

## MAXIMIZING DATABASE POTENTIAL

Data mining, the process of analyzing data from a variety perspectives and summarizing it in a useful way, is an important method for enterprises to increase revenue, cut costs, and maintain competitive edge. Using specialized data software, users can analyze data from many different dimensions or angles, categorize it, and summarize useful correlations and patterns that can be used for business intelligence.

## THE PROBLEM DOMAIN

Data storage is a crucial element in the data mining process. Typically, enterprises stored their mission-critical information in legacy databases or flat files, requiring complex queries and powerful computers to access the data for reporting and analysis. As innovation in computer processing power and data storage continue to change the paradigm, there is a call for sophisticated summary and ad-hoc reports from the data store. With the need for summary and analysis applications comes the need for more sophisticated database models. Enterprises are seeking new options for database modeling and data sharing. Data gathered from Online Transaction Processing (OLTP) is suited towards providing highly optimized transactions for CRUD (create, read, update, and delete) operations, but doesn't provide analysis functions. As connectivity becomes more and more important, web-based access to summary reports, ad-hoc reports, general reports and analysis tools is a highly desired option for data analysis.

## THE ODS SOLUTION

A solution comes in the form of the Operational Data Store (ODS). The ODS has characteristics of both OLAP (online analytical processing) and OLTP (online transaction processing). An ODS can respond to complex queries from knowledge users, data mining software, and search engines like OLAP, while processing at a high transaction rate as it is fed transactions in real-time from many intermediary systems. The ODS can be programmed to define a process for data extraction, cleansing, and translation, resulting in a flexible, powerful system for data storage, retrieval, and processing.

The ODS can also sort and program data into facts and multiple dimensions to facilitate creation of a star schema or data cube for complex queries and analysis. Both ad-hoc queries and customized canned queries can be used with a ODS system. The ODS can be used to provide web-based access to canned queries, reports, and summaries using industry standard tools. Easily customized for the specific needs of an enterprise, data synchronization and validation processes can be implemented to massage extracted data to store additional attributes for each business fact or concept. An ODS can be customized with a universal schema to standardize reports and queries.

## BENEFITS

Extraction, cleansing, and translation of data into an ODS provides uniform data for analysis. The ODS programming of data into facts and dimensions transforms raw data into useful data, facilitating the creation of

star schema or data cube models, an intuitive and effective method for data reporting and analysis. The simplest of the data mart schemas, star schemas minimize creation of multiple table joins and relationship traversal. Additional benefits of ODS include easy implementation of complex queries with drilldown capabilities, and application of periodic dimensions like monthly, semi-annual, and annual projections can be easily accomplished.

## DRAWBACKS

Traditional query optimization techniques provided by a relational database management system (RDBMS) may not work for star schema models. Due to complexity of the ODS data retrieval, cleansing, and sorting process, the process time for loading facts and dimensions may take longer. Although the facts and dimensions in a star schema provide an intuitive view of the data, they also introduce complexities through multidimensional joins. Users writing their own queries need to understand performance overheads of multi-dimension joins. Finally, online reports that perform complex, time-consuming queries may drain IT resources.

## A HYBRID APPROACH FOR MAXIMIZING PERFORMANCE

- When query performance, data extraction, and load time are critical, keeping the ODS schema as close as possible to or same as the OLTP schema will increase performance and minimize load time.
- Identify reports that do not need the complex querying capabilities provided by star schemas and have them go against the OLTP schema.
- Program the ODS with canned queries and summary views of the data to support more custom and sophisticated analysis.
- Make the extraction, cleansing and translation process parallel and queued for performance.
- Long-running reports like financials can be run in Batch mode to utilize offline time.
- The ODS and Star schema approach should only be used if the need for complex querying and analysis is of a high priority.

## ABOUT PROKARMA

ProKarma delivers integrated technology and business process outsourcing solutions for over 150 global leaders in a wide range of industries and markets. ProKarma is co-headquartered in Portland, Oregon and Omaha, Nebraska, with sales and delivery centers in the United States, India, Argentina and Peru. ProKarma was selected as a Global Services 100 Provider for 2012 and ranked as the fastest growing IT services company in America by Inc. 500.