Abstract

Over the last few decades, business intelligence has emerged as one of the highest priority items on CIO agendas. Businesses and government agencies know that mining information from the increasingly large volumes of data they collect is critical to their business or mission. During this same period, a number of other factors have contributed to the high rate of growth of business intelligence (BI) and data warehousing (DW) technologies including: • Many more users with diverse needs • Need for ad hoc queries vs. standard canned reports • Need for more “real time” information • Growth of the number of databases within an organization, with need for consolidation of information. • Rapidly growing volumes of data • Growth of internet
and web-based applications, including self-service applications. • Regulatory/legal requirements
Traditional approaches to data warehousing have significant drawbacks in terms of effectively
delivering a solution to the business for such diverse requirements. These drawbacks include
high licensing and storage cost, slow query performance against large data volumes, and
difficulty in providing access to all of the data stored in multiple databases. At the same time,
these growing issues are placing a high burden on DBAs and IT organizations to implement,
tune, and manage databases that supply BI information. The result is that business users are
frustrated with how long it takes to get critical analytical information needed for the success of
the business. As the need for BI and DW has grown, various new products and technologies
have been introduced to address different needs. Many are best suited for workloads that
consist of a high volume of planned, repetitive reports and queries. An example of such an
application would be a data warehouse used to support a retail call center. Each time a
customer calls, the system calls up his or her account. This is a repetitive OLTP-like query that
benefits from a specifically designed and engineered system to optimize the performance of
these queries. Data warehouses using a traditional index-based architecture are well suited to
this workload. But another growing area for data warehousing and BI is analytics. Examples
may include, marketing, finance, sales, compliance, risk management, or operations groups
performing ad hoc queries such as: “How did a particular 2007 Christmas sales campaign
perform compared to our 2006 campaign?” or “Let’s analyze why there are more mortgage
defaults in this area over the last 12 months versus the last five years.” The ad hoc nature and
diversity of these requests make row-oriented, index-based architectures a poor choice for an
analytical data warehouse. By definition, DBAs don’t know what users will need in the future
and are therefore unable to determine what indexes to create. As adding an index adds to both
the size of the database and the time needed to load data.

References

- C-Store: A column-oriented DBMS, Stonebraker et al., Proceedings of the 31st VLDB
  Conference, Trondheim, Norway (2005)
- Izak et al., Brighthouse: an analytic data warehouse for ad-hoc queries, Proceedings of
  the 34th VLDB Conference, Auckland, New Zealand (2008).
- Infobright open source data warehousing: Working smarter, not harder. Infobright IEE
  Data Warehouse for Ad-hoc Queries. PVLDB 1(2) (2008) 1337-1345.
- Izak, D., Wróblewski, J., Eastwood, V., and Synak, P. Rough Sets in Data
- Agnew, N.M., and Pyke, S.W. The science game: An introduction to research in the
- Johnson, J.A., and Johnson, G.M. InfoBright for analyzing social sciences data.
- D. ?I?zak, T. Kim, Y. Zhang, J. Ma, & K. Chung (Eds.), Communications in Computer

Index Terms

Computer Science
Engineering and Technology

Keywords
Infobright Optimizer  Data Packs and Data Pack Nodes