Manage Data Growth While Increasing Revenue with Fast and Efficient Data Replication

Traditional data warehouses currently in use at many businesses are having difficulty keeping up with the pace of user reporting and analysis needs. Businesses are seeking a better way to connect data in different locations, as well as a faster, more efficient method of extracting their data.

The traditional Extract, Transform, and Load (ETL) solutions can take a long time for certain types of data (e.g., web server logs). The speed at which you can extract information from Big Data, for example, depends on how quickly and efficiently rapidly changing information can be loaded into data warehouses. Many organizations are moving to a data lake approach, where data is stored as natively as possible for greater efficiency. Then when needed, the data can be extracted into different views.

Whether your goal is to expand your organization’s real-time data capabilities, data warehouse, ability to develop and deploy custom applications quickly and efficiently, or build a more scalable, responsive database infrastructure, AWS provides an extensive set of cloud services that can help you drive your business forward. AWS Marketplace has the tools, templates, and resources to help you get started. There are solutions that will help you automate bulk data loading from database sources, such as Oracle, Microsoft SQL Server, and MySQL to Amazon Redshift, Amazon S3, and Amazon RDS, continuously, with zero downtime. You will also be able to avoid the undifferentiated heavy lifting of manually extracting data, transferring via API or script, chopping, staging, and importing.
Data Replication Energizing Your Business Needs

Data replication platforms can be valuable for businesses who want to improve or need the following:

- **Supplemental transactional systems**: Supplemental transactional systems help improve business performance by detecting and reacting to business occurrences in real time.
- **Mobile apps or big data analytics projects**: Projects of this kind require up-to-the-minute data to maximize business results.
- **Data synchronization**: This is necessary to ensure critical operational systems leverage the latest information. Anyone with master data management projects, data warehouses, analytical data marts, and mainframe environments will want to leverage data synchronization.
- **Continuous availability**: Necessary to protect high availability for the most widely used transaction data stores, for disaster recovery, and for cross-site workload balancing. Fail-over recovery environments are no longer the norm, and IT executives need continuous availability for critical workloads.

Common Business Use Cases for Ongoing Data Replication

- Transferring data incrementally data to AWS
- Migrating data from a data warehouse or an SAP environment
- Moving data to a Big Data system based on Hadoop or Spark for batch and real-time data analytics
- Backup, disaster recovery, and archiving

Database Replication Methods

- **Basic Read-Only Replication**: With basic replication, data copies provide “read-only” access to the table data that originates from a primary (master) site. Applications can query data from local data copies to avoid network access regardless of network availability. However, applications throughout the system must access data at the primary site when updates are necessary.
- **Snapshot Replication**: Data on one server is copied to another server, or to another database on the same server. A read-only table snapshot is a local copy of table data that originates from one or more remote master tables. An application can query the data in a read-only table snapshot, but cannot insert, update, or delete rows in the snapshot.
- **Transactional Replication**: Transactional replication is the automated periodic distribution of changes between databases. Data is copied in (or near) real time from the primary server to the receiving database. Transactional replication also offers a backup for frequent, daily databases changes. In most cases, transactional replication begins by taking a publisher snapshot, which is then copied to the subscriber. Then, any publisher changes are logged in real time and replicated.
- **Data Synchronization**: This is the process of comparing the source data and destination data to ensure that they match. The key benefit is that such solutions are generally free or inexpensive. The downside is that the process of synchronizing them is quite system-intensive, and consequently this process generally runs infrequently.
AWS Database Migration

The AWS Database Migration Service works by setting up and then managing a replication instance on AWS. This instance extracts data from the source database and loads it into the destination database. It can be used for a one-time migration followed by ongoing replication, to support a migration that has minimal downtime. During this process, the AWS Database Migration Service handles many of the complex details associated with migration, including data type transformation and conversion from one database platform to another, for example, from Oracle to Amazon Aurora. Also, by transitioning from Oracle to Amazon Aurora, you are no longer bound to a proprietary database technology. The service also monitors the replication and the health of the instance, notifies you if something goes wrong, and fails over to a replacement instance if necessary.

Conclusion

To increase the scalability of existing infrastructures, a cloud-based data warehouse can be an appropriate solution. The scalability of cloud-based services means organizations don’t need to spend cycles updating databases. Additionally, they get all the benefits of utility-based pricing. AWS Marketplace has data replication solutions to address your company’s requirements, including automating data transfers.