Yellowbrick

BENCHMARKING

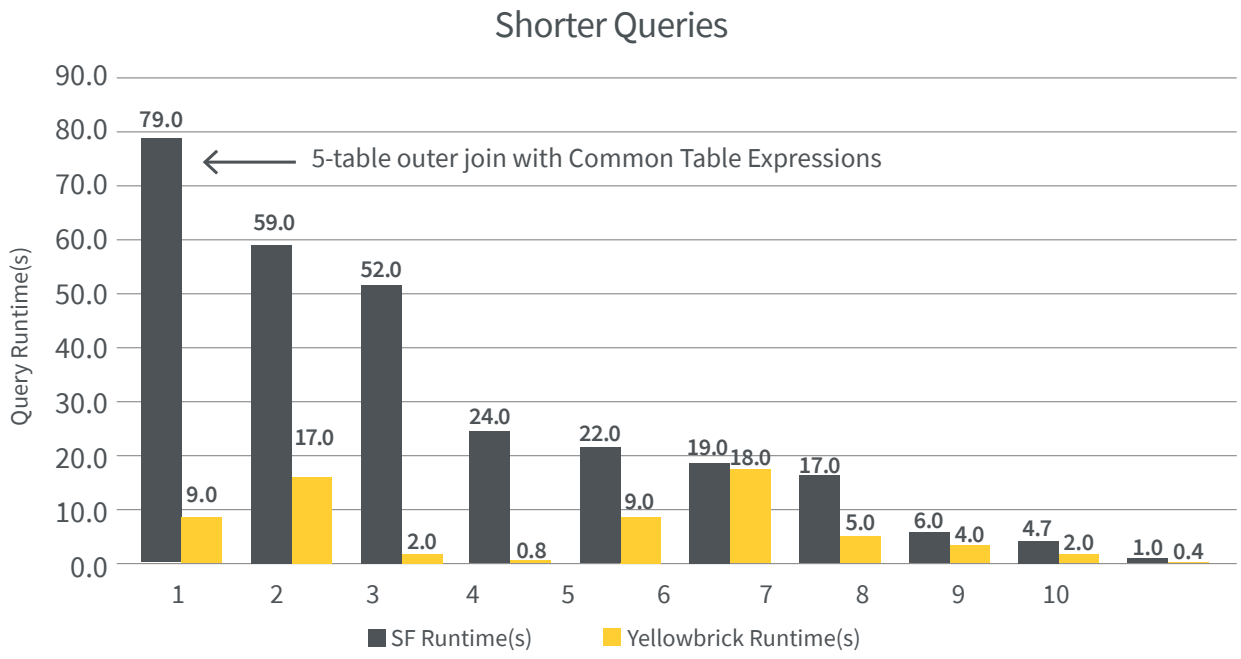
Yellowbrick vs Snowflake Query Performance Comparison

Yellowbrick is regularly benchmarked against other data warehousing solutions during the procurement process. In this document we present query performance comparisons from real world benchmarks against Snowflake. This benchmark was conducted by a customer using its own data and query workloads.

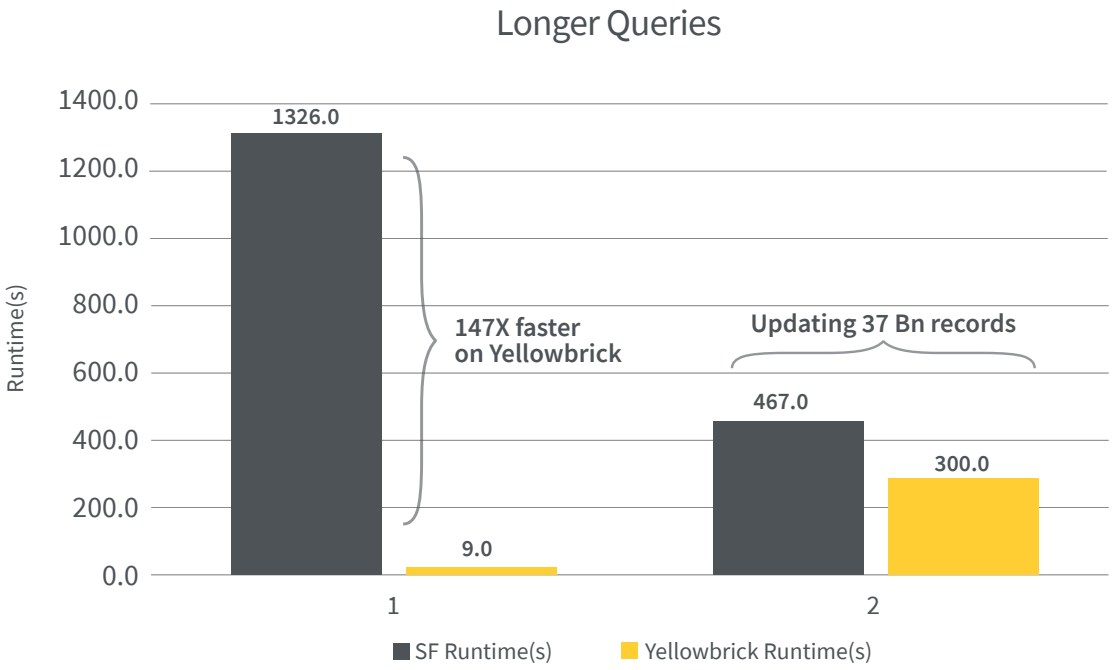
Snowflake is an MPP cloud data warehouse, which is delivered "as-a-service" on all three major public clouds. Advocates of Snowflake cite its elasticity features, its ability to share data easily and its compelling user experience. While it is certainly true that Snowflake has made data warehousing more accessible to many smaller enterprises, its performance struggles when data volumes and concurrency requirements are higher and mixed workloads get more complex. Snowflake's answer to higher concurrency and query volumes is to spin up more virtual data warehouses, which means partitioning up your users and workloads manually. This quickly becomes cost prohibitive, particularly if you are running business-critical "always-on" applications.

A customer in the digital marketing space wanted to reduce the time to insight into campaign performance for its clients, who are primarily in the financial services sector. The customer decided to replace its existing 40-node Vertica data warehouse platform to drive the improvements in performance needed from its advertising platform. The customer ran proof of concept exercises with Snowflake and Yellowbrick to identify the replacement. As part of the POC, it conducted query performance tests using 12 representative SQL workloads against its own data. The data consisted of 3.5 trillion records distributed across 82 tables, with a volume of approximately 500 TB.

The query tests compared a 32-node 2XL Snowflake cluster with a 30-node Yellowbrick cluster. Yellowbrick was 6X faster on average across all 12 queries. Ten of the queries had runtimes less than 2 minutes on both platforms:



The queries ranged in complexity from simple aggregations to complex joins and large updates. Yellowbrick was 147X faster in the case of one of the longest running queries involving a semi-join scanning over a year of data:



In addition to outperforming Snowflake in terms of query performance, Yellowbrick was also able to demonstrate a significant improvement in performance of critical ELT tasks, taking ELT times down from over 10 hours to just 45 minutes. Near real-time data ingestion was also important to the customer and tests demonstrated that Yellowbrick was able to ingest data from Kafta topics at 2 million records per second, far exceeding the 900,000 records/second requirement.

After selecting Yellowbrick, this customer can now update campaigns every 2 hours, enabling it to modify underperforming campaigns in flight, rather than having to let an obviously underperforming campaign run all day. This has helped the customer increase the performance of the upsell campaigns for their clients.

Test drive Yellowbrick today!

www.yellowbrick.com/test-drive