utilidad obtenida del análisis de la información en la toma de decisiones acertadas en las que la empresa sale beneficiada, en este punto: saber que de una decisión adoptada - en base al análisis de datos a través del datawarehouse - se produce mayores ingresos y/o ahorros, es la forma mas adecuada de estimación, para ello se debe considerar, la facilidad de obtención de reportes así como la flexibilidad de su configuración. Se considera - para generar aprendizaje de las decisiones adoptadas cada vez - la importancia de almacenar la forma e información de los agregados y reportes usados a través del tiempo, como una forma de demostrar su valía para la empresa.

Esta lista, con un adecuado uso, permitirá una toma de decisiones mas acertada por parte de la gerencia del proyecto, dueños y clientes, al conocer las posibles consecuencias de una u otra determinada estrategia o medida y prevenir errores frecuentes, permitiendo un ahorro de costos, tanto en la formulación, desarrollo y puesta en marcha del proyecto.

2. Listado de 10 Errores mas Comunes en la Construcción de un Datawarehouse Corporativo

1er Error: Asumir la solución de los problemas que puedan presentarse como una cuestión netamente técnica

El datawarehouse requiere de la participación activa de los usuarios gerenciales tomadores de decisiones.

2do Error : No asignar un presupuesto adecuado para el proyecto completo.

Una adecuada asignación de capital y recursos que permita dar soporte y funcionamiento a la plataforma e infraestructura tecnológica que requiere un datawarehouse corporativo debe de ser uno de los aspectos iniciales al considerarlo como proyecto.

3er Error: Falta de compromiso de la alta dirección.

El éxito del datawarehouse corporativo requiere del apoyo completo de la alta dirección basada en la seguridad y confianza brindada a la gerencia del proyecto y su equipo de desarrollo, para permitir que el trabajo sea fluido en todas las áreas organizacionales involucradas en el proyecto.

4to Error: No tener una infraestructura adecuada que de soporte a ello.

El datawarehouse corporativo requiere una adecuada infraestructura tecnológica como también organizativa. La arquitectura de sistemas de un datawarehouse abarca desde servidores de base de datos propios, servidores de transformación y limpieza de datos, nodos de los usuarios gerenciales front-end dispuestos a lo largo de la organización. En software se requieren aplicaciones en capas servidor, cliente, de negocio que trabajen eficientemente en ambientes de consultas n-dimensionales y en procesamiento paralelo.

5to Error: Bases de Datos redundantes, no transparentes y sin documentar.

El estado de las Bases de Datos transaccionales desde donde se extrae la información que será transformada y almacenada en el datawarehouse, generalmente no es considerada en los estimados de tiempo iniciales, sin embargo, ello puede representar significativa demora, lo ideal en cualquier caso es que los estados de las entidades como cuentas, clientes, deudas, pagos, etc. sean almacenados en tablas históricas por periodos de tiempo significativo según la frecuencia promedio de variación (mensual, quincenal, diario) y en forma estándar,

sin embargo en la práctica esta información se encuentra como parte de tablas usadas para reportes gerenciales, los cuales ya han pasado por un proceso de filtro, perdiendo así la coherencia de la data histórica y cambios en sus estados a través del tiempo a nivel de todo el negocio. En algunos casos se recomienda hacer una reingeniería de bases de datos independiente a nivel transaccional, en la que se pueda almacenar en forma limpia los cambios en los estados de las entidades como un paso previo para considerar la extracción de la data desde sus orígenes.

6to Error: No fomento de un ambiente de cooperación completa entre DBA´s y el equipo de datawarehouse.

Cuando se empieza un proyecto de datawarehouse en una empresa, generalmente ya existe el área de Administradores de Bases de Datos DBAs, por ello se recomienda crear el Área de datawarehouse al mismo nivel de los DBAs - y no bajo el control de los DBAs. Muchas DBAs son responsables del mantenimiento de las bases de datos para dar soporte a las transacciones del día a día. El crear una plataforma alterna suficiente para el desarrollo del datawarehouse, tener acceso a las fuentes de información del dominio del datawarehouse en forma directa y no como intermediarios los DBAs, es uno de los factores importantes que contribuye a un desarrollo mas rápido del proyecto y que apunte hacia el éxito.

7mo Error: No usar una metodología de especificación de requerimientos adecuada para gerencia.

En las actividades de especificación de requerimientos por parte de los usuarios gerenciales, el no usar una metodología que permita que el usuario pueda expresar sus requerimientos en forma fácil y dispuesta para una realimentación posterior es uno de los errores frecuentes. Se recomienda el uso del prototipeo así como la elaboración de documentos de especificación funcional de requerimientos por áreas gerenciales.

8vo Error: Desconocimiento de la cadena de valor, flujos de información en las actividades del negocio.

Identificar de las actividades cuales son factores críticos de éxito, así como observar el flujo de las actividades de competencia central que permiten que los productos o servicios ofertados brinden valor al cliente, es una labor de conocimiento del negocio que los diseñadores del datawarehouse no deben pasar por alto, esto para la elección de una estrategia de desarrollo del proyecto – partiendo de un modelo general desarrollar datamart por procesos o áreas organizacionales – que permita establecer prioridades según el plan estratégico de la empresa.

9no Error: No tener una perspectiva de integración con otras tecnologías relacionadas.

Un sólido basamento teórico en tecnologías de información, y de la amplia gama de tendencias y soluciones existentes, permitirá el desarrollo de un datawarehouse con una visión al futuro. En esta perspectiva, tecnologías relacionadas a el, son: OLAP que permite el análisis de la información histórica para determinar patrones de comportamiento, por otro lado, el datamining que permite el descubrimiento de patrones de comportamiento pero en forma automática mediante modelos y algoritmos (decisión trees, clusteres, network neurals, fuzzy logic, regresión linear, etc.),

Algo mas alejado esta la implementación de un balanced scorecard corporativo, donde el datawarehouse le sirva información de los indicadores por periodos de tiempo históricos.

10mo Error: Mala gerencia del proyecto y discontinuación del plan de desarrollo del proyecto.

La capacidad del gerente del proyecto es expuesta en toda su real dimensión, tanto de la formación técnica, humana y producto de las experiencias en proyectos similares. El desarrollo del datawarehouse debe ser cíclico y por etapas, siendo las actividades de análisis, diseño, desarrollo, pruebas; repetitivas en forma de hitos, tratando de evitar el trabajo repetitivo.

3. Conclusiones

El presente artículo forma parte de una serie de artículos que el autor presenta, sobre tecnologías de información, entregando particularmente en este artículo, una lista de 10 de los errores más comunes, que se presentan en el desarrollo de proyectos de construcción de un datawarehouse corporativo.

El basamento teórico en el que se fundamentan esta clase de proyectos parte de la concepción de las Bases de Datos de Información Multidimensionales (Ontología, Modelo Conceptual de Datos, Interpretación Semántica, Teoría de Conjuntos y Relaciones), de las que se espera obtener información exacta, precisa y resumida para el análisis y apoyo para la toma de decisiones gerenciales.

La base práctica fue obtenida como resultado de la participación del autor en diversos proyectos de datawarehouse en empresas comerciales y financieras importantes peruanas, cuya experiencia lograda es impregnada en el presente artículo, y es presentada al lector como una herramienta de consulta y basamento teórico-práctico, para el éxito futuro en el desarrollo de su datawarehouse corporativo.

4. Bibliografía

1. Building the data warehouse Autor W. H. Inmon Editorial John Wiley & Sons, Inc. New York, NY, USA 1996
2. The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing and Deploying Data Warehouses Autores Ralph Kimball, Laura Reeves, Warren Thornthwaite, Margy Ross, Warren Thornwaite Editorial John Wiley & Sons, Inc. New York, NY, USA 1998
3. Genetic Algorithm for Materialized View Selection in Data Warehouse Environments Editorial Springer Berlin / Heidelberg 1999
4. Data Warehouse Modeling and Quality Issues Autor: Panos Vassiliadis - Knowledge and Data Base Systems Laboratory Computer Science Division - Department of Electrical and Computer Engineering - National Technical University of Athens-Zographou 157 73, Athens, GREECE http://www.dbnet.ece.ntua.gr/~pvassil/publications/PhD/pv_eng_phd.pdf
5. Maintenance of Data Warehouse Views Using Normalisation Autores: Mukesh Mohania, Kamal Karlapalem, Millist Vincent In D.

Ram, editor, Data Management, pages 32--50. Springer Verlag, 1997.

6. A methodology for datawarehouse design: conceptual modeling
Authors José María Cavero Universidad Rey Juan Carlos, Spain, Esperanza Marcos Universidad Rey Juan Carlos, Spain, Mario Piattini Universidad de Castilla-La Mancha, Spain, Adolfo Sánchez Cronos Ibérica, S.A., Spain . Publisher IRM Press Hershey, PA, United States 2002
7. A Strategy for Managing Data Quality in Data Warehouse Systems
<http://www.computing.dcu.ie/~mhelfert/Research/publication/2001/IQ01HelfertMaur.pdf>

Acerca del autor



Nació en Lima en 1974, es Ingeniero de Sistemas graduado en la Universidad Nacional del Callao, actualmente esta elaborando su tesis para obtener el Grado de Maestría en Ciencias con mención en Ingeniería de Sistemas en la Universidad Nacional de Ingeniería en Lima-Perú. Los últimos años viene desempeñado sus actividades en importantes instituciones financieras, comerciales y educativas peruanas. Tiene actualmente fundada la empresa CPS-Tech www.cps-tech.com que se dedica a brindar consultoría de sistemas para negocios globales.

Mail: jmvalenzuelap@thematika.com

Este artículo también puede ser encontrado en www.thematika.com.