Reducing Elapsed Times: One of the secrets to Data Warehouse Success

*By **Carlos Alves**, IT Director, Consult Inc.*

With data volumes developing rapidly for IT departments everywhere, data warehousing dba's continually face the challenge of making high-quality data accessible, broadly and quickly. This means such performance issues as the *excessive elapsed* times encountered in the running of data-intensive applications can become real problems. Once again, the old cliché is proved true: time is money.

Fortunately, a number of enterprise tools are available that can help data warehouse dba's overcome these problems. This article discusses what features to look for in a total data management solution: one that will not only help cut down elapsed times, but also optimize business processes like database loads, batch processing, ETL, Web log processing and data mart implementation.

Whether your daily processing volumes fall in the megabyte or terabyte range, excessive time can be spent in the data staging phase of operations. Using the right tool, you can reduce the number of passes through the data by speeding up or combining several key steps, including summarizing, copying, extracting, filtering, joining, merging, ordering, pattern-matching, partitioning, segmenting, transforming, and validating. These steps, along with sorting, will speed loading and accelerate extract processing, thus expediting CRM, BI, and DSS applications.

If you have large relational databases, too much time can be spent loading and then extracting data for handoffs or reports. With the appropriate tool, bulk loading for databases like Oracle can be accelerated by using some of the techniques mentioned above. In this manner, 50 percent or more can be cut from the average load time, providing you with the necessary information for analytical purposes sooner.

A good data management solution must also offer the ability to handle a variety of record formats and data types. This gives you the flexibility to process data from disparate sources and to output data in the form required for downstream operations.

Implementing a dedicated aggregation engine will allow you to process large numbers of records in minutes and not hours, giving you those data mining results that you need, as quickly as you need them.

In my data warehouse development days, I have determined that successful projects share the following characteristics:

* **They are business-driven**: Business analysis focuses the project on high-payback data (e.g., 20/80 rule where it concentrates on the 20 percent of the data sources that supply 80

percent of the business needs) and concrete business results/reports that are traceable to the data through designs, models, requirements, and needs.
Periodic return-on-investment evaluations are conducted based on pre-build calculations (traditional cost-to-savings ratios) and post-build calculations (business-based or application-based cost justifications).

* **They are incrementally built**: Early, small successes are emphasized to gain user support and acceptance.

* **They are based on proper modelling**: The data warehouse logical model is normalized to the maximum practical degree to support data extracts of subsets for data marts used for decision support and analysis.

* **They are data quality-centred**: The data model and the "extract, transform, and load" application are employed to boost overall data quality by improving many characteristics of the data such as accuracy, consistency, and minimized redundancy.
Addressing data quality from the beginning is very important because ensuring this can account for as much as 60 to 80 percent of the total cost of the data warehousing project.

We operate in an information-driven economy. This means that a company’s most valuable strategic asset is its data. By ensuring that a complete data management solution is a part of your data warehouse enterprise, you can secure the continuing business value of this very data.