Help alleviate batch windows with reliable, timely data delivery

Access to timely and accurate information has become a critical driver of business success. Organizations need access to timely and accurate information to make better business decisions, serve their customers better and respond to changes in the market. But increasingly, they are hitting roadblocks in their quest for up-to-date, accurate business information.

Today, most midsized-to-large corporations have a data warehouse capable of analyzing information from production systems on a daily or weekly basis. However, many are finding that extracting information from overtaxed mission-critical production databases is becoming increasingly time-consuming and complex. At the same time, the demand for more timely information is increasing to the point where daily or weekly updates are no longer sufficient.

The demand for real-time integration

What keeps today's chief information officers (CIOs) up at night? It is likely to be a long list of challenges, and chances are that data integration ranks high amongst those.

Let's look at some business challenges that organizations face:

Increasing demand for real-time information for reporting and analytics. Traditionally, reporting was done from warehouses that were updated on a daily or weekly basis. For many types of reports, that data is current enough. For others, though, nothing short of up-to-the-minute will suffice, such as inventory data where product inventory is very high or very low, or billing information where billing is done by the minute or every fraction of a day.

Large volumes of information are difficult to handle in a batch window. As more information is gathered—such as online transaction data, inventory data and customer information—the effort involved in moving it to the warehouse increases drastically. Many organizations are finding that an 8-hour batch window is no longer sufficient for traditional extract, transform, load (ETL) tools to integrate all of the needed data.

Necessity to conduct business 24x7 is reducing batch windows. As more business is done across time zones and over the web, many organizations face the problem of shrinking batch windows, which makes it more difficult for traditional ETL tools to extract data in the short time available.

Growing need to detect and react to business events as they happen. Many organizations are looking for ways to detect business events in production systems and have those events trigger a response in another system. For example, a cell phone company would like to send a text message to a customer running low on minutes asking if he would like to purchase more time.

The need to track all changes for auditing purposes. Organizations must comply with regulations; this often requires them to continuously track all changes to data and not just the net result of those changes.

Increasing need to keep data in sync across the enterprise. Customers want up-to-the-minute access to order, payment and inventory data so they can buy products, pay bills and check delivery status online. Employees need much of the same so they can better service customers and make wise business decisions. To accomplish this, eCommerce data needs to be in sync with business applications and data needs to flow in real time across the enterprise.

Organizations want to deploy new applications using data on legacy systems without paying for an increase in workload. Often, legacy systems are already maxed out and new capacity is very expensive. Organizations want new applications on UNIX or Microsoft Windows to avoid this cost, but integrating the data from those legacy systems without increasing the load on them is a key challenge.

Why optimize ETL processes?

The majority of organizations use ETL tools to extract data in bulk from their production systems and load it into other systems, such as data warehouses. The main strengths of traditional ETL tools are that they extract data from many different applications, perform complex transformations on that data, and then bulk load large volumes of data into data warehouses. But when it comes to extracting data from production systems, there are certain specific requirements of an ETL tool, such as:

ETL is a batch process that typically requires a batch window—ETL tools can move batches of data out of production systems into data warehouses. For many corporations, running a nightly batch process when systems are not running at full capacity works well. However, as businesses are increasingly becoming global and as more and more business is conducted 24x7 over the Internet, finding a time when production systems can be taken offline to run a batch extraction of data becomes increasingly difficult.

ETL batch windows increase with the amount of data extracted—Even where corporations can tolerate an overnight batch window, many are finding that with increasing amounts of data to be extracted, the batch window required to do the extraction is expanding beyond the hours that their production systems are offline.

ETL tools typically reflect net changes rather than incremental changes—For example, traditional ETL tools would not track if the inventory level of a certain product decreases due to purchases but then increases due to restocking, or if the status of an order changes.

Data replication enhances ETL

A data replication solution provides a way to capture changes on production systems so that they can be applied elsewhere without directly querying the database. Coupled with an ETL solution, this provides an effective way to alleviate batch window restrictions. Many corporations combine the strengths of data replication and ETL tools. For example, real-time data replication technology accesses the data directly from database logs and replicates data from operational systems without requiring batch windows or interruptions to mission-critical systems. The data is then loaded into operational data stores (ODS), with the ETL tool feeding the data warehouses or other applications. Production systems can be kept up and running 24x7 and real-time reporting can be supported off of the ODS.

An ideal data replication solution for ETL has many benefits:

Reliable data delivery. Only data changes are replicated, with transactional integrity—resulting in much greater scalability through less data being moved across systems, and reliable, consistent data for making trusted business decisions.

Minimal impact on the performance of production systems. Native, log-based data replication reads database log files to detect changes rather than querying the database directly. Hence there is minimal processing impact on mission-critical production systems.

Flexible implementation. Depending on IT requirements, there could be many different approaches to delivering trickle feeds to ETL processes, including staging tables, message queues, flat files or a direct connection through the network.

Changes are captured in real time so information is always up-to-date. Data replication tools capture changes continuously as they occur. The result: information is always up-to-date rather than being only as current as the last batch window, leading to more reliable information for intelligent business decisions.

Does not require changes to the source system. Because the data replication tool is reading only the log, it does not require changes directly to the source database, yet it can detect all transactions including descriptive information about the change (user, application, time and so on).

Logs all changes to the system, not just the net results of those changes. For audit and compliance, all insert, update and delete actions are recorded rather than just the net results of those actions.

Conclusion

By reliably trickle-feeding real-time data changes to ETL processes, transactionally consistent data can be delivered throughout the day, with minimal source system utilization, instead of waiting for a batch window. This helps accelerate data integration processes and provides the enterprise with readily available trusted information for improving business outcomes.

For more information

To learn more about IBM solutions, please contact your local IBM representative or Business Partner, or visit: **ibm.com**/software/data/integration

IBM Software Group
Route 100
Somers, NY 10589
U.S.A.

Produced in the United States of America
May 2011

IMW14004-USEN-03