SQREAM

WHITEPAPER

HOW FINANCIAL SERVICES COMPANIES OVERCOME BIG DATA CHALLENGES

sqream.com

INTRODUCTION

The financial landscape has changed drastically over the last twenty years, as technological advancements and the internet revolution have enabled new and more complex financial business models. These changes brought with them new ways of interacting with customers and other financial institutions, and facilitated game-changing competition by narrowing certain barriers to entry. The digital transformation of financial services produced never-before-seen volumes of data growing at an unprecedented rate.

A NEW LANDSCAPE BRINGS NEW CHALLENGES

Along with bringing great potential, the openness and accessibility of the new landscape increases the challenges around fraud detection, data protection, risk management, money laundering, customer retention, and more. These challenges compound the strain of fulfilling ever-changing regulatory compliance and governance regulations, which require financial organizations to instantly respond to and report on discrepancies and anomalies of relevant data.

Financial firms often hit and sometimes slam into the wall, trying to meet these challenges with antiquated data architectures that were not designed to handle the massive and growing data stores, and the complex queries required in today's financial environment.

Financial firms often hit and sometimes slam into the wall, trying to meet these challenges with antiquated data architectures that were not designed to handle the massive and growing data stores, and the complex queries required in today's financial environment. Moving data from one end of the pipeline to the AI models or applications that drive business can be painstakingly slow, with potentially negative business ramifications. Often, the amount of data that can be analyzed is only a subset of the data store - a result of inefficiencies in the data pipeline, or technological restrictions that make critical insights simply unobtainable.

LEVEREGING BIG DATA TO MEET EMERGING DEMANDS

This paper explores these challenges, and explains how financial institutions can transform their massive data stores from a major burden to a business driver by delivering critical insights faster. It will discuss how to propel forward data-driven systems such as fraud detection and risk management to meet new and emerging organizational demands, compliance requirements, and governance regulations. It will also delve into how to expand the usage and analysis of customer data for significantly increased competitive advantage.

CHALLENGE: FEEDING FRAUD DETECTION AI PREDICTIVE MODELS

Varying estimates have losses from fraudulent activities and claims in the financial industry reaching billions or trillions of dollars. While fraud is not a new problem, the challenge of fraud detection is growing significantly as services move online, and the means at the disposal of the criminals become increasingly advanced. Using AI predictive models based on customer account segmentation, approved and rejected transactions, cross validations, risk models and identified anomalies, companies are able to identify behavioral patterns and suspected fraudulent activities earlier. The data used to train the AI models are obtained from historical data, and the larger the data store used to feed the model, the better and more effective the predictions that can be achieved.

The challenge facing financial institutions is to ensure that they can prepare and segment their massive data stores efficiently and effectively on an ongoing basis to feed their fraud AI predictive models. When data stores reach the dozens and hundreds of terabytes and to petabyte levels, this challenge often becomes a major obstacle.

> ***The larger the data store used to feed the model, the better and more effective the predictions that can be achieved.***

CHALLENGE: FORWARD-LOOKING RISK MANAGEMENT BASED ON HISTORICAL DATA

Financial institutions face a growing and ever-changing set of regulations aimed at protecting customers, investors, and the organizations themselves. Regulations are both local and international, crossing nations and continents, creating a huge challenge for organizations to track, monitor, and report in compliance. Many of these regulations, including the Comprehensive Capital Analysis and Review (CCAR) and Dodd-Frank Act Stress Tests, require banks to analyze the risks hidden in their massive data stores (such as deposits and loans), and provide accurate results to be used for calculating asset liability and assessing liquidity risk.

> ***The challenge of analyzing exponentially growing data stores to provide reports on millions and billions of records simply cannot be met using old technologies that were not developed to handle these volumes of data.***

The challenge of analyzing exponentially growing data stores to provide reports on millions and billions of records is overwhelming. It simply cannot be met using old technologies that were not developed to handle these volumes of data, and are incapable of handling the segmentation and analysis of raw data required as new regulations emerge.

What is needed is technology that enables the combination of raw data from disparate sources, and the ability to accelerate data preparation and analysis, significantly reducing the time for risk management and regulatory compliance.

CHALLENGE:
INCREASING CUSTOMER LOYALTY IN AN ERA OF INDUSTRY TRANSFORMATION

Once upon a time, customers were loyal to their financial institution, taking pride in having chosen the right firm with whom to trust their business and assets. Today, customers work with multiple financial organizations, and moving between companies is much simpler and more commonplace than it once was. Historically, churn was a key success indicator primarily for communication companies. While financial organizations tracked customer movements, churn was not an issue of the magnitude that it is today. Banks took for granted that customers generally remained loyal, and if not for loyalty, the cumbersome process of changing banks itself was enough to retain customers.

But times have changed. Competition and customized products combined with an ease of movement not seen before have significantly reduced customer loyalty, and increased the importance of customer retention. Big data analytics enable financial institutions to better know their customers' preferences and behaviors, and provide improved products and financial structures while detecting patterns and early indications of trouble brewing in particular accounts. Data is used to simulate business scenarios such as changes to account conditions and interest rates, and to provide the basis for upselling new products and services.

Big data analytics enable financial institutions to better know their customers' preferences and behaviors, and provide improved products and financial structures while detecting patterns and early indications of trouble brewing in particular accounts.

The data needed to facilitate effective customer satisfaction and retention programs are available in the bank's data lakes, databases and data warehouses. Yet while financial executives are increasingly viewing big data analytics as the means with which to win in the highly competitive environments, exponentially growing data stores are proving to be a substantial obstacle. Financial organizations are challenged with rapidly ingesting, preparing and analyzing their data effectively and efficiently, so that they can effectively use that data to keep customers satisfied.

SQREAM PROVIDES BETTER INSIGHTS, FASTER.

Data and analytics are at the heart of the financial services industry. However, many financial organizations particularly struggle with disjoint data silos, complex data management, lengthy data preparation tasks, and never-ending queries and reports, all of which place a heavy burden on the organization. Data teams need a single, readily accessible data source that will empower them to understand the operations of business.

To address modern data needs, SQream DB provides financial organizations a modern data analytics solution that consolidates data sources and empowers data and business professionals to achieve crucial insights. SQream DB accommodates terabytes to petabytes of data, with a fully featured SQL interface, superior scaling, and a robust architecture based on accelerated hardware.

BUILT GROUND UP FOR HIGH-PERFORMANCE

SQream DB was created from scratch to empower data consumers, harnessing the raw brute-force power and high throughput capabilities of the GPU, with MPP-on-chip capabilities and a fully relational SQL database. Unlike in-memory databases or SQL translation layers for Hadoop, for example, SQream DB is a powerful data warehouse, designed for larger-than-memory, constantly growing data.

INDEPENDENT SCALING

SQream DB scales storage and compute independently, providing businesses with the precise storage and compute resources for their dynamic needs. The system's persistent storage can run on virtually any file system – whether local, distributed, on-premise or in the cloud - ensuring reliability and performance. SQream DB's optimized columnar storage system is partitioned both horizontally and vertically for best performance for heavy analytic operations like joins, aggregations, summarizations, and sorting.

Every SQream DB instance can be thought of as an MPP database in itself, with shared-data architecture. Each instance has full access to all data in the persistent storage layer, with permissions managed by the service layer above the SQream DB instance based on the user's role. These instances can be launched or shut down at will as requirements change.

SQream DB's architecture further boosts performance by relying on storage and operating system caches to transparently and automatically cache data, which can be re-used by other SQream DB instances as needed.

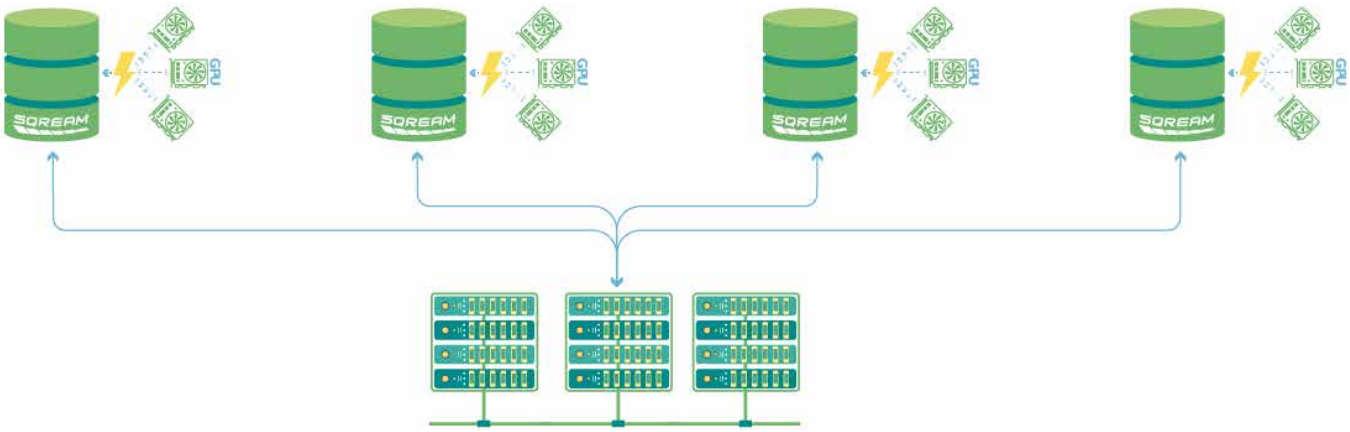

Figure 1 - SQream DB can scale both compute and storage separately, maximizing flexibility.

AUTOMATIC TUNING, SELF-MANAGING

SQream DB contains hundreds of optimizations and automations designed to let businesses focus on data, rather than data management.

Unlike traditional databases that require a team of administrators to finesse and manually tune processes, maintain indexing, update views and projections, SQream DB was designed for frequently changing, modern workloads. Built to handle worst-case scenarios, it is optimized for the huge datasets common in financial organizations, where typical database optimizations struggle.

ANALYZE RAW DATA DIRECTLY AND EASILY

Automatic tuning is a key enabler for analyzing data without intermediate steps, and is a major part of SQream DB. The brute power of the GPU allows SQream DB to analyze data immediately after load. This is in stark contrast to most data warehouses, which require time-consuming and insight-limiting processes like indexing, cubing, projecting, etc.

During data ingest, SQream DB automatically and transparently prepares all data for immediate, fast analysis – with no user intervention required.

INTEGRATION AND FAST DATA INGEST

One of the most common tasks for any analytics database is loading data from an external source. SQream DB ingests up to 3.5 TB per hour per GPU from a variety of sources, either directly from flat files like CSV or Parquet, or through a variety of industry accepted ETL tools. SQream DB can also read data directly from external sources using the external table syntax, avoiding premature data loading.

It is common for SQream DB to provide the analytics database, where Apache Kafka serves as the messaging queue system, and Apache Spark provides transformations. In such installations, SQream DB will be the layer bridging the applications, with persistence store for analysis.

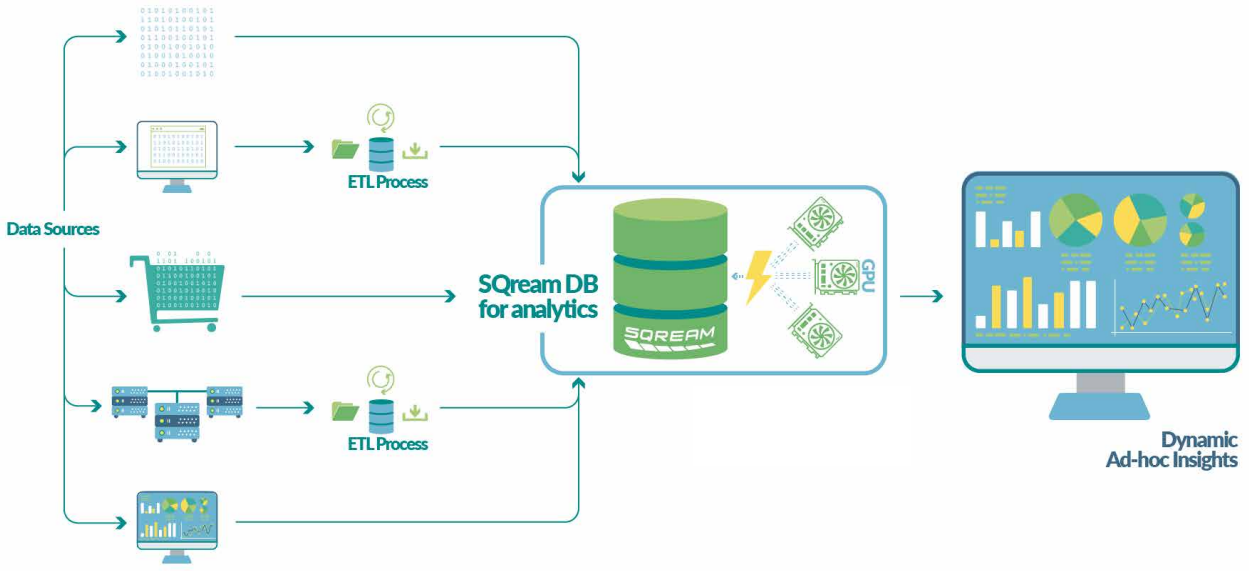

Figure 2 - A typical SQream DB implementation, with no intermediate data preparation needed

PROVEN RESULTS WITH GLOBAL FINANCIAL ADVISORS

In a recent report published by Citihub Consulting, a global financial market IT advisory firm, SQream DB outperformed leading databases on high-volume queries. The test, built by Citihub and a tier 1 investment bank, generated a large FSI risk dataset suited specifically for FRTB reporting and analysis. The report found SQream DB to be the best performer for massive datasets by a large margin.

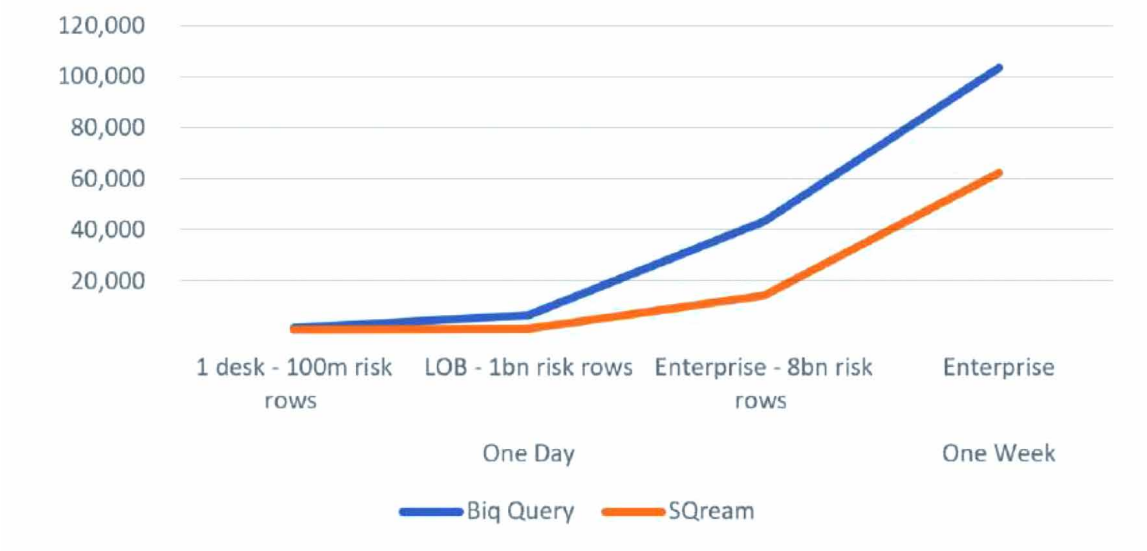

Figure 3 - Moving average over 5 days

"...we were surprised at how quickly SQream's performance overtook BigQuery as the volume scaled up"

"...SQream does not suffer from the query optimisation issues that we have seen elsewhere"

Citihub Consulting

MAJOR FINANCIAL INSTITUTION SOLVES THEIR DATA DILEMMA

One of the organizations successfully using SQream DB is a leading global financial services provider that needs to conduct long-term historical analytics on tremendous amounts of data to support the development of advanced AI fraud detection techniques.

The company's Data Science team initially used a shared Hadoop infrastructure for analytics, which required time-consuming data preparation, and limited data discovery. Long-running queries were frequently interrupted or de-prioritized by other system users, and queries on particularly large datasets couldn't be executed at all. The company sought a solution that would alleviate these challenges and enable unrestricted data exploration.

THE SOLUTION – AN ACCELERATED DATA SCIENCE ENABLER

After testing several solutions, the Data Science team integrated SQream DB into their Spark and Hadoop-based architecture to power accelerated mass-data analytics. The new architecture offloads important data to a SQream DB cluster, enabling the Data Science team to model, test, and deploy advanced AI models in a fraction of the time and resources, with on-the-fly ad-hoc querying abilities.

With SQream DB, fraud detection queries scanning thousands of merchants and retailers were completed in minutes, where the previous system would take several hours. Previously infeasible queries were completed in minutes, with minimal data preparation. SQream DB accelerated load performance by a factor of twenty, and reduced disk use by 30%.

The Data Science team is now able to broaden the timeframe of historical queries about credit card transactions from months to years, resulting in more accurate models for predicting fraud, protecting customers and retailers around the world.

> ***We are always looking for new and better ways to analyze our huge data stores. SQream DB allows our Data Science team to generate the insights we need with a lean and performant addition to our existing Hadoop system.***

SQREAM
GPU DATA WAREHOUSE

ABOUT SQREAM

Visit sqream.com to learn more about how financial organizations around the globe turn their data challenges into business opportunity.

Bring the power of SQream DB to your business:

info@sqream.com
sqream.com
@SQreamtech