Success Story

This Fortune 50 retailer needed to migrate large volumes of data from legacy database platforms to a distributed cloud solution.

Solution Background

This Atlanta-based Fortune 50 company is the largest home improvement retailer in the United States. The retailer needed a data integration solution for migrating large volumes of data from legacy database platforms to CockroachDB a distributed cloud database solution with real-time replication and bi-directional synchronization enabled.

Business Challenge

Our client had numerous database tables shared across multiple applications. Data synchronization between source and target database platforms was implemented during the early adoption to meet specific business objections. This limited the ability to perform initial data loads that were required when new applications were onboarded, because any impacted tables needed to remain online. The typical solution would be to file imports into staging and common table expressions into the destination. However, this solution does not scale due to the size of certain tables and the complexity of identifying batches for parallel execution. A thousand tables would require a thousand unique implementations, each one with its own combination of keys to delineate segments of data that can be migrated in parallel. The time it would take to implement a solution for each migration was not feasible. Additionally, the runtime to execute a single table migration with 600MM records using CTEs was tested at around 5-10MM records per hour, which would take several days to complete. The size of the segments were not small enough and the number of parallel processes available were not large enough to make this a workable solution.

ADAPTURE THINK FORWARD.

Solutions Requirement

The retailer needed a solution that was easily repeatable across several migrations for multiple applications and tables. Implementation time needed to be minimal to meet a deadline before multiple migrations planned each month. It was required that execution time per migration was no more than 10 hours to fit within a typical production deployment window for an application.

ADAPTURE Delivered

ADAPTURE scripted and deployed an opensource data integration platform called Apache NiFi across three small node clusters using commodity hardware running on Google Cloud. Our team of experts used NiFi to develop a simple data pipeline that picked up any data files sourced from Oracle and mapped the file to a table in CockroachDB. The pipeline handled splitting large files, batching records for asynchronous processing, data transformations, back pressure, error handling, and transactions. The process guaranteed no data loss and enabled us to monitor progress in real time. The solution was seamless to deploy, easy to spin-up and down as needed, and required very little effort to include new files and tables for each migration. The pipeline was tuned for throughput and handled a hundred gigabytes of data at a time, loaded files for several tables at the same time, and allowed us to process 100,000 inserts per second for a rate around 350MM per hour. The solution scaled horizontally, so throughput was achieved simply by adding additional nodes in the cluster.

Benefits

Our client now has a data migration solution with the ability to transfer large data volumes. It is easy to deploy in their environment, can be quickly adopted, and scales to meet their needs. This tactical solution enabled the home improvement retailer to move forward with their application migration to the cloud. In addition, they started loading data and moving their applications into production.

READY TO MIGRATE APPLICATIONS TO THE CLOUD?

Contact an ADAPTURE Solutions Architect today.