

Moving Large Data at a Blinding Speed for Critical Business Intelligence

A competitive advantage

Intelligent Data In Real Time

How do you detect and stop a “Money Laundering” transaction just about to take place?

How do you detect and stop a fraudulent credit Card transaction taking place?

How do you transfer all the stock transaction data to be reconciled before the cut off time of 4:00 PM?

How do you get the integrated health data of a patient before making any major decision?

How do you get hundreds and even thousands of your vendors accessing information in real time to adjust the inventory, production, and shipment?

How do you broadcast several marketing campaigns everyday to move certain products and services?

One of the critical part of the answer to all these questions is:

By moving large amount (100s GB to TBs) of data at very high speed throughout the day without affecting the normal Online Transactions

The origins of database go back to libraries, governments, businesses and medical records. There is a very long history of information storage, indexing, and retrieval. Properly indexed and labeled data have always been necessary for businesses to get ahead.

Companies started using computers in 1960s with increasing storage capability of computers. In 1970-72 relational database model originated, which defined the logical organization of a database from the physical storage methods. In 1980s SQL (Standard Query Language) became prevalent and fueled DB market for business and has become the standard since then.

Looking at the current state of industry, it's clear that strategic applications are rapidly scaling up the amount of data they manage and the number of users they support. This trend is growing at a rapid rate. Handling such growth successfully and managing such a wealth of information intelligently requires new and innovative ways of thinking. OLAP (Online analytical processing) and BI (Business intelligence) are taking new dimensions

to knead the data, which provide the enterprise the means to identify strategies to realize the complete business value.

Intelligence in Data

BI can help organizations analyze business strategy and behavior. The strategy analysis scope can vary from external analysis such as customers, competitors, environment and government regulations, to internal analysis such as operational performance analysis, business activity monitoring (BAM), financial performance and investment returns.

For example:

The operational managers and staff level employees at the reservation desk of a reputed airline company uses the business information and the intelligence as a tactical tool to provide relevant information in a most effective way to the customers.

A leading financial institute does data analysis and prediction to resolve problems, involving both fraud detection and risk management. They use the database of customers who are late in paying their mortgage. The organization developed the profile of customers likely to default their mortgage payments.

Health care organizations use data mining to overcome wide range of business issues such as predicting length of stay in a hospital, forecasting treatment costs, and predicting total cost of patient care.

Database paradigms

The business intelligence and knowledge discovery on OLTP data is very important to keep pace in the accelerating business world, however is it wise to do such analysis on the OLTP systems, or is there need to move the data to data warehouse or data marts for analysis? The following paragraphs discusses the OLTP and data warehouse architecture.

OLTP systems are the information gathering backbone of an enterprise. These systems administer and handle a business's daily activities. The data in the OLTP systems is organized for the business systems (application oriented) to speed up the transaction processing and is optimized for data integrity. Typically the data is in normalized form to efficiently allow the small data read/write access by large number of users. The data in the OLTP is not historical in nature and is not stored in the structure that supports ad-hoc reporting and analysis.

Data warehouse provides framework to use and analyze data to make effective business decisions. The data in these systems are optimized for fast retrieval and is usually in De-Normalized form (Star Schema or Dimensions and Facts are common database paradigm

in these systems). The data is optimized for small number of large transactions, which involve ad-hoc and complex queries. Typically data in these systems offers a single, consistent version of the data, unlike OLTP systems, which offer distinctly formatted data. The users of these systems usually summarize and consolidate the data along multiple levels of abstraction and from different angles to analyze and achieve the knowledge, which is at very high level, compared to primitive level data.

Data Architecture in an Enterprise

The previous section explains why it is important to have a separate data warehouse or data mart for analysis and reporting purposes. In this section we will see, how the business transactional data is collected within the centralized warehouse. The process of making data available for BI tools goes through various stages.

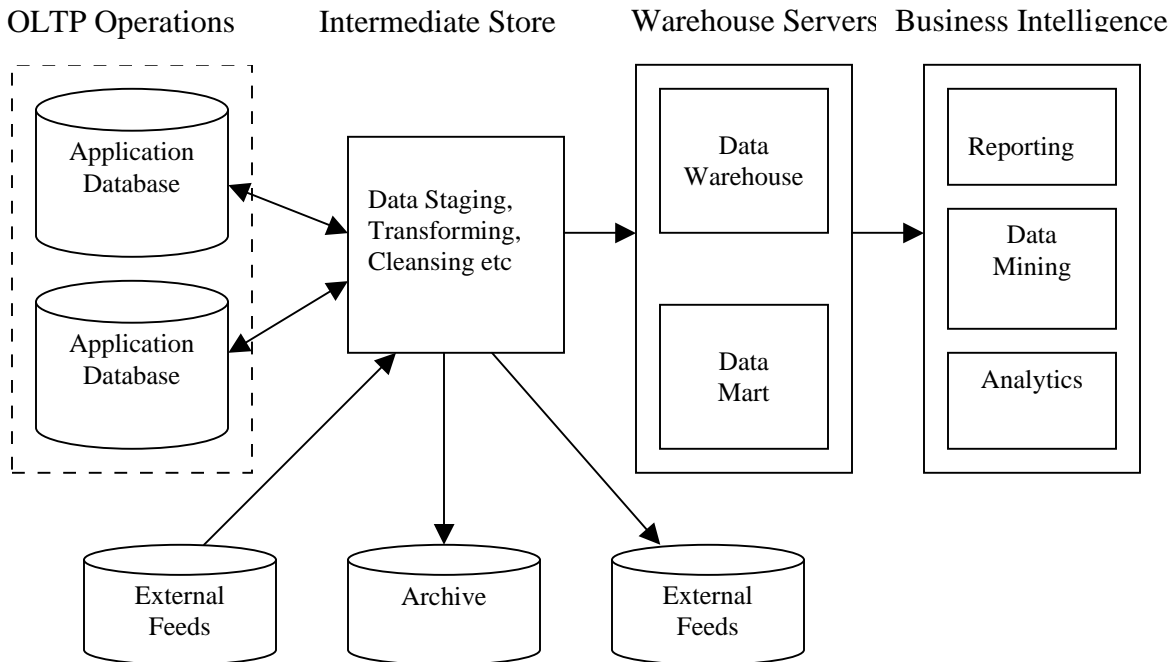


Figure 1 *Data Architecture in an Enterprise*

The OLTP data and external 3rd party data are extracted into the data staging area. The data is then transformed and cleansed into a format suitable for warehouse servers. It is then loaded into the data warehouse. Business intelligent tools use the data loaded into the data warehouse for analysis, reporting or mining. Typically the ETL (Extraction Transformation and Loading) process is done on periodic basic depending upon the business needs (every few hours to once a month). The data in data staging area is also loaded into the OLTP systems to create new data sets or it is archived or given to external feeds for other applications.

In the staging area, the related data from different OLTP systems are merged. Different encoding mechanisms in different OLTP systems are resolved. The data is also cleansed before it is moved to data warehouse to remove the inconsistencies such as different addresses for the same employee or customer.

OLTP systems have all transaction details. OLAP queries usually need summary or aggregate data for reports. For example, a query to retrieve quarterly sales totals for the products over the last two years run faster if the data has aggregated rows showing daily or hourly data, rather than the query with data having each transaction records. The amount of aggregation depends upon different factors such as speed requirements, granularity required for the analysis.

The OLTP data is organized for business systems, and is usually difficult and time consuming for analytical processing. When the data is moved from OLTP to OLAP systems, it must be transformed so that it supports the analysis.

Data Migration

The huge, periodic data migration happens between various systems in an enterprise for different purpose. The most critical of them would be the data moving out or into the OLTP system, as it affects the business transactions directly. It is very important to move the data from OLTP systems to other platforms or vice-versa fast, intelligently and efficiently while minimizing the impact on processing systems. It becomes even more crucial when the OLTP data is moved more frequently.

Following examples illustrate few scenarios of huge data exchange for knowledge discovery.

A consumer goods manufacturing company has 48 plants worldwide. They replicate their production data into the headquarter database twice a day, where it's used for cost-management planning and reporting. They approximately move half terabyte data each time from all units to the headquarter database.

An online brokerage firm profiles data of 8.5 million customers to help cross-sell brokerage, banking, and loan products and services. They run approximately 65 producer campaigns a week. They move approximately 200 GB of production data to the data mart for the campaigns, 65 times a week.

A retail store has a huge operational data store, which serves as a current, real-time quality business data of various production systems in an enterprise. Different groups use this operational data for the analysis specific to their group. An online marketing group, uses the guest data to predict the online business trends and the customer behavior to score the customer and offer promotions or discounts. This group extracts incremental data (approximately 30 GB) to the data marts every 3 hours. Another group does the

ad-hoc-analyses on guest data to analyze the products sale in the stores by moving approximately 200 GB data twice a week. The vendor-reporting group collects approximately 40 GB data every day from the operational store to do vendor specific reporting and analyses.

One of the largest health-care organization serving more than 3 million subscribers integrates the data to create an enterprise data warehouse and bring together and align the information from four affiliated, but separate plans. They extract, cleanse, align and integrate data from 24 discrete source systems. Many of these contained data dating back over 20 years. Collectively these applications contain more than 8 terabytes of data.

A fortune 1000, building manufacturing company deploys a data integration solution to ensure employees, customers and vendors to have access to the business and product information. They populate two different systems regularly with most up-to-date information. They integrate more than 27 million records daily from 20 applications.

An Airline Company has created an advanced customer database solution sourced from the airline's reservations. They combine a carrier's passenger and transaction data with proprietary demographic and other external data to provide carriers with a new source of marketing data to grow customer loyalty and improve marketing decisions.

How Do Genus products help?

NonStop Enterprise Division (NED) of Hewett-Packard has played major role in the On Line Transaction Processing (OLTP) business and is still the preferred vendor for such application. Approximately 95% of trade transactions, 80% of ATM transactions, 75% of the EFT network, 67% of credit card transactions and 70 million wireless subscribers use NonStop SQL for doing business. A huge amount of data (in terabytes) has been created as a result of such successful OLTP business. It is very important for organizations to analyze the business transactional data for planning, market trends, budgeting, profitability analysis, performance measurement, and finding patterns.

Genus has suite of products (Fig. 2) to help move huge amount of data at high speed from (to) NonStop SQL to (from) other platforms or efficiently aggregate the huge data set.

Genus High Speed Data Connectivity Infrastructure (HSDCI) leverages scalability, availability features provided by NonStop Servers to move data into and out of NonStop Server fast, efficiently and in parallel without disrupting online applications. Genus Data Extraction Tool and Genus Data Loading Tool, which is part of High Speed Data Connectivity Infrastructure, can seamlessly integrate with leading ETL tools on remote platforms such as Unix or Windows. The periodic transfer of huge data from OLTP applications to data staging server or vice-versa in real-time is cumbersome process, Genus tools helps to make this process lenient, efficient and cost-effective.

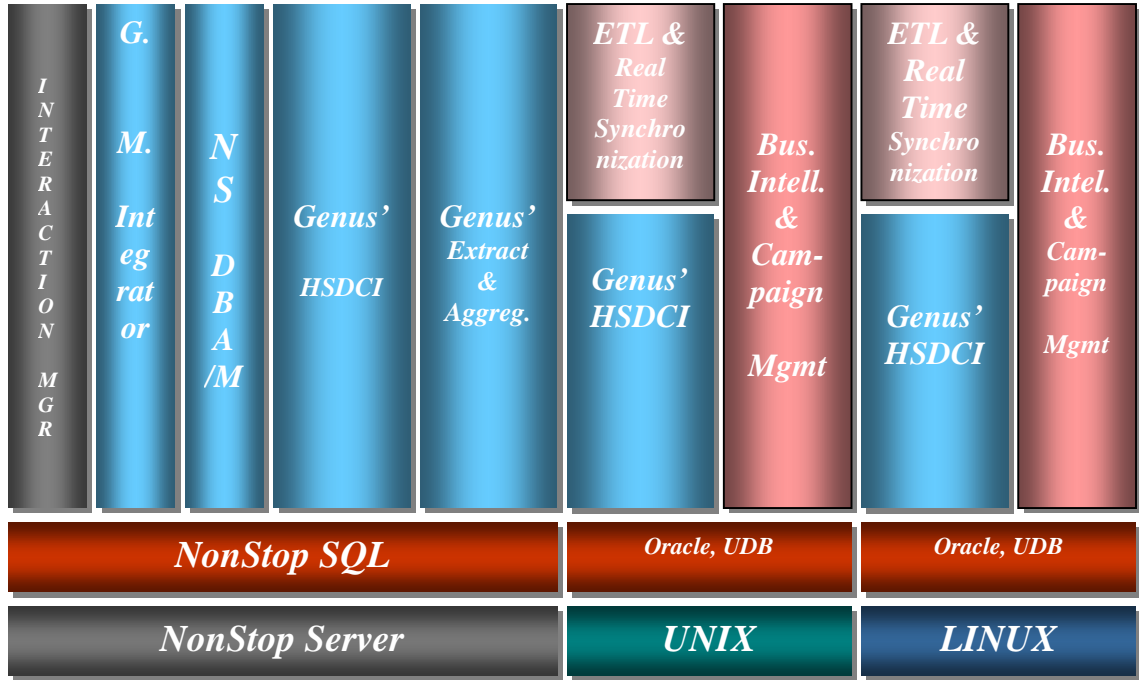


Fig 2. Genus' Tools in an Enterprise Environment

In certain situations it is very wise and efficient to aggregate the data on application platform, instead of moving huge data off platform and do aggregation there. 'Genus Data Extractor and Aggregator Engine' creates customized restructured data subset of data derived from NonStop SQL database in an optimized and scalable manner.

Conclusion

Analyzing the data for effective business needs is very important. However the process is complex and time-consuming. One of the challenges in this process is moving the huge business data fast, efficiently, in timely manner to analytics platform. Without the use of proper tools and techniques, it becomes inextricable. Use of tools which are scalable, reliable and manageable, will maximize the efficiency of the enterprise.

Contact: Genus Software, Inc. Phone 408 257 8722, E-mail – products@genussoft.com, Website www.genussoft.com
Lloyd Niccum, Phone 651 983 3400, lloyd@genussoft.com, Mark White, Phone 314 249 6275, mark@genussoft.com