SQREAM

Don't settle for a sample. Ask Bigger.

Faster and better ML data preparation with SQream

Challenge:

With the ever-growing complexity of Machine Learning (ML) models, the need for ML Operations (MLOps) has become essential in deploying successful models into production and continuously improving them with new data.

On average, **data scientists spend 80% of their time preparing data**. Attempts to deal with this challenge, such as training smaller datasets or conducting fewer cycles of experimentation after deployment, only **reduced ML models' accuracy** and **forced companies to compromise**.

Solution:

Leveraging the ability to **perform numerous tasks across several GPU cores in parallel**, SQream's petabyte-scale analytics engine was designed to deal with complex data preprocessing and wrangling. With [PySQream](#), data scientists can seamlessly integrate SQream's processing engine within their existing Python MLOps environment.
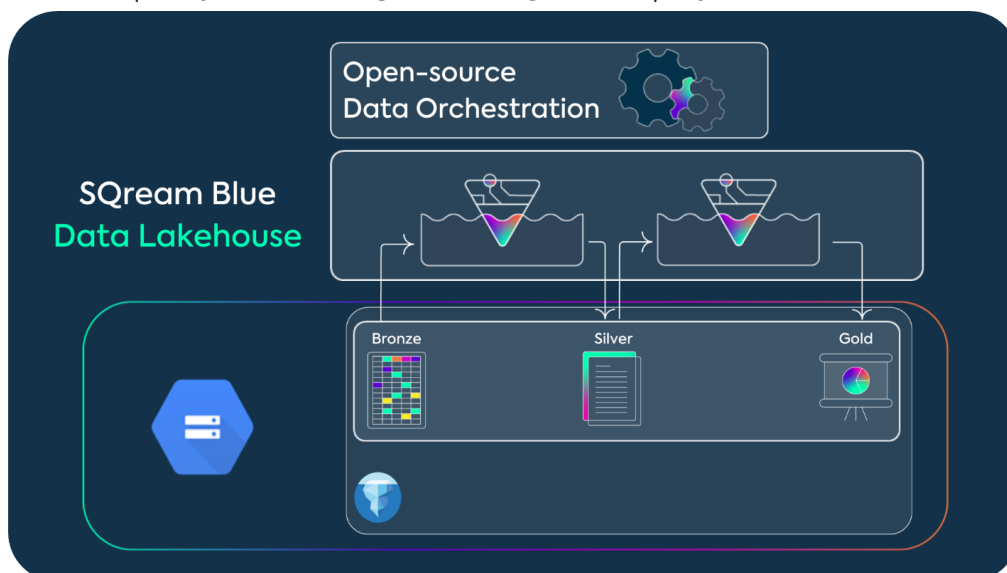
When **all data is prepared and accessible in one place**, data scientists can work **independently** on multiple projects without any engineer's assistance.

SQreamDB (On-premise) - Connect your notebook to SQreamDB and enable petabyte-scale data preparation with simple SQL. Save your prepared data as a DataFrame and move it on for further training and deployment.

SQream Blue (Google cloud) - Prepare data for ML directly on your cloud storage with Blue, the lakehouse for data preparation. Continuously feed your pipeline with data and improve its quality for more agile training and deployment.
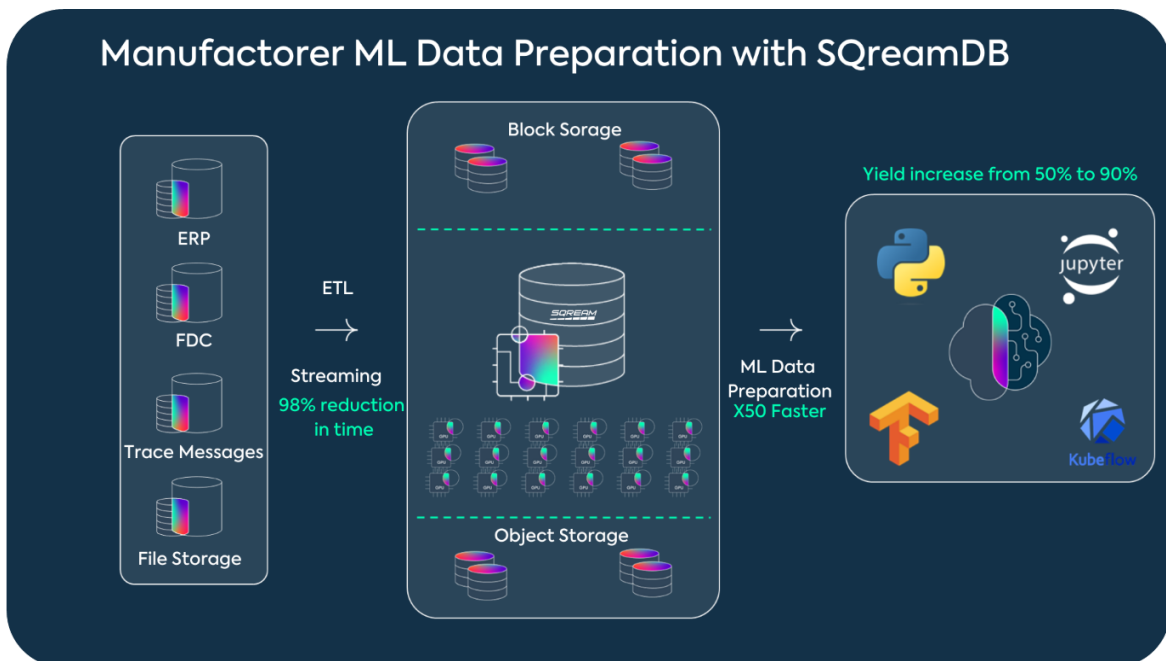
Case study - anomaly detection in manufacturing

A prominent electronics manufacturer in Asia was looking to improve its business performance and decided to embark on a project to centralize all data and create a machine learning (ML) platform. The goal was to identify and categorize production faults and defects, using anomaly detection algorithms, and maintain and adjust the production floor machines accordingly.

The manufacturer chose to implement SQreamDB as its central data repository and build its ML platform around it. With over 30TB of new data ingested daily SQareamDB is responsible for all structured data preparation for ML. With SQreamDB, the manufacturer could conduct **data preparation** for ML **50X faster** and improve its model precision by enlarging the test set.

This eventually led to substantial yield growth - from 50% to 90%!

Manufactorer ML Data Preparation with SQreamDB

Integrates with the MLOps ecosystem

Trusted by global leaders

About SQream

SQream makes it possible to ask bigger questions of extremely large, complicated data sets. With no off-limits questions, analysts get unprecedented new insights at exceptional speed. For too long, high costs and complexity have caused big data projects to fail at an alarming rate. To succeed, companies need a shortcut for asking bigger data questions to make better decisions. With SQream, you can finally "dig deeper, go faster, and reach anywhere" so you're able to ask those bigger questions – on prem, in the cloud, or at the edge. This is why organizations ranging from fast-growth startups to Fortune 100s all rely on SQream. Ask bigger.™
To learn more, visit sqream.com or follow us on twitter @sqreamtech.