

# SQream vs. Snowflake 300TB performance on the TPCx-BB with AWS

## Introduction

TPC Express Big Bench (or TPCx-BB) is a benchmark that was developed in order to objectively compare Big Data Analytics System (BDAS) solutions. SQream's big data analysts ran an internal field test derived from the TPCx-BB in September 2021 to understand its performance on large datasets in comparing to one of its cloud competitors – Snowflake. For more information regarding TPCxBB, please see [https://tpc.org/TPC\\_Documents\\_Current\\_Versions/pdf/TPCx-BB\\_V1.5.2.pdf](https://tpc.org/TPC_Documents_Current_Versions/pdf/TPCx-BB_V1.5.2.pdf).

## Platforms Analyzed

SQream (currently running only on private cloud), Snowflake.

## Scale Factor

We ran the benchmark with a scale factor of 300,000, which creates a dataset of ~300TB, as SQream was designed to handle large datasets.

## Hardware Used

The main consideration for customizing the hardware stack was the right balance between cost and performance. Obviously, we took into account Snowflake's recommendation depending on the size of the chosen dataset (300TB) and maintained an equal number of nodes for SQream.

	Environment	Configuration	Compute cost (hour)	Storage cost (TB)
Snowflake	AWS	X-Large	\$32.00	\$40 (on-demand)
SQream	AWS	16X g4dn.8xlarge	\$34.8	\$23

## The Process

After configuring the chosen cloud environment for the field test and generating the 300TB dataset, we were ready to begin. Out of the 30 queries included on the TPCx-BB, we tested only 17 use cases as a reflection of the functionalities that were supported by SQream's platform as of September 2021. Those queries were 5-7, 9, 11-15, 17, 20-26. As we were running the different use cases, we focused on two metrics for comparison :

### Performance:

- Ingestion – time elapsed during the process of transporting the data from its source to the DB / DWH.
- Query – time elapsed during the process of executing the 17 queries (using concurrent streams, aka 'Throughput Test').
- Total Time To Insight (TTTI) – Ingestion + Query.

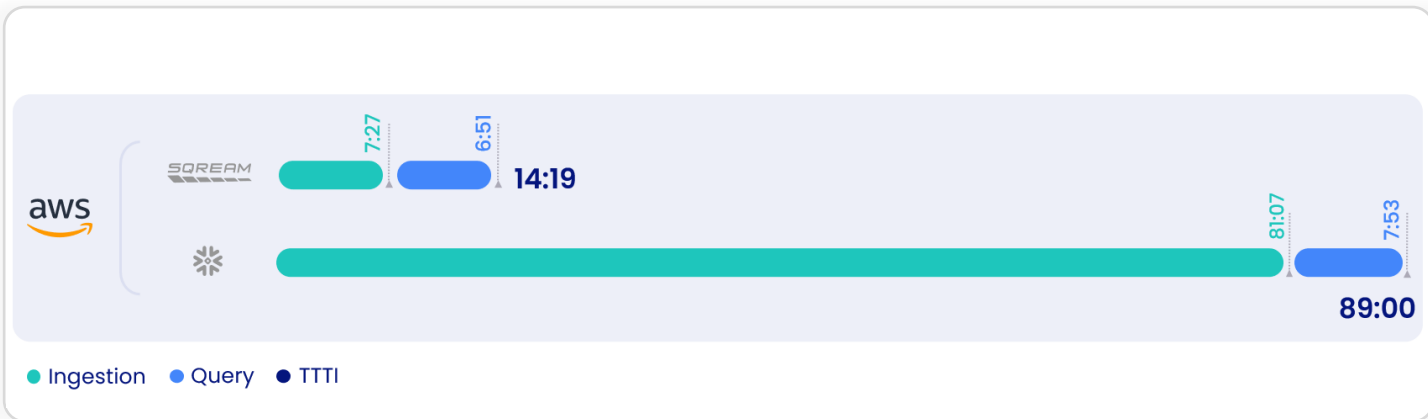
### Cost :

- Storage – the cost of storing the compressed data on the relevant cloud vendor service (\$/TB).
- Compute – the cost of resources used to ingest the raw data from its sources and complete the 17 queries (\$/Hour).

## The Results

The following chart shows the overall performance of each platform for the given workload, in terms of total time for Ingestion and Query in the TPCx-BB field test:

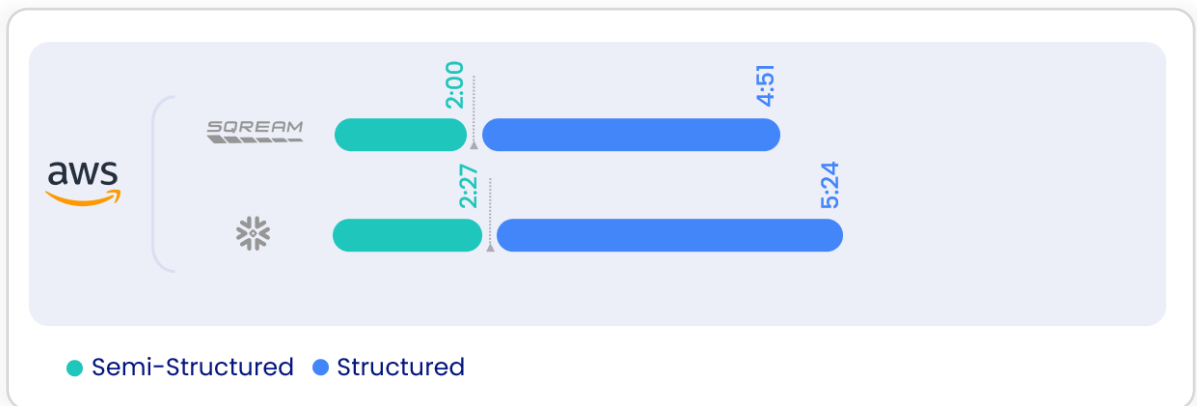
## TPCx-BB 300TB Benchmark – Performance HH:MM (lower is better)



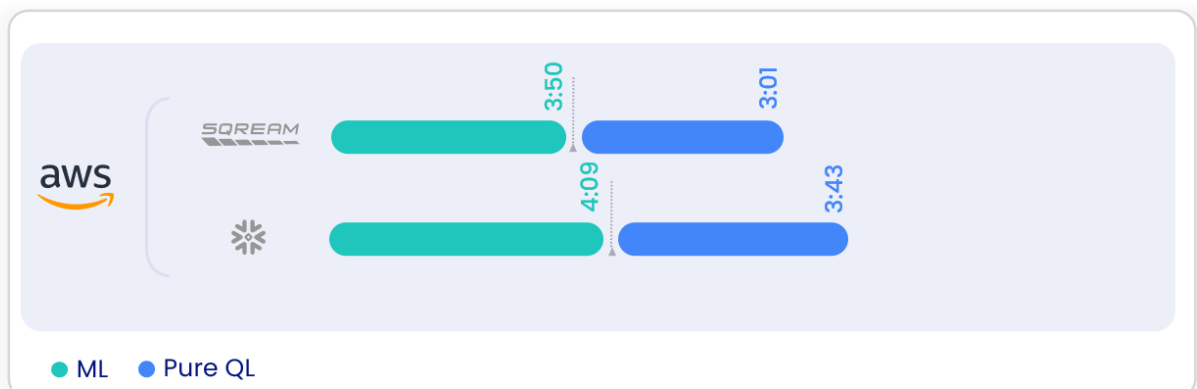
<sup>1</sup>Due to not strictly running TPCx-BB, we didn't use its metrics for performance and cost/performance.

The results revealed several performance differentiators between the competing products. Overall, SQream presented a much better TTTI, X6.2 faster. As for average execution time of the 17 queries, both platforms presented almost the same results (with a slight advantage for SQream). When segmenting the results into more specific use cases or data types, SQream maintained its advantage:

### Query time performance (MM:SS) – per data type (lower is better)

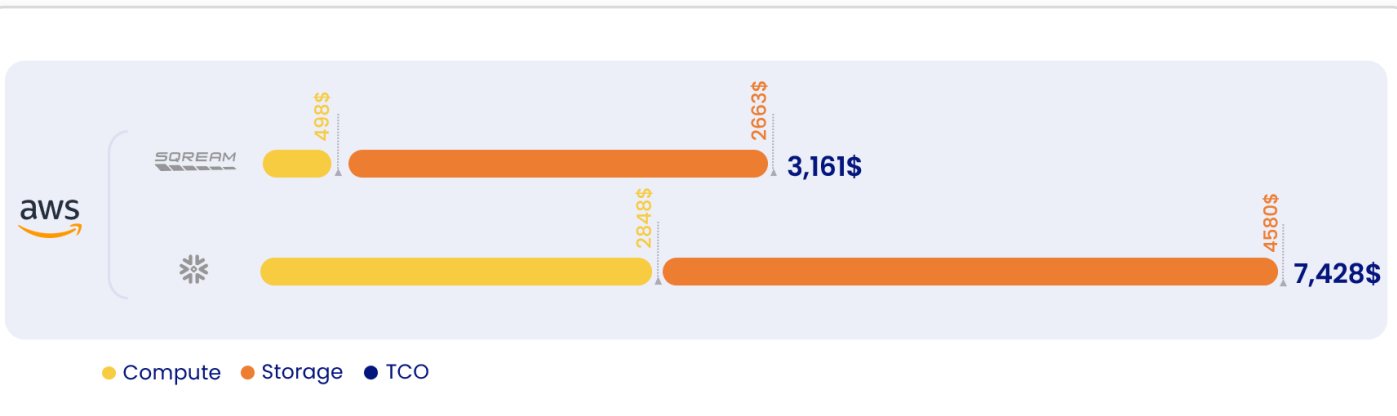


### Query time performance (MM:SS) – per use case (lower is better)



Even though the compute cost of machines with GPUs (which is SQream's case) is usually much higher, the outstanding performance of SQream during the field test (and especially in the ingestion part) staging it also as the most cost-effective option:

### TPCx-BB 300TB Benchmark – Cost (lower is better)



<sup>1</sup>On-demand pricing depends on the selected vendor / cloud environment.