

3. Lectura del artículo “A Comparison of Data Warehousing Methodologies”

Debido al alto crecimiento en los almacenes de datos dentro de las tecnologías de la información en los últimos tiempos, un gran número de herramientas y metodologías han sido desarrolladas para manejar proyectos. En este trabajo se hará un pequeño análisis al igual que una comparativa, por otra parte es importante mencionar que no se hablara sobre los temas ya tratados o conocidos durante clase.

Aunque existen diversas metodologías para los almacenes de datos, solo se enfocarán en 15 de estas. Las fuentes de estas metodologías puede ser clasificadas en tres categorías: proveedores de tecnología básica, proveedores de infraestructura y compañías de modelado de información. Se puede presentar un conjunto de atributos que capturen las características esenciales de cualquier metodología de los almacenes de datos, estos son:

- **Core Competency Attribute.** Cuyas metodologías podrían tener diferentes fases dependiendo hasta el segmento que estén. **Los proveedores de tecnología básica** son estas compañías que venden maquinas de bases de datos. Estos proveedores usan esquemas de almacenes de datos que toman ventaja de las matices de sus máquinas de bases de datos. Las metodologías que se revisan incluyen NCR's Teradata-based methodology, Oracle's methodology, IBM's DB2-based methodology, Sybase's methodology, and Microsoft's SQL Server-based methodology.
 La segunda categoría, **proveedores de infraestructura**, incluyen estas compañías que estan en los negocios de infraestructura de los almacenes de datos. Una herramienta de infraestructura en los almacenes de datos podría ser un mecanismo para manejar metadatos usando repositorios, para ayudar a extraer, transferir, y cargar datos en el almacén de datos, o para ayudar a crear soluciones al usuario final. Las metodologías propuestas en esta categoría, por lo tanto, son DBMS-independent. tales metodologías incluyen SAS's methodology, Informatica's methodology, Computer Associates' Platinum methodology, Visible Technologies' methodology, and Hyperion's methodology.
 La tercer categoría, **proveedores de modelado de información**, incluye proveedores ERP (SAP and PeopreSoft). Aunque las metodologías usadas por estas compañías difieren en detalles, ellas están enfocadas en las técnicas de capturar y modelar los requerimientos del usuario en una manera significativa. Por lo tanto, la competencia de esta categoría es modelar información de necesidades de clientes.
- **Requirements Modeling Attribute.** Las metodologías capturan requerimientos de negocio y modelos de información de desarrollo basado en estos requerimientos.
- **Data Modeling Attribute.** Se enfoca en las técnicas de modelado de datos que las metodologías usan para desarrollar modelos logicos y fisicos.
- **Support for Normalization/Denormalization Attribute.** La normalización o desnormalización procesada es una parte importante de una metodología de almacenes de datos. Para soportar consultas OLAP, bases de datos que requieren joing de tablas frecuentes, lo cual puede ser muy costoso. Para mejorar la realización de consultas, una metodología debe soportar desnormalización. Todos los proveedores de DBMS soportan esta actividad.
- **Architecture Design Philosophy Attribute.** Un número de estrategias están disponibles para diseñar una arquitectura de almacén de datos, desde el diseño del almacén para toda la empresa hasta diseño de data mart. La organización necesita determinar cuál objetivo será más importante antes de adoptar una metodología.
- **Implementation Strategy Attribute.** Dependiendo de la metodología, la estrategia de implementación podría variar entre un tipo SDLC y un tipo RAD:
- **Metadata Management Attribute.** Casi todos los proveedores se enfocan en el manejo de metadatos, un muy importante aspecto de los almacenes de datos.

- Query Design Attribute. Gran número de tablas de los almacenes de datos toman un gran tiempo para procesar, especialmente si deben hacer join con otras. Porque las consultas realizan un importante problema, algunos proveedores hacen demasiado énfasis como las consultas son diseñadas y procesadas, varios agregan paralelismo a estas para mejorar la eficiencia.
- Scalability Attribute. Aunque todas las metodologías soporta escalabilidad, esta depende en el tipo de DBMS que es usado.
- Change Management Attribute. Varios cambios afectan los almacenes de datos, podrían implicar reorganización de desarrollo, replantear prioridades, definir objetivos del negocio y otras actividades relacionadas. Además de cambios en el mundo físico. Estos cambios tendrían que ser manejados, recordados y evaluados. Con las herramientas front-end de OLAP, podría haber varios cambios a la interfaz de front-end, tal como adición de nuevos objetos front-end inicialmente no disponibles, cambios en definiciones de objetos, y eliminación de objetos de front-end obsoletos. El manejador de cambios es un problema importante para considerar en una metodología de almacén de datos. Sorprendentemente, muy pocos proveedor incorporan esto en sus metodologías.

Finalmente, las metodologías están evolucionando rápidamente pero muy extensamente porque el campo de los almacenes de datos no es muy maduro., ninguna de estas metodologías ha sido reconocido como estándar aun. Pero en la industria hay una convergencia de las metodologías. Las metodologías basadas en proveedor son apropiadas para las organizaciones que entienden su problemas de negocio claramente y pueden crear modelos de información En otro caso, las organizacion deberían adoptar las metodologías basadas en modelo de información. Si el enfoque es de acuerdo a la infraestructura del almacén tal como metadatos o el diseño del cubo, es aconsejable usar las metodologías basadas en infraestructura.