DataWalk Performance Test Results
Deep Queries

Executive Summary

A common issue with any type of database (e.g., SQL, NoSQL) is that some complex queries never return a result. Considering this, DataWalk executed benchmark tests to measure performance results with deep queries (i.e., many SQL joins), to measure where a traditional SQL database may fail, and to measure how DataWalk performs with a large number of joins.

Note that while DataWalk is a full-stack system for data analysis and not a database, DataWalk includes a powerful facility for easily generating queries by traversing data sets and connections through a visual interface. These queries can then easily be saved and modified, and then instantly re-run by the push of a DataWalk "Easy Button", or they can be initiated from other systems via the DataWalk API.

The benchmark results indicated that for a sample customer data set, the SQL database was unable to return a result if nine joins were required. In contrast, DataWalk delivered linear performance as the number of joins increased, up to the maximum test of 600 joins.

This reflects that with DataWalk's unique technology, not only will complex queries be quickly computed, but any query will always return a result, which is a significant advance relative to alternative systems.

Depth traversing comparison

Legend:
- DataWalk jumps
- Standard joins

X-axis: Depth
Y-axis: Time in minutes
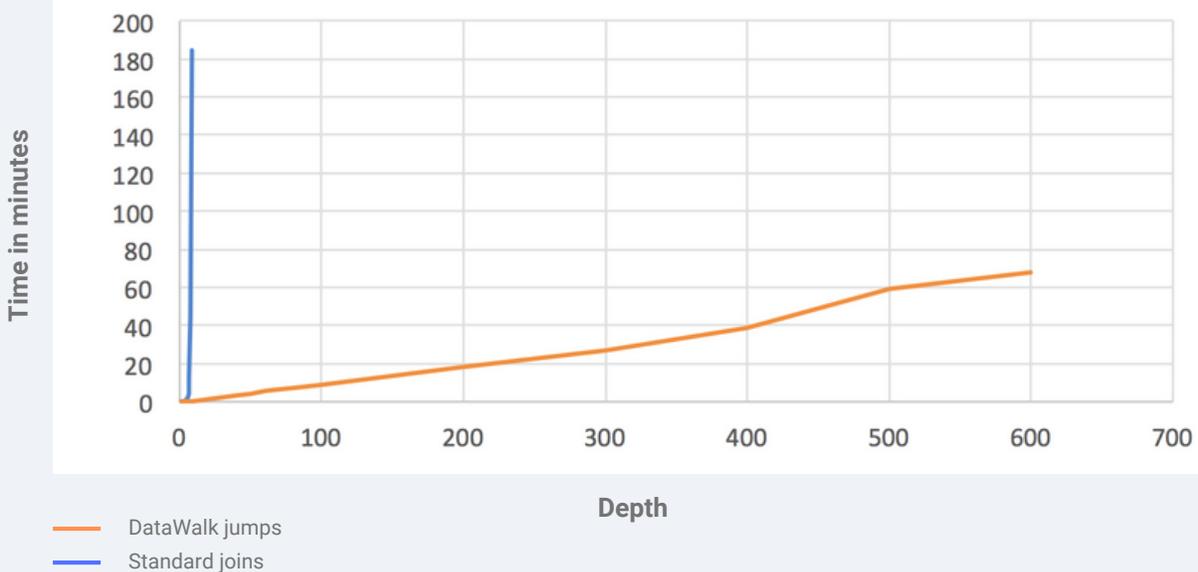
The Test

This test was to query the systems while incrementing the depth of the query based on the "inner join" and "left join" statements. The specifications for the queries were generated based on sample reports as used in customer environments.

Configuration of the system under test was identical for both tests, with eight commodity servers and a 1.3TB database.

Test Results

The SQL-based system showed exponentially increasing response time and resource consumption as the number of database joins was increased. For this workload, the SQL system was unable to do more than eight joins. Attempting to join nine structures resulted in the system declaring insufficient resources after 3 hours of processing, and attempting to do any further SQL querying did not generate any results.

On the DataWalk system, queries are generated via the Universe Viewer, and DataWalk technology transparently divides complex queries into multiple smaller pieces. For the same test referenced above, the DataWalk software was able to join up to 600 structures with linear performance as the number of joins increased. This represents a dramatic improvement of >>60X in the ability to do deep, complex queries.
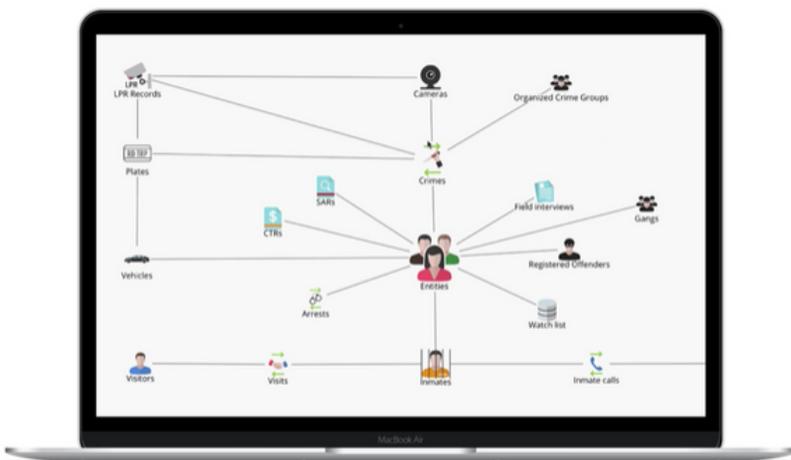
DataWalk is a full-stack data analysis platform for revealing patterns, relationships, and anomalies for large-scale, multi-source data. DataWalk allows organizations to rapidly import and blend vast amounts of data from multiple sources into a singular data view, and then analyze data and connections via simple visual interfaces.

For additional information, see www.datawalk.com.

www.datawalk.com   info@datawalk.com