

Trends in Data Warehousing: A Practitioner's View

William O'Connell

IBM Toronto Lab
boconnel@ca.ibm.com

Abstract

This talk will present emerging data warehousing reference architectures, and focus on trends and directions that are shaping these enterprise installations. Implications will be highlighted, including both of new and old technology. Stack seamless integration is also pivotal to success, which also has significant implications on things such as Metadata.

Trends in Data Warehousing

Industries experience with data warehousing over the last decade has provided important lessons on what works in today's business intelligence (BI) solutions. It is not only these lessons, but also the emerging trends which are also shaping our industry directions in business solutions. As a result, our emerging reference architectures used in building these enterprise data warehouse solutions are changing to meet business demands.

This evolving reference architecture used in building solutions will be overviewed, followed by the implications of these changes. It is these evolving reference architectures that are putting new demands on the databases that are used in warehousing. An important point is that although many of these concepts are not new, databases are being pushed in new ways which are requiring further technology invention.

With the emergence and evolution of the intranet, as well as more businesses exploiting semi-structured data, the more traditional business models are evolving with respect to such things as data accessibility, delivery, and concurrency. Technology such as XML and webservices become more critical as databases integrate

with web portals and BI tooling. Moreover, additional demands on more broad decision making within enterprises are causing heavy consolidation and non-traditional mixed workloads (heavily mixing OLTP and DSS) beyond what has been conventional in the past. Service level agreements, as well as normal operational characteristics are not the same (e.g., backups). Moreover, in many cases the consolidation is not an option and or desired. In such latter cases, the business question still needs to be run. As a result, federation augmentation is also very real in enterprise systems. Query management in a federated environment is still a challenging task. A combination of consolidation and federation augmentation is being seen.

In addition to heavy consolidation and federation augmentation, both real-time (right-time) and active data warehousing systems are being built. These systems present interesting challenges to traditional maintenance and extract/transformation/load operational procedures. Specifically, in large multi-terabyte systems which are 24x7x365. Queries in such systems that execute over aggregated data (including materialized views) need to be very close in time to a consolidated operational data store (ODS) in the same enterprise data warehouse. The maintenance challenges are pushing the technology. Finally, the closed loop processing in an enterprise-wide solution, allows warehouses to play an even more crucial role. Not only are operational systems creating events, so are data warehouses; they play a crucial active role in an enterprise. One such example of events produced in a warehouse is measures, which may be key business indicators (KPIs) used in business performance monitoring through portals.

In addition to this talk presenting emerging data warehousing reference architectures, trends and directions shaping these enterprise data warehousing installations will be overviewed. In doing so, some key implications to databases will be highlighted. In addition to the database itself, any warehouse solution consists of a solution stack. Implications on the whole stack will be touched upon, including such things as metadata and interoperability via standard interfaces such as XML.

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the VLDB copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Very Large Data Base Endowment. To copy otherwise, or to republish, requires a fee and/or special permission from the Endowment

**Proceedings of the 30th VLDB Conference,
Toronto, Canada, 2004**