

## **Data Warehousing on IDS**

### ***Motivation behind the new warehousing offering from IDS***

By Fred Ho  
Program Director, Informix Competitive Technologies

A data warehousing workload is inherently different from an online transaction processing (OLTP) workload, both in terms of the customer application and in its impact on the underlying database management system (DBMS). OLTP workloads typically consist of short transactions accessing random (usually a few) records. A database schema for an OLTP application is typically designed to minimize redundancy using Entity-Relationship modeling techniques - since changes to the underlying tables generate index updates that are expensive to maintain. The optimally designed OLTP schema is in what's known as 5th Normal Form (5NF): a level of database normalization that reduces redundancy. For data warehousing (or Decision Support Systems), queries to the database usually involve accessing a large number of records, involving joins across dimension tables and the fact table followed by aggregation (Group By) and ordering (Order By). To maximize performance for these types of queries, the schema is typically designed using Dimensional Modeling techniques and data is often de-normalized to 3rd Normal Form (3NF) or less.

IDS inherently possesses a number of features that make it very suitable for a Decision Support System (DSS) environment. For example, IDS' multi-threading, Dynamic Scalable Architecture (DSA) and rich fragmentation (a.k.a. partitioning) schemes, along with the ability to do fragment elimination, allow for efficient parallel query processing. It also has a very efficient hash join algorithm which is essential when joining smaller dimension tables to large fact tables. Finally, its ability to do efficient Add/Drop fragments allows for easy roll-on/roll-off of table and index fragments, critical to most DSS environments where time-cyclic data management is deployed.

Customers with OLTP applications are accustomed to the reliability, high availability, and high performance characteristics of IDS. Fortunately, customers running DSS workloads can expect exactly the same characteristics. In fact, a significant number of IDS customers have already deployed IDS as their data warehouse platform, many of which in the Terabyte range. Even the recent VendorRate survey that declared IDS as the Vendor of Year for 2008 rated IDS highly in the category of data warehousing. So clearly, IDS as a database is a viable platform for the DSS environment—and the new warehousing offering from IDS builds on that platform to deliver a complete solution.

A complete data warehousing solution is much more than just query performance at the database level. There is significant ongoing effort collecting data from various data sources, merging them, cleansing them and finally loading them into the warehouse. There is also the effort of building reports, BI portals, dashboards and scorecards that are expected in today's on-demand and fast changing business environment. Up until now, customers who use IDS for data warehousing have had to rely on custom code, scripts, and application programs to maintain this environment. The new Informix Warehouse Feature offering from IDS is designed to significantly reduce these pain points and to make it not only viable, but easy and efficient to do data warehousing on IDS.

The detailed features of the new offering can be found in an article located at the newly created Informix website for [Informix Warehouse Feature](http://www.ibm.com/informix/warehouse) (<http://www.ibm.com/informix/warehouse>). Briefly, the SQL Warehouse (SQW) component includes a Design Studio based on the popular Eclipse Interactive Development Environment (IDE) that supports physical data modeling. It offers a palette of operators an application designer can use to build the necessary "data flow" operators such as join, merge, and transformation, all without coding a single SQL statement. These data flow operators can then be packaged up into "control flow" operators that are scheduled and run in the database according to a schedule determined by the DBA. Note that by transforming a schema designed for OLTP to a STAR schema more suitable for DSS workload, query performance will vastly improve without any change to the underlying DBMS.

To complete the picture of comprehensive BI and analytic functionality for IDS customers, the IBM Cognos product suite comes with a comprehensive suite of tools for reporting, drill-down/drill-up analyses, and dashboard and scorecard capabilities. Its industry-leading BI capabilities allow it to do both Multi-dimensional Online Analytical Processing (MOLAP) and Relational OLAP (ROLAP) that adapts to the size of the datasets involved. While other BI tools in the industry can also readily access IDS databases, the Cognos product, along with SQW and IDS, provide a one-stop shop for Informix customers seeking a data warehousing solution.

As we continue to enhance this offering, there are plans in place to offer additional warehousing functionality and enhancements to the IDS server including popular XPS features, data compression, further Star Join capabilities and an appliance offering. The active-active clustering capabilities in IDS with MACH 11 also provide options for distributing different types of query workload across different machines. This method makes the most of using real time operational data and can help you maximize your Informix infrastructure even further. These will be discussed in future articles.

Comments and suggestions for feature enhancements are welcomed. Please direct your comments to Fred Ho ([hof@us.ibm.com](mailto:hof@us.ibm.com))

