# FAST, FINE-GRAINED ANALYTICS WITH SQREAM DB AND TABLEAU

SQREAM DB WHITE PAPER

v1.0

 sqream.com

# PREFACE

This white paper is for any data professional – on the infrastructure, data engineering, data science or BI side - who is experiencing slow Tableau dashboards. This paper will describe the root causes of Tableau performance issues and explain how a new GPU-based data warehousing approach can alleviate many of these challenges.

# INTRODUCTION

Organizations worldwide are facing the challenge of effectively analyzing their exponentially growing data stores. Today's businesses increasingly rely on leveraging their growing data stores to extract actionable insights into customer behavior, security anomalies, risk analysis, stock and inventory predictions and more. This shift to a data-driven approach has caused many organizations to store their data in large data lakes, yet their analytical platforms struggle with the growing volumes and velocity of data.

The BI pipeline - from the database infrastructure down to the BI tool - has become strained and slow, resulting in several problems:

- Slow access to data
- Out-of-date dashboards that cannot be frequently refreshed
- Inflexible infrastructure that prohibits ad-hoc queries

The main culprit behind these issues is the lack of advancement in data processing and analytics technology. In the past 30 years, most advancements in this field were small innovations that focused on optimizing specific use-cases or workload sizes. These innovations, in both the infrastructure and the end-user BI visualizers, would become obsolete or outpaced within just a few years due to tremendous data growth.

SQream DB brings a new technological approach that eliminates data professionals' struggles at the source, resulting in fast, accurate, up-to-the-minute dashboards from which vast new insights can be extracted. SQream's GPU-accelerated data warehouse is designed for fast, unrestricted access to an organization's full scope of data, even when data grows exponentially. SQream's technology enables Tableau users to explore and productize interactive dashboards that power many aspects of modern data-driven organizations.

## DATA PROFESSIONALS USING TABLEAU WITH SQREAM DB WILL BENEFIT FROM:

- Fast access to previously unobtainable insights, for smarter and more accurate business decisions
- Near real-time, fine-grained, drill-down dashboards
- Access to more data, without the cost of scaling expensive solutions
- Increased employee productivity through fast, unrestricted data access
- Significant cost savings on hardware, software, and management
- Simplified administration and operational efficiency

# COMMON CHALLENGES TO EFFECTIVE ANALYTICS

## 1. TIME-TO-ANALYSIS: SLOW DATA INGEST AND PREPARATION CYCLES

Data consumers in general, and particularly data scientists and BI analysts are increasingly frustrated by the speed of accessing data in the organization. Many analytic workloads require lengthy ingest processes and data preparation tasks, including correlations, pre-computations, filtering, enrichment, and other techniques. These data preparation tasks are often performed during off-peak times to avoid overloading the data infrastructure.

Unfortunately, this data preparation severely limits the analysts' ability to perform data exploration. The dimensionality of the data is typically reduced (e.g. summarizing customer transactions, collapsing many transactions into one) to allow for faster querying later on. This trade-off between coarser data granularity and faster querying comes at the high cost of making future fine-grained drilldown analysis impossible.

## 2. COMPLEXITY OF THE DATA PIPELINE

Poor performance, inflexibility, and difficult scaling ultimately lead to complex solutions. Many companies now not only have database administrators, but also data engineers, data custodians, and data stewards. Data engineers work on the data-management side, maintaining the organization's data infrastructure, while database administrators now focus almost exclusively on fine-tuning database performance. Database administrators of traditional data warehouses have a large array of tweaks and optimizations that have been built over 40 years of patching and addressing scalability issues with these solutions.

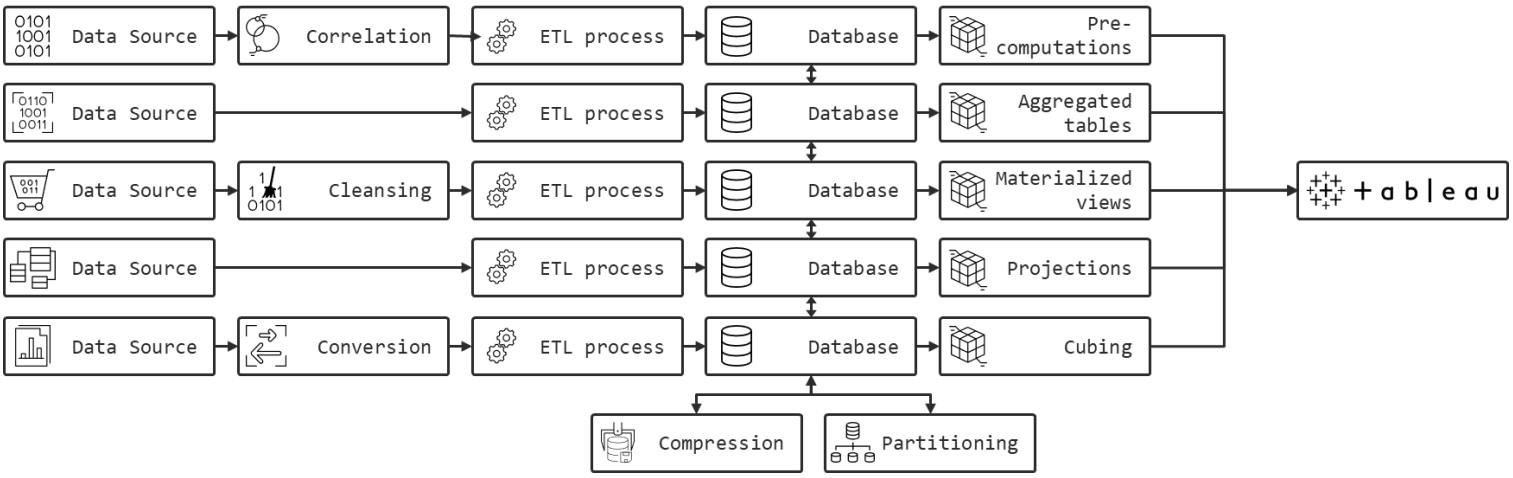


*Figure 1 - Common data pipeline with traditional data warehousing methods are quite complex*

These problems trickle down to the BI tools and visualizers. For example, one misguided attempt to create a new dashboard could generate complex queries that can bring the data warehouse to its knees, causing outage for other lines of the business. These outages are painful and frustrating to infrastructure teams and data consumers alike. However, as BI analysts are the masters of the data in their domain, they are the ones most frustrated by the business impacts of data being inaccessible due to complexity, sub-sampling, and inflexible tools and infrastructure.

## 3. SCALABILITY: INSUFFICIENT SUPPORT FOR GROWING DATA

Traditional databases and even distributed systems have trouble scaling to support exponentially growing volumes of data. While a traditional database may struggle with data scale, a distributed database may have trouble running joins, and is harder to maintain.

With distributed data systems, performance issues are typically dealt with by adding nodes and re-distribute the data. With a traditional data warehouse, the typical solution to performance issues is to buy a bigger appliance, with more RAM and CPU. Both "solutions" are not actual solutions, but rather a temporary band-aid until the data grows again.

# THE NEW APPROACH: A SIMPLIFIED DATA ARCHITECTURE

SQream DB is a new-generation data warehouse that minimizes time to analysis, eliminates complex data pipelines, and addresses the poor scaling of traditional solutions.


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

## MINIMIZING TIME-TO-ANALYSIS: 'LOAD-AND-GO' ARCHITECTURE

SQream DB's GPU-accelerated architecture and automatic optimizations are a key enabler for analyzing data without intermediate steps. The raw, brute power of the GPU allows SQream DB to analyze data immediately after loading. This capability is in stark contrast to most data warehouses, which require time-consuming and insight-limiting processes like indexing, cubing, and projecting. During the ingest process, SQream DB automatically and transparently prepares data for immediate, fast analysis – with no user intervention required.

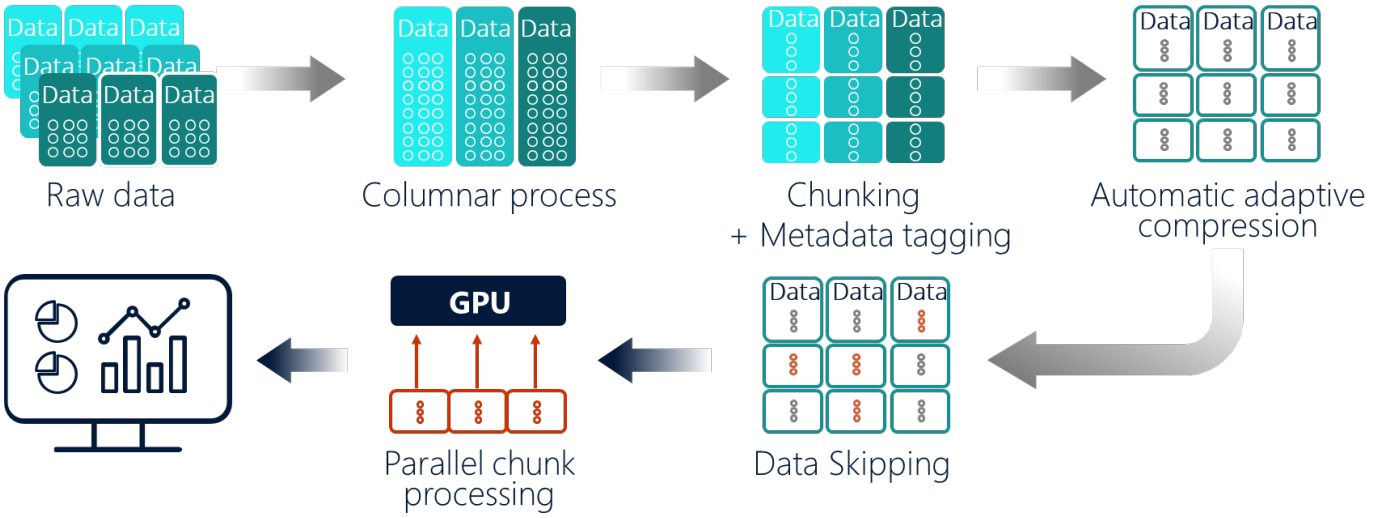


*Figure 3 - Load-and-Go architecture. Every step of the process is completely automated.*

The SQream DB table is optimized for fast bulk ingest, loading tables at speeds in excess of 3TB/hour - a fraction of the time it takes other solutions.

Tableau users can take advantage of SQream DB's raw-data-is-king approach. Because SQream DB implements standard SQL and ODBC connectivity, users can connect and begin data exploration within seconds, without limiting the dimensionality or depth of the query. Joins are available on every column, with no indexing needed.

## FLEXIBLE ANALYTICS: CUTTING THE PREP WORK

SQream DB brings a new approach to data warehousing. From the user's perspective, SQream DB looks like a standard analytic SQL database, but the underlying architecture is that of a GPU-accelerated data warehouse with a highly automated "Load-and-Go" process designed to accelerate and simplify data processing. Data preparation is further reduced, since SQream DB stores data on-disk in a shared-data architecture that eliminates the need for distribution keys or creating projections.

## FASTER QUERIES, FASTER DASHBOARDS

Traditional data warehouses rely on a fixed set of resources for running all scenarios. In contrast, SQream DB can allocate additional resources to handle a varied workload by combining available CPU, GPU, RAM, and storage resources.

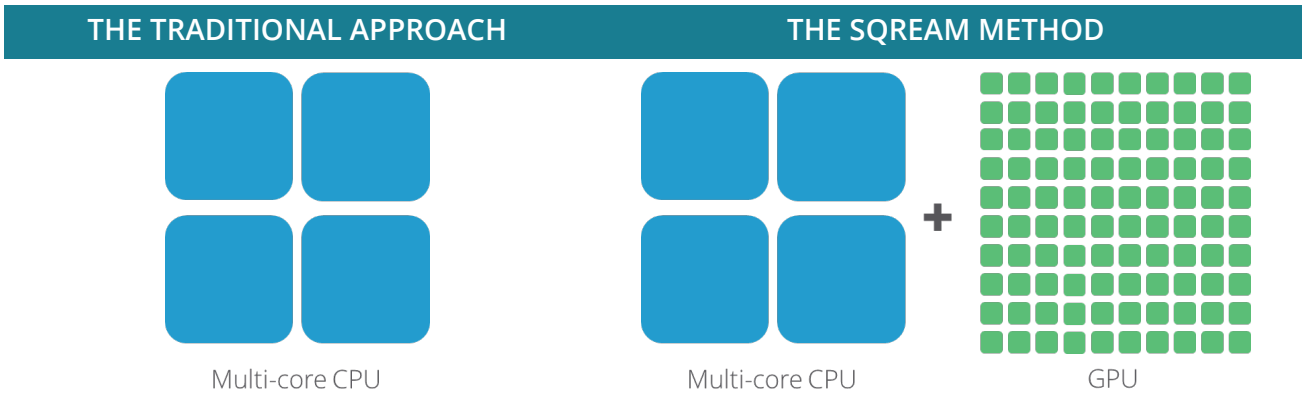


*Figure 4 - CPU technology vs. GPU technology*

This balance of CPU and GPU operations is key to ensuring optimal performance. GPUs excel at performing repetitive operations on large amounts of data in many streams. The result is a faster response times, even in the most complex dashboards.

## UNLIMITED SCALABILITY: ANALYZE MORE DATA FOR BETTER INSIGHTS

SQream DB can scale to unlimited data sizes, many data consumers, or both. Because compute is decoupled from storage, scaling is available for just storage, just compute, or both. This is a key factor in providing excellent performance for organizations of any size, while reducing costs. Growing the system in either direction doesn't affect the data availability or integrity, which means that SQream DB can scale to virtually unlimited data sizes.

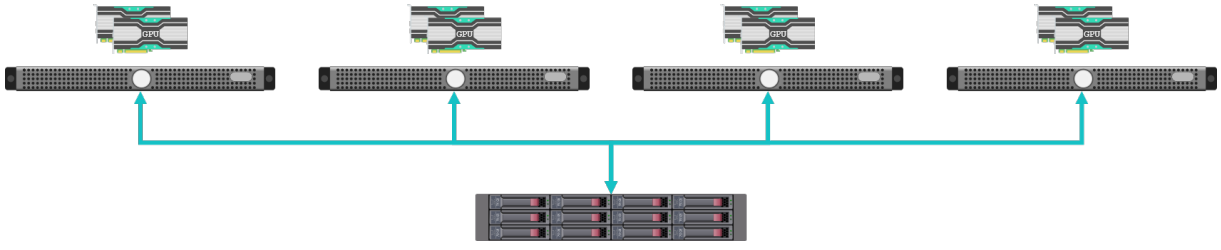

*Figure 5 - SQream DB has a shared-data architecture that enables independent scaling of storage and compute*

Because of its unlimited scalability, SQream enables data consumers to analyze significantly more data without having to invest in complex, distributed solutions. SQream DB supports data consumers and infrastrcture teams alike by transparently optimizing, compressing, and partitioning data with no intervention needed.


*Figure 6 - The SQream DB hyper-partitioned table, designed to scale up to trillions of rows with ease*

SQream DB's tables support scalability by hyper-partitioning data in multiple dimensions – a process we call chunking. Chunking is automatically and transparently performed during ingest. A user can query and interact with all of their data, just like a regular table. This capability allows SQream DB tables to grow to sizes that other databases can't support, while retaining familiar management functionality.

## SIMPLE YET FLEXBILE

Owing to the automated approach of its Load-and-Go architecture, automatic adaptive compression, and dynamic workload management (WLM), SQream DB helps simplify data architectures. Unlike other data warehouses, SQream DB dynamically responds to changes in the analytic workload, automatically tuning queries and system resources on-the-fly. Compression is automatic, as is the SQream DB hyper-partitioned table. All operations are performed via standard SQL interfaces and standard connectors for maximum flexibility.

SQream conforms to the ANSI SQL-92 standard, making interaction with the system no different than with any other RDBMS. With SQream DB, however, data professionals will benefit from many advantages under the hood. When a query is issued in Tableau via ODBC, the SQL command is instantly parsed and converted to Relational Algebra for further processing and optimizations inside the SQream DB query engine.

# GETTING THE MOST OUT OF SQREAM DB AND TABLEAU

## LET SQREAM DB DO THE WORK, FAST

Some best-practices tell you that you should create an "extract" in Tableau to speed up subsequent processing and prevent overloading your data warehouse. With SQream DB, extracts are a thing of the past.

SQream DB is optimized for exploration and ad-hoc querying. The relevant data for the query is identified by SQream DB, and all relational operations and transforms are run on the GPU. Within seconds, the resulting data set is sent back to Tableau in "Live" mode, ensuring that the results are up-to-date.

## FORGET INDEXING

With SQream DB, there's no need for pre-computations like creating and managing indexes, creating projections or cubes, or even materializing views. SQream DB handles the query optimization transparently, and even automatically handles the load balancing so you can just use Tableau normally in "Live" mode.

## FILTER IN ADVANCE / PAUSE THE WORKBOOK FOR EASIER WORK

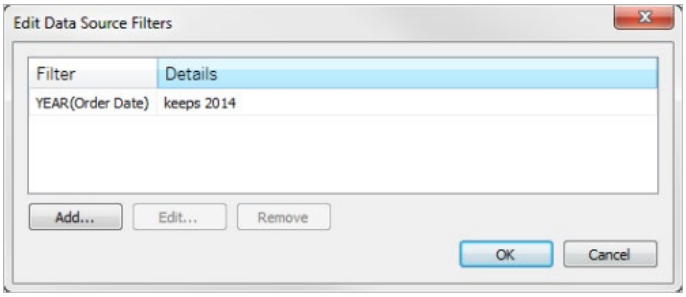


*Figure 7 - A Tableau data source filter can be used to reduce query time*

Clever filtering is a primary part of what makes SQream DB so fast. However, when you first drag-and-drop a set of columns into your worksheet, Tableau will start firing off queries to the database. At scale, when working with trillions of rows, this can take some time! It's best to add data-source filters or pause the worksheet as you add columns until you've created the necessary filter to limit the scope of the queries before you're ready to run them. With filters in place, trillions of rows can now be accessed and analyzed in seconds.

## TRACK PERFORMANCE TO PINPOINT ISSUES

Use built-in Tableau tools to monitor query performance. Tableau contains a great built-in tool called the Performance Recorder. Many SQream DB users use the performance recorder in conjunction with SQream DB's own logging system to identify slower queries and optimize performance.


*Figure 8 - The Tableau Performance Recorder can be used to identify processes that slow you down*

## ISOLATE DESIRED COLUMNS

Like many columnar databases, SQream DB benefits from selecting only the relevant columns. When creating a group from a selection, be sure to only select the relevant columns. Adding additional columns can add dimensionality but will decrease performance.

# FIX TABLEAU ISSUES AT THE SOURCE

The common challenges users have with Tableau are mostly caused by legacy data pipelines which weren't designed for modern scales of data. SQream DB and Tableau are the ideal solution for getting great visual analytics out of your big data. With SQream DB as Tableau's analytics back-end, users will benefit from fast ingest, dramatic time savings on data preparation, and get high-performance visual analytics on any scale of data.

# ABOUT SQREAM

SQream DB combines performance, flexibility and ease-of-use, empowering and accelerating your data science and making discovery insights in your data fast, so that you can focus on the core of your business, instead of on the infrastructure. Bring the power of SQream DB to your business with a free trial in the cloud or on-premise at https://sqream.com/try-sqream-db.