Revolutionizing Data Analytics: SQream Blue Shatters Performance Benchmarks at a Fraction of the Cost.

SQream Blue sets a new standard in big data analytics with its cloud-based solution, dramatically outperforming Databricks in the TPCx-BB benchmark. Processing 30 TB of data in under an hour and at $\frac{1}{3}$ of the cost. SQream Blue delivers unparalleled ROI by leveraging GPU parallelism technology for large-scale enterprises.

3x faster at a 1/3 of the cost

SQream Blue tackled TPCx-BB's most complex queries and proved to be a cost-performance leader for big data analytics.

Testing involved processing and analyzing 30 TB of data, which SQream Blue completed the

benchmark 3x faster at a $\frac{1}{3}$ of the cost compared to the leading data lakehouse alternative. That's like reading every cataloged book in the US Library of Congress in under an hour - and then buying them all for less than $25.
And that outstanding cost-effectiveness scales up rapidly with the cloud-based SQream Blue, for an even higher long-term ROI.

What is TPCx-BB?

The nonprofit Transaction Processing Performance Council (TPC) developed TPC Express Big Bench (or TPCx-BB) as a benchmark for objectively comparing Big Data Analytics System (BDAS) solutions. It covers essential functional and business aspects of big data use cases and provides verifiable performance, price/performance, and availability metrics.

Outstanding cost-effectiveness scales up rapidly with the cloud-based SQream Blue, for an even higher long-term ROI.

SQream and the TPCx-BB

TPCx-BB's benchmark queries represent a range of data processing tasks typically encountered in real-world retail scenarios. One such query, for example, requires the solution to build a model for predicting if a visitor to an online store would be interested in a given item category based on their online activities and demographics.

We ran the benchmark on AWS cloud with a scale factor of 30,000, which creates a dataset of around 30 TB, to test the relative capabilities of SQream at scale. All generated data was stored as Apache Parquet files on AWS S3 and the queries were processed without pre-loading into a dat abas e.

For comparison, SQream Blue was tested against a leading data management solution, Databricks, running its most advanced query engine - Photon. SQream Blue used a Pro cluster of g6.8xlarge machines, while the Databricks configuration was an All-Purpose Compute (Photon) cluster of r7gd.4xlarge machines. Each configuration cost about the same, which would allow us to balance performance with affordability and compare cost-effectiveness fairly.

The Results: High Speed and Low Price

The results for the TPCx-BB 30 TB benchmarking on AWS clearly indicated that SQream Blue provides a significant cost-performance advantage for big data analysis.
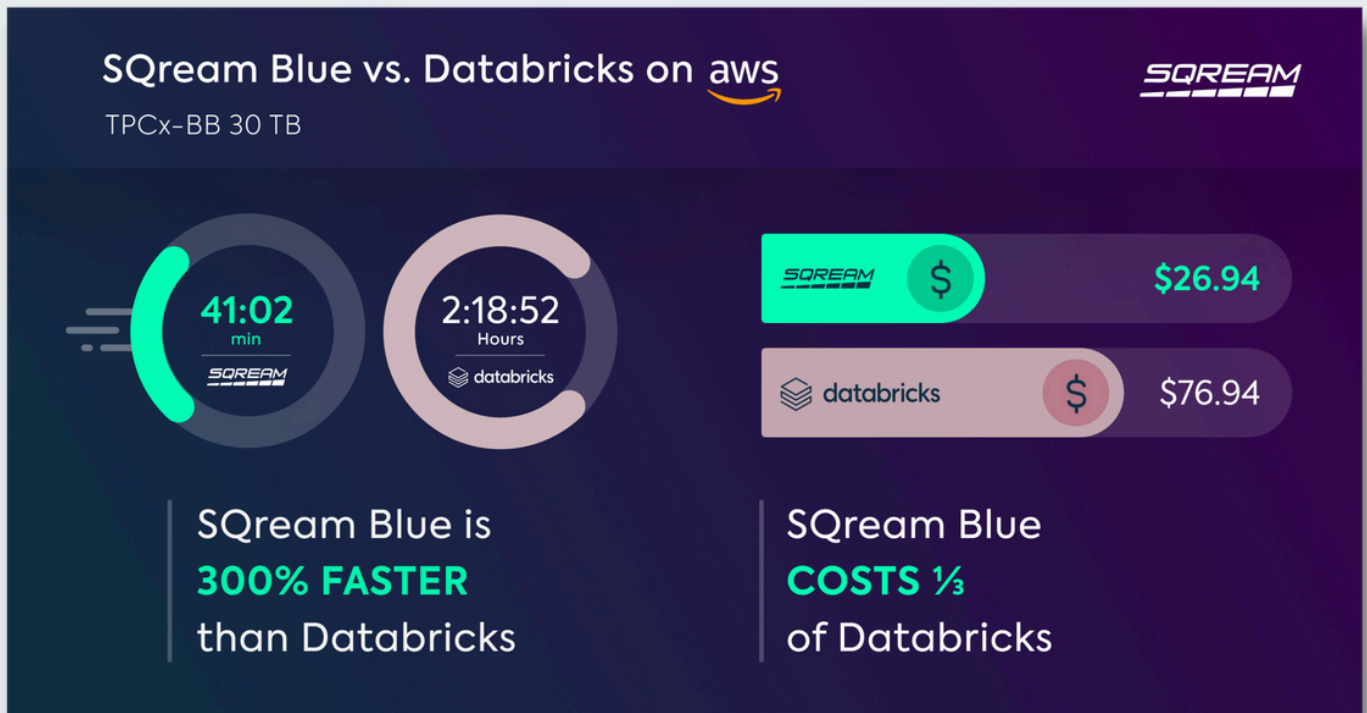
SQream Blue's total runtime was 2462.6 seconds (which is 0:41:02). The total cost for processing the data end-to-end was $26.94.

Databricks' total runtime was 8332.4 seconds (which is 2:18:52), at a cost of $76.94.

SQream Blue introduced a 70% reduction in time at the cost of 1/3 of Databricks. This level of efficiency indicates that achieving SQream-level performance with other solutions would rapidly become extremely cost prohibitive.

SQream Blue vs. Databricks on aws
TPCx-BB 30 TB

41:02 min — SQREAM
2:18:52 Hours — databricks

SQREAM $ $26.94
databricks $ $76.94

SQream Blue is **300% FASTER** than Databricks

SQream Blue **COSTS ⅓** of Databricks

How SQream Blue Does It

SQream Blue is a cloud-based SQL data lakehouse that empowers organizations to accelerate data processing and queries with a unique multi-level GPU parallelizing technology and patented compression algorithm. Moreover, the solution can allocate available resources to ensure dynamic workloads are handled most efficiently. The interface layer's statement compiler analyzes each query and proactively divides data processing between the GPU and the CPU, with some queries running solely on the CPU to avoid unnecessary overhead.

By distributing operations across multiple GPU cores and other resources, balancing parallelism and concurrency as needed, SQream Blue provides unprecedentedly rapid results and optimal performance.

SQream Blue provides unprecedentedly rapid results and optimal performance.

The solution's architecture also contributes to its streamlined efficiency, as data does not need to be ingested or moved at any point during the preparation cycle. Rather, SQream Blue directly accesses data in open-standard formats at the customer's low-cost cloud storage, maintaining privacy and ownership, preserving a single source of truth, and eliminating the need for data duplication. Processing is further optimized by leveraging Apache Parquet's column-oriented structure and metadata to eliminate unnecessary data reads that often increase run times and computing costs..

To ensure the best possible support for organizations engaged in big data analytics, SQream Blue easily integrates with common open-source workflow management and orchestration tools (Apache Airflow, Dgaster, Prefect). SQream also integrates seamlessly with industry-standard ODBC, JDBC, and Python connectors.
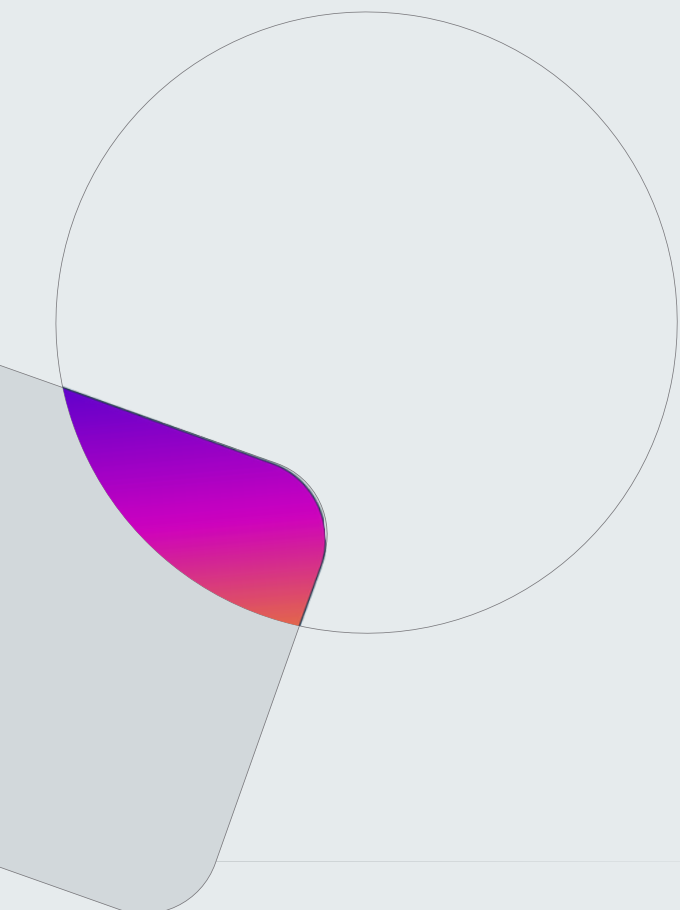
What about the cost?

While a single GPU server used by SQream Blue is more expensive than a CPU server, the GPU has far better cost-performance at scale. Hundreds of CPU servers would be necessary to handle the massive terabytes-to-petabytes of information enterprises typically need to process, while the same server power can be derived from just 10-20 GPUs with SQream's technology.

Thus, SQream Blue's speed, flexibility and scalability make it the most cost-effective tool for quality insights from large, complex datasets. This translates into high ROI, low total cost of ownership (TCO), and price-performance that leads the pack – just as the latest TPCx-BB results unequivocally demonstrated.

Contact us for the full TPCx-BB results.

About SQream

SQream empowers companies to get value from their data that was unattainable before at an exceptional cost-performance. Our data processing and analytics acceleration platform utilizes a patented, GPU-based SQL engine that accelerates the querying of extremely large and complicated datasets. By leveraging SQream's advanced supercomputing capabilities for analytics and machine learning, enterprises can stay ahead of their competitors while reducing costs and improving productivity.