![Astera - DATA INTEGRATION MADE EASY]

Whitepaper

Maximizing Data Warehouse Capabilities with Software Tools

What is a Data Warehouse?

Advances in technology have allowed organizations to collect and store massive amounts of data from many sources. The challenge today is to find a reliable, consolidated, integrated reporting and analysis system that enables them to leverage their data for different applications, from customer service, to partner integration, to top-level executive business decisions.

The data warehouse is a repository of an organization's electronically stored data. Data stored in the data warehouse is different from data found in the operational environment in that it is organized in such a way that relevant data is clustered together to facilitate reporting for day-to-day operations and analysis to determine business trends and to make decisions based on this information.

While the data warehouse itself is focused on data storage, software tools for extracting, transforming, and loading, retrieving and analyzing, and managing the data are also key components of a data warehousing system. These tools are the underlying framework that ultimately ensures the quality and usability of the data. A broader definition of the data warehouse therefore must include business intelligence tools, tools to extract, transform, and load data into the repository, and tools to manage and retrieve data.

Data Warehouse Advantages

- Improved access to a wide variety of enterprise data
- Increased data consistency
- Additional documentation of the data
- Lower computing costs
- Increased productivity
- Centralized place to combine data from disparate sources
- Creation of infrastructure that supports changes in computer systems and business structures
- Empowers users to perform ad-hoc queries or reports without impacting performance of operational systems and resources

The data infrastructure of most organizations is a collection of heterogeneous systems. For example, an organization might have one system that handles customer-relationship, one for human resources, one for sales, production, finance, partners, etc. These systems are often poorly or not at all integrated with each other and simple questions can be very hard to answer, even though the information is available "somewhere" within the disparate data systems. Data warehouses solve these issues by creating a single database of homogeneous data that can be more easily accessed and manipulated. The number of sources can be nearly limitless, provided the system can handle the volume.

While larger companies typically use data warehousing to analyze big sets of information for enterprise purposes, smaller companies and enterprises wishing to review just one subject often utilize data marts. Data marts are more specific and targeted in their storage and reporting. Data warehousing can include smaller amounts of data grouped into marts that are connected together as part of the larger system. In this way, companies can use both data warehousing and marts, letting users choose the best source and functionality depending on their current needs.

Centerprise Data Integrator for Data Warehousing and Data Marts

- Sophisticated ETL Features in a Unified, Intuitive Environment
- Performance: Key to Data Warehousing Success
- Slowly Changing Dimension Write Strategy
- Loading Fact Tables
- Lookup Transformation
- Change Data Capture
- Aggregate Transformation
- Connectivity

Sophisticated ETL Features in a Unified, Intuitive Environment

Centerprise Data Integrator brings together high-performance data warehousing ETL features in a unified and intuitive package. Designed to be used by business experts and developers alike, Centerprise is widely acclaimed for its superior usability, high performance, and scalability. Centerprise offers a number of features and optimizations to support data warehouse loading, including a high-performance slowly changing dimensions (SCD) component, lookup caching, robust parallel-processing engine, and optimized database writes.

Performance: Key to Data Warehousing Success

Centerprise is designed from the ground up as a multithreaded, parallel-processing engine that ensures minimal blocking and starvation of threads, thus delivering a high degree of parallelism. When combined with today's multicore and multi-processor hardware, this approach results in a data transformation engine that can scale to handle high data volumes. The Centerprise engine increases throughput in direct proportion to increases in processing power, which, for data warehouses with exploding data volumes, ensures continued scalability of the warehouse and the ETL processes.

Slowly Changing Dimension Write Strategy

Centerprise features a dedicated write strategy for loading and updating slowly changing dimension tables. The SCD Transformation uses Centerprise's data synchronization engine to efficiently handle large dimension tables. The SCD Transformation supports Type 1 and Type 2 SCDs and provides multiple row-versioning patterns, including effective and expiration dates, active/inactive value, and version number fields.

Loading Fact Tables

Loading fact tables efficiently is vital to a successful data warehouse project. With ever-increasing data volumes and shrinking transfer windows, it is imperative to load data quickly and correctly. Centerprise supports native bulk inserts for popular databases. Additionally, a number of other technologies and features combine to deliver throughput to handle large data sets.

Lookup Transformation

With a large number of records and dimension table lookups, the need for a high-performance lookup transformation cannot be overemphasized. Centerprise specializes in unique lookup transformations that provide the technology necessary to handle high data volumes, including fast, intelligent caching and parallel execution. Dimension table versioning and late arriving facts are supported via effective/expiration dates, active/inactive row, and version number fields.

Change Data Capture

A technology that improves throughput considerably is change data capture (CDC), a set of approaches where incremental changes are applied to destination tables. Centerprise supports two distinct CDC patterns: incremental read from source using audit fields and incremental update at destination using the CDC hash function. These approaches, when used appropriately, can substantially reduce transfer runtimes.

Aggregate Transformation

The Centerprise aggregate transformation applies aggregate functions such as sum, count, minimum, maximum, average, and other aggregate functions to elements. Additionally, users can specify group-by elements and create output grouped by the specified fields. As with all Centerprise components, the aggregate transformation uses Centerprise's parallel processing technology to enable users to process high data volumes.

Connectivity

Centerprise's ever-expanding library of Centerprise Connectors offers a plethora of integration options that support popular databases and file formats including Oracle, SQL Server, DB2, Sybase, and MYSQL. Centerprise also supports the most popular file formats including delimited, fixed length, Excel, COBOL, XML, and others.

Centerprise Solutions

Data Warehousing

High-performance data warehousing ETL features in a unified, intuitive environment.

CDC

Choose either a batch or real-time change data capture strategy for your particular requirements.

Data Mapping

Transforms advanced mapping, validating, and cleansing tasks into basic drag-and-drop or single-click commands.

ETL

Extract data from any source, transform it to suit your needs, and load it into your database or warehouse.

Data Conversion

Complex doesn't need to be complicated. Visual, code-free parsing, transforming, and loading of data from any source.

Data Integration

A single platform for complex, hierarchical integration the requires no coding.

Data Migration

Unique hierarchical data processing technologies automate and streamline data migration projects.

EDI

Full electronic data interchange functionality combined with Centerprise complex data mapping capabilities.

Astera
DATA INTEGRATION MADE EASY

www.astera.com
Contact us for more information or to request a free trial
sales @astera.com   8888-77-ASTERA