20x to 100x Faster Analytics Through Data Warehouse Augmentation

Bring Critical Analytic Workloads Into the Modern Age

SingleStore

Table of Contents

Introduction: Putting Today's Data Warehouses in Context

Data-warehouse-as-a-service market is expected to reach 4.3 billion USD by 2026

The data warehouse is an indispensable tool for many modern enterprises—and their popularity shows no signs of slowing. According to a February 2021 report by Mordor Intelligence, the data-warehouse-as-a-service market was valued at USD 1.44 billion in 2020 and is expected to reach USD 4.3 billion by 2026, representing a compound annual growth rate of 20 percent.

This sustained popularity is no surprise: on-premises and in the cloud, data warehouses have become effective tools for performing complex data analytics, reporting, and historical comparisons. Many of today's data warehouses power business intelligence (BI) and reporting workloads that enable organizations to quickly aggregate and analyze large amounts of data from multiple sources to drive insights.

4.3B

The data-warehouse-as-a-service market is expected to reach 4.3 billion USD by 2026.

Source: Mordor Intelligence, February 2021

Traditional Data Warehousing Flow

In most of these architectures, data is drawn from online transaction processing (OLTP) applications or other data sources, usually in batch mode via some sort of ETL or ELT process that runs at set intervals such as every 2 hours, 4 hours, 6 hours, 12 hours, or 24 hours, depending on the business needs. As part of this integration process, the data is aggregated, transformed, and loaded into a common database schema for easy access via SQL statements--or via point-and-click BI tools that generate SQL statements under the hood. This allows users to easily query the warehouse and view the results through dashboards, reports, and other front-end applications. *(Figure 1)*

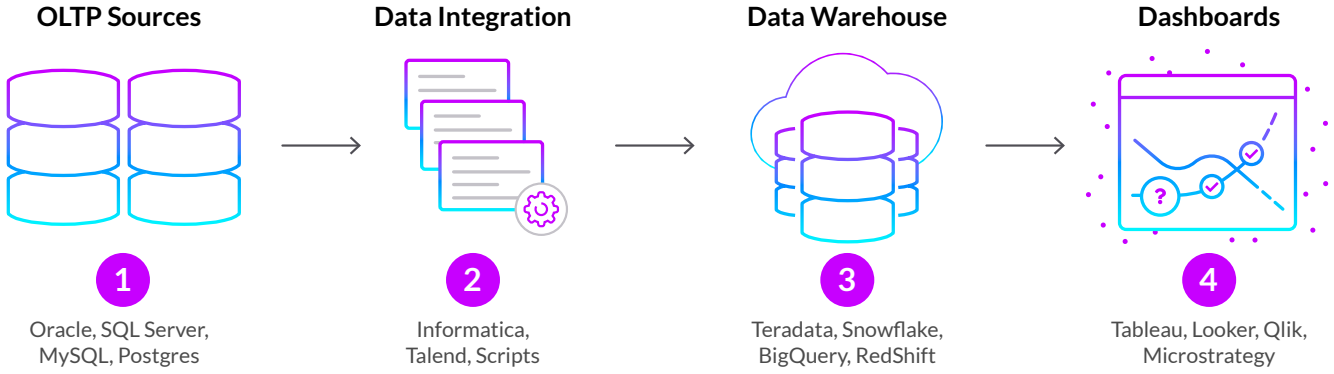


Figure 1: Common data flow for analytics and data warehousing

Understanding the Limitations of Traditional Data Warehouse Architectures

Traditional data warehouse architectures were not designed to handle the speed, scale, and agility that today's enterprises need to succeed. As data grows in complexity and scope, yesterday's data engineering workflows struggle to handle new types of data and real time analysis scenarios. New forms of real-time data require streaming data ingestion and immediate, low-latency analytics to be valuable.

Unfortunately, popular data warehouses--including Teradata, Snowflake, Google BigQuery, and Amazon RedShift—typically depend on rigid, batch-oriented ETL or ELT technologies to capture, ingest, cleanse, and transform data into a structured format that fits a predefined schema before it is available for analysis and reporting. This, in turn, negatively impacts the application and user experience.

As a result of these rigid, traditional workflows, enterprises encounter four primary data bottlenecks that impede the performance of the data warehouse. They include:

1. **Streaming Ingest and Analytics:** Because they were built for complex queries over large structured data sets, these data warehouse architectures are not optimized to ingest, process, and analyze fast moving streaming data, which is necessary to drive insights and actions in real-time or near real-time.

2. **ETL Batch Windows:** In most cases, complex data-integration and transformation processes must be completed before a data warehouse can drive intelligence to downstream users and applications. These ETL batch windows could range anywhere from two hours to 24 hours, depending upon the business priorities. During this time, data is "held hostage," preventing applications and users from obtaining visibility into the ever-changing dynamics of the business.

3. **Low-Latency Queries:** Traditional data warehouses are great at running known queries against pre-aggregated data sets, but they are not optimized for fast query performance or ad-hoc analytics. Inherent query latencies prevent business users from obtaining timely insights.

4. **High Concurrency:** Traditional architectures tend to break down under the duress of high-concurrency workloads, in which a large number of users and a high number of queries are simultaneously executed to populate interactive dashboards, applications, or reports. Scaling data warehouses to support high concurrency workloads can be extremely costly.

What if you could achieve

100x

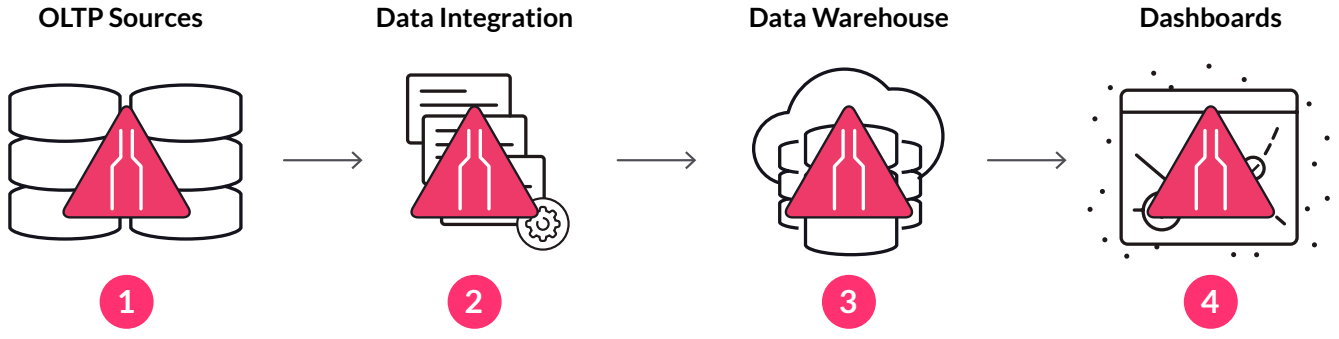
faster analytics and performance compared to your data warehouses and associated data pipelines while driving significant cost reductions?

In this eBook, you will learn how you can dramatically increase data warehouse performance and accelerate time-to-insights by enhancing your data ingestion capabilities, increasing query speed, and providing exceptional concurrency for all types of analytic activities—often at only one-third the cost of running legacy infrastructure.

** These bottlenecks and challenges are summarized in Figure 2*

Common Data Warehouse Bottlenecks

Traditional data warehouses are hindered by four primary bottlenecks:



Limited support for streaming ingest: Data warehouses were not architected for parallel, high-throughput ingestion of streaming, real-time data.

ETL batch windows: Batch windows inject significant delays into the data flow, are often scheduled during off hours and often take too long to complete. That means dashboards and reports reflect data that is hours or days old.

Query latencies: Data warehouses were not optimized to handle low-latency queries, such as is required for fast analytics applications and interactive dashboards.

Concurrency limitations: Traditional data warehouses break down under the duress of high concurrency workloads supporting large groups of users, and can be expensive to scale.

Figure 2: Common bottlenecks associated with the data warehousing flow

A Unified Database for Fast Analytics

"SingleStore can process complex queries with large data sets in 1 to 3 milliseconds. The closest Snowflake or BigQuery can get is in the 200 millisecond range."

- B2B Startup

SingleStore is built from the ground up as a distributed, highly-scalable, unified database that can deliver maximum performance for both transactional and analytical workloads. It unifies transactional and analytical processing on diverse data (unstructured, semi-structured, and structured) in a single engine—with the ability to use standard SQL to join these diverse native data types. With **20x to 100x** the performance at one-third the cost of legacy infrastructures, SingleStore delivers unmatched speed, scale, and agility in a powerful, cloud-native relational database.

Up to 100x faster

Drive 20x to 100x faster analytics by augmenting your data warehouse with SingleStore.

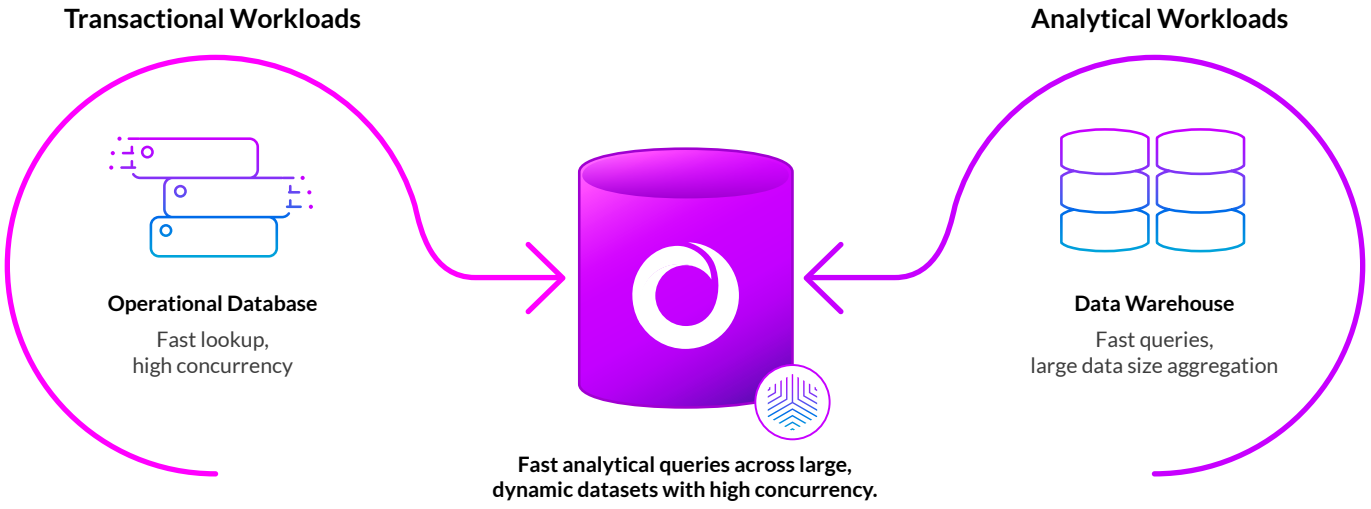


Figure 3: SingleStore's unified database with patented Universal Storage

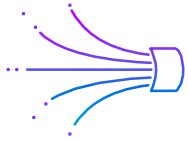
SingleStore is ideal for running fast analytical queries across large, dynamic data sets, with consistently high performance. SingleStore's patented Universal Storage delivers a breakthrough in database storage architecture that allows both operational and analytical workloads to be processed using a single table type. It consists of two key components:

- An in-memory rowstore that easily handles intensive data-processing demands, allowing massively concurrent updates with exceptional response times of just a few milliseconds and
- A memory- and disk-based columnstore that accommodates billions of rows of data, utilizing an 80 percent compression ratio

This unique Universal Storage architecture brings together the best of both worlds: the exceptionally fast transactions and lookup performance of an operational database, together with the scalable analytics of a data warehouse. While the in-memory rowstore is great for super low-latency queries, the columnstore ensures fast reads—even for analytical operations that involve scanning billions of rows of data.

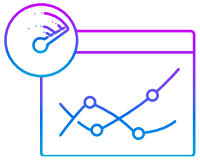
Data Warehouse Augmentation with SingleStore - Key Capabilities

SingleStore is the unified database that is optimized for parallel streaming data ingestion, super-low-latency queries, and high concurrency to help you process, analyze, and act on data instantly.



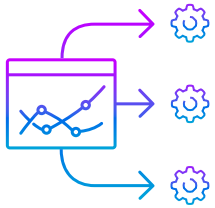
Parallel, high-scale streaming data ingest

Ultra fast ingest:
SingleStore's parallel, high-throughput engine can easily handle millions of events per second from distributed data sources such as Apache Kafka, Amazon S3, Azure Blob, Filesystem, Google Cloud Storage, and HDFS data source. This is a common bottleneck for traditional as well as cloud data warehouses and processing engines—but not for SingleStore.



Blazing fast queries

Super low latency:
SingleStore delivers ultra-fast query response for both live and historical data using familiar ANSI SQL. Query latency of 10 milliseconds or less is typical, even with thousands of concurrent users.



Unparalleled scalability

High concurrency:
SingleStore's elastic, scale-out architecture includes a distributed, massively parallel data processing engine. It delivers consistent, predictable response rates, even with high data ingest and concurrency of tens of thousands of users. SingleStore powers reliable, highly responsive dashboards with plenty of capacity for interactive analytics.

Fast analytics on dynamic data for complex analytical queries

Figure 4: SingleStore key capabilities for enabling fast analytics

Augmenting Data Warehouses with SingleStore—Key Patterns

Making significant improvements to your data warehouse doesn't necessarily mean starting over. Leading organizations are **augmenting** their data warehouses with SingleStore to power fast dashboards and intelligent, data-intensive applications.

A growing number of organizations are augmenting their data warehouses with SingleStore to enable faster analytics at lower costs, both for on-premises systems and for cloud data warehouses. Many SingleStore customers experience **20x to 100x** performance gains and rapid time-to-insights by augmenting Teradata, Snowflake, Amazon Redshift, and Google Big Query data warehouses with SingleStore to power their analytics, applications, and dashboards.

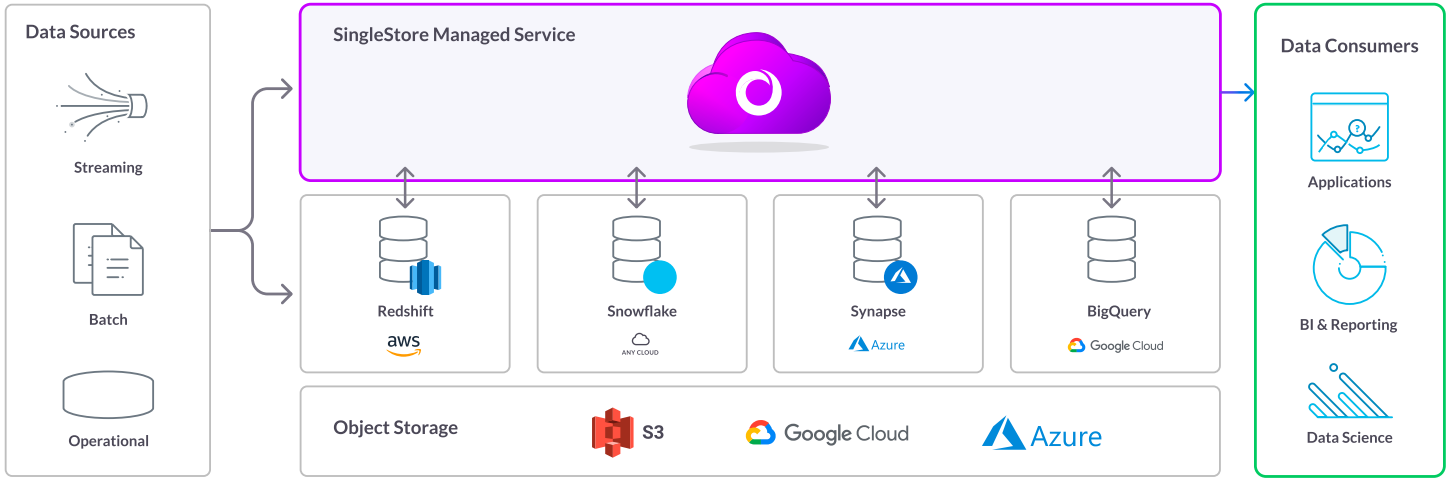


Figure 5: Augmenting Data Warehouses with SingleStore

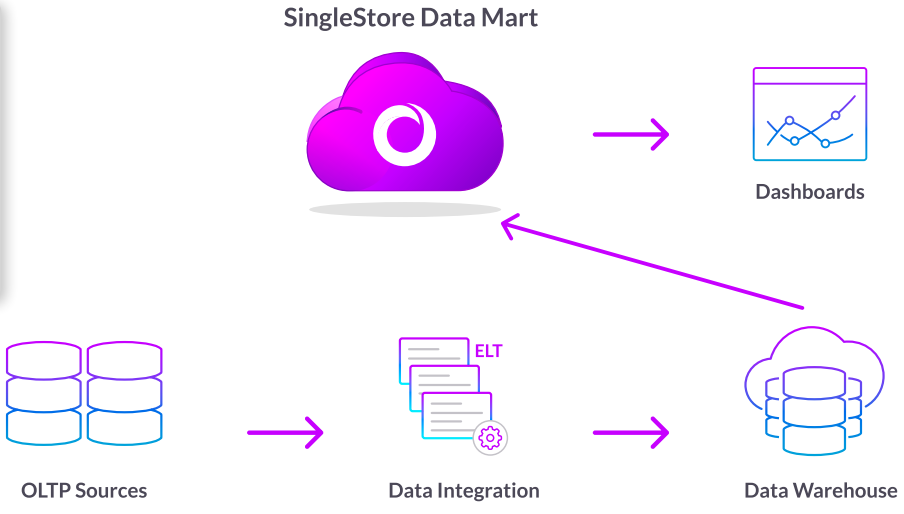
SingleStore

Most SingleStore customers follow **three** popular augmentation patterns.

Augmentation Pattern 1: SingleStore as a Data Mart

When is this pattern ideal?

Ideal for **improving the performance of key applications and dashboards**—including query latency, concurrency, and total cost of ownership (TCO).

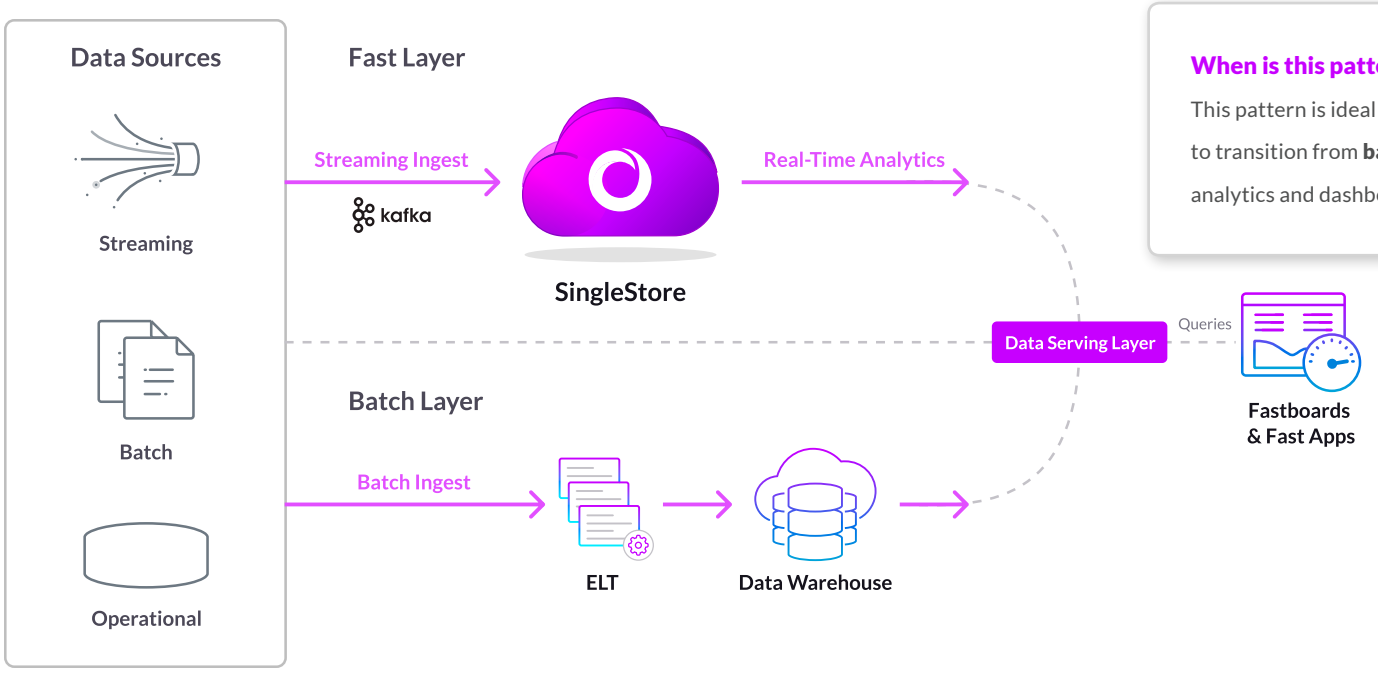


One popular augmentation pattern involves utilizing **SingleStore as a data mart** to power fast analytics, dashboards, and applications.

This pattern involves moving relevant datasets from the data warehouse into SingleStore that is optimized for fast queries and high concurrency. With schema mapping and continuous data loading, SingleStore augments critical analytic workloads to enable fast analytics while keeping other workloads intact.

With SingleStore, it is easy to pull the data you need for fast dashboards from your data warehouse into a SingleStore instance, yet continue to use the data warehouse for other workloads, such as routine financial reporting and data science use cases. This augmentation pattern is a proven way to improve the performance of your analytic applications, while driving down the total cost of ownership related to your data warehouse.

Augmentation Pattern 2: The Lambda Architecture



When is this pattern ideal?

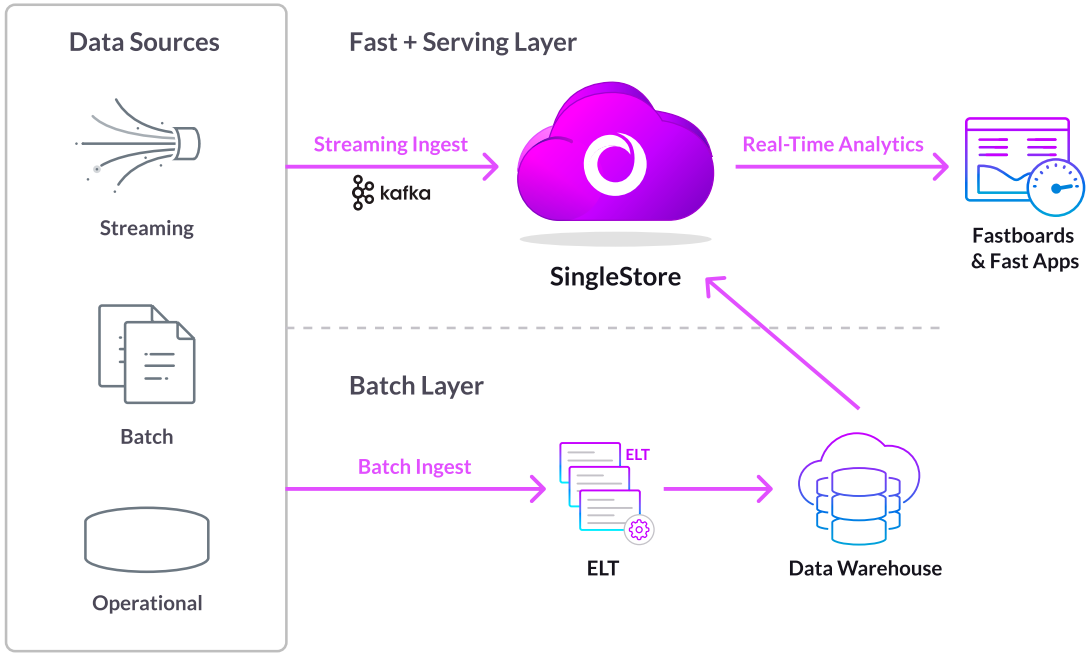
This pattern is ideal when you need to transition from **batch to real-time** analytics and dashboards.

The Lambda architecture processes large amounts of data by providing a platform to concurrently access both batch-processing and real-time streaming methods. The Lambda architecture forks data into two paths: a streaming path or fast layer; and a more conventional batch layer.

The Lambda pattern is optimal when your service levels stipulate a narrow window between the time a piece of data is born and the time that it must appear in a dashboard or application. Time-sensitive data or real-time data can be directly streamed into SingleStore using SingleStore Pipelines, while the rest of the data is loaded into the data warehouse via a batch-ingestion process. When queried, a serving layer merges both views to generate appropriate results.

As shown in the figure above, streaming data is ingested directly into SingleStore via the fast layer, while batch data follows the traditional route into the data warehouse via the batch layer. When queried, the serving layer merges the speed views and batch view to generate appropriate results.

Augmentation Pattern 3: Fast Lambda or Lambda+ architecture



When is this pattern ideal?

This pattern is ideal when you want to transition from **batch to real-time** analytics while improving query latencies and **boosting performance.**

This **Lambda+** pattern combines Patterns 1 and 2 to enable streaming ingest while simultaneously driving low latencies and high query performance. It allows you to combine older curated data with newer streaming data to obtain consistent analytics from batch and streaming data.

In this pattern, SingleStore performs the functions of the fast layer and the serving layer of the Lambda architecture. Customers use this pattern when they are transitioning from batch to real-time analytics ingestion, while supporting high-concurrency queries for dashboards and data-intensive applications.

Customer Case Studies

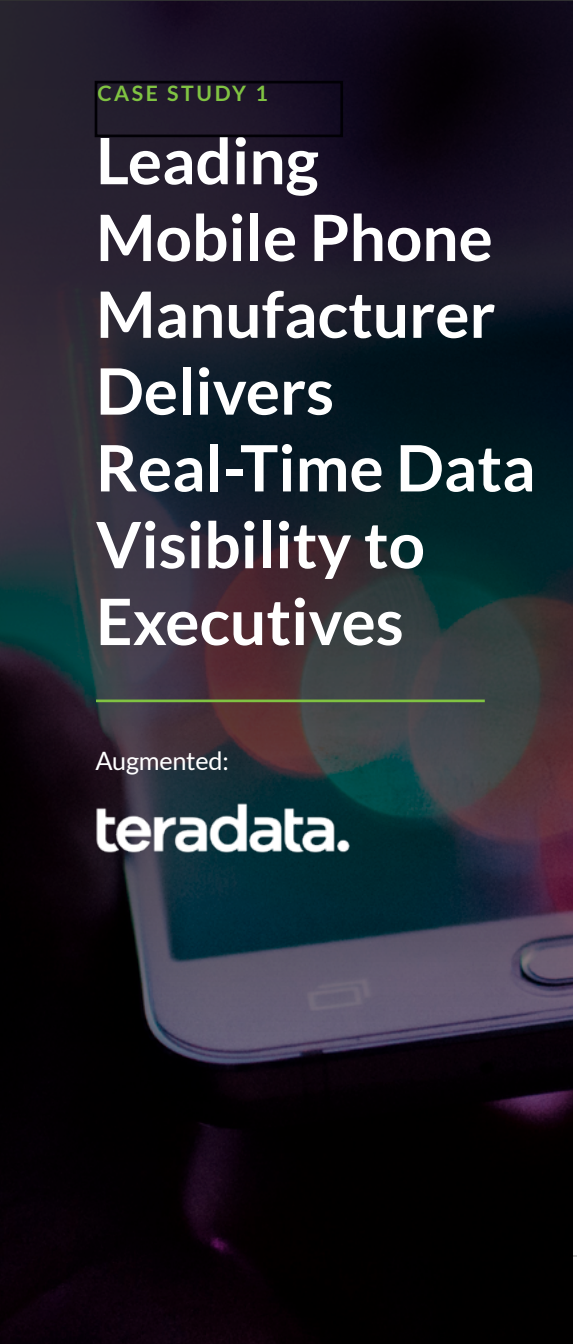
SingleStore

CASE STUDY 1

Leading Mobile Phone Manufacturer Delivers Real-Time Data Visibility to Executives

Augmented:

teradata.



LEADING GLOBAL MOBILE PHONE AND ELECTRONICS MANUFACTURER

Situation

Senior executives at this fast-moving electronics manufacturer rely on a Tableau dashboard to monitor the real-time sales and market movements of mobile devices, which requires visualizing data by device, region, price point, product attribute, and many other dimensions.

Challenge

Slow and lagging performance of the executive dashboards meant executives had to wait many hours to obtain new insights. These delays adversely impacted product launches, marketing campaigns, and supply chain operations. For example, managers could not quickly determine how much raw materials were required to satisfy fluctuating consumer demands.

Teradata, which powered this executive dashboard, couldn't scale to handle the data growth and concurrency requirements of **400+ queries per second.** Additionally, the electronics manufacturer had to ingest **4 billion** rows of new data each day and this led to significant delays: as long as 10 hours to process and display the latest data in the dashboard.

Solution

Augmenting Teradata with SingleStore enabled this company to deliver real time insights by boosting data-ingestion rates to **12 million rows per second.** SingleStore significantly improved performance: delivering queries in less than 100ms and transforming day-old analytics into real-time insights for the executives. SingleStore's native connection to Tableau made it easy to populate the real-time dashboards via **MySQL wire protocol,** enabling a direct Tableau-to-SingleStore interface.

4B+
Rows of new data ingested daily

100ms
query response / for 150K+queries per second

Results

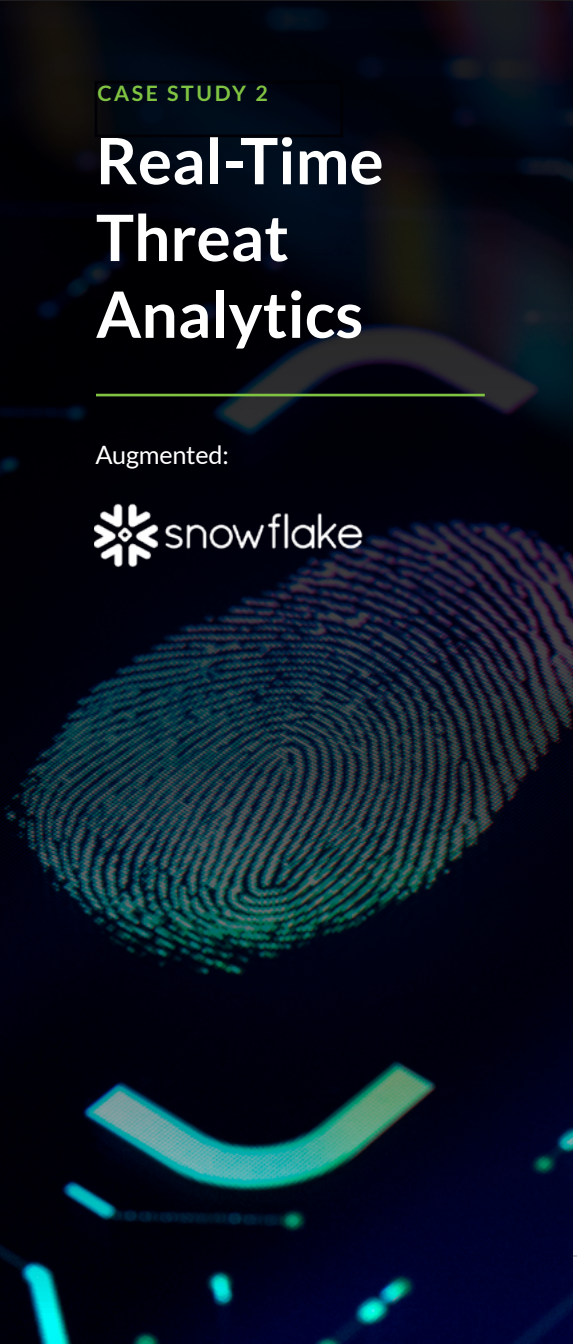
- Executives obtain operational insights to sales and market movements **in near real-time**—no more "flying blind"
- The architecture can cost effectively scale out to support more than **4 billion** new rows of data per day
- Queries are returned in less than **100 milliseconds** to enable fastboards
- The data warehouse can now deliver consistent performance, even with high concurrencies of more than **160,000** queries per second

CASE STUDY 2

Real-Time Threat Analytics

Augmented:

snowflake



LEADING CYBERSECURITY ORGANIZATION

Situation

Every millisecond counts when you are tasked with monitoring and reporting on potential security breaches, malware attacks, and other threats to network security. This organization depended on Snowflake as the data warehouse to power threat analytics and reporting of cybersecurity incidents.

Challenge

There was a significant lag between the time when a potential threat was detected to when the incident was reported--sometimes as long as three to five minute delays—eroding this firm's competitive position in the market.

Technically, this latency was driven by a combination of factors including difficulty supporting a growing volume of queries and issues with streaming ingestion. With concurrent loads of 1,000 queries per second, Snowflake just couldn't keep up.

Solution

Since augmenting Snowflake with SingleStore, the cybersecurity team has been able to dramatically **reduce the time it takes to report on and analyze threats.** SingleStore ensured real-time streaming ingestion from Amazon S3, together with less than 500ms latency for all queries--even with thousands of users concurrently accessing the application.

SingleStore

15x
improvement in speed of ingestion

100x
improvement in time to report on new data

Results

- Customers receive threat-detection alerts and reporting in **less than one second** versus approximately three minutes before
- **180x** improvement in time to report on new threats, improving the customer experience
- Reduced data-ingestion latency by **15x** for millions of records
- Less than **500ms** latency for all queries, even with more than **1,000** concurrent users

CASE STUDY 3

Media Company Boosts Ad Sales with Fast Dashboards

Augmented:

LEADING NORTH AMERICAN MEDIA CONGLOMERATE

Situation

More than 100 sales reps at this large North American media company depend on a Looker dashboard to understand ad inventory and performance in order to sell ad slots to customers. Unfortunately, the Amazon RedShift data warehouse that powered the dashboard was too slow to process transactions and display results, leading to delays of as much as two hours between when ads were sold and when they were reflected in the dashboard.

Challenge

It took an average of two hours to ingest new data from Amazon S3 into Redshift. Furthermore, because hundreds of sales reps were accessing the same dashboard at the same time, it took more than 5 minutes to return queries when the dashboard was filtered or refreshed. Ad executives inadvertently found themselves closing deals for ad spots that had already been sold by their colleagues. With ads accounting for 32 percent of total revenue, this problem was not only damaging customer relationships, but also negatively impacting the bottom line.

Solution

Augmenting RedShift with SingleStore enabled the media company to continuously ingest new records from S3 in less than two seconds. Query response times have improved in tandem: ad execs can refresh their dashboards in less than one second, as opposed to five minutes before.

99%
improvement in
speed of ingestion

300x
improvement
in query latencies

Results

- **Fast, interactive dashboard** for sales reps, with real-time data updates to enable new sales
- **300x** improvement in query latencies: Less than **1 second** latency for dashboard updates, versus 5 minutes with RedShift
- Data ingested in less than **2 seconds**, as opposed to 2 hours with RedShift
- Supports **1,000+** users concurrently with no performance degradation
- Measurable increases in ad sales and effectively **zero** double-booked ad spots

The Value of Data Warehouse Augmentation

Is your organization stymied by an outdated data warehouse architecture? Not sure?

Ask yourself these questions:

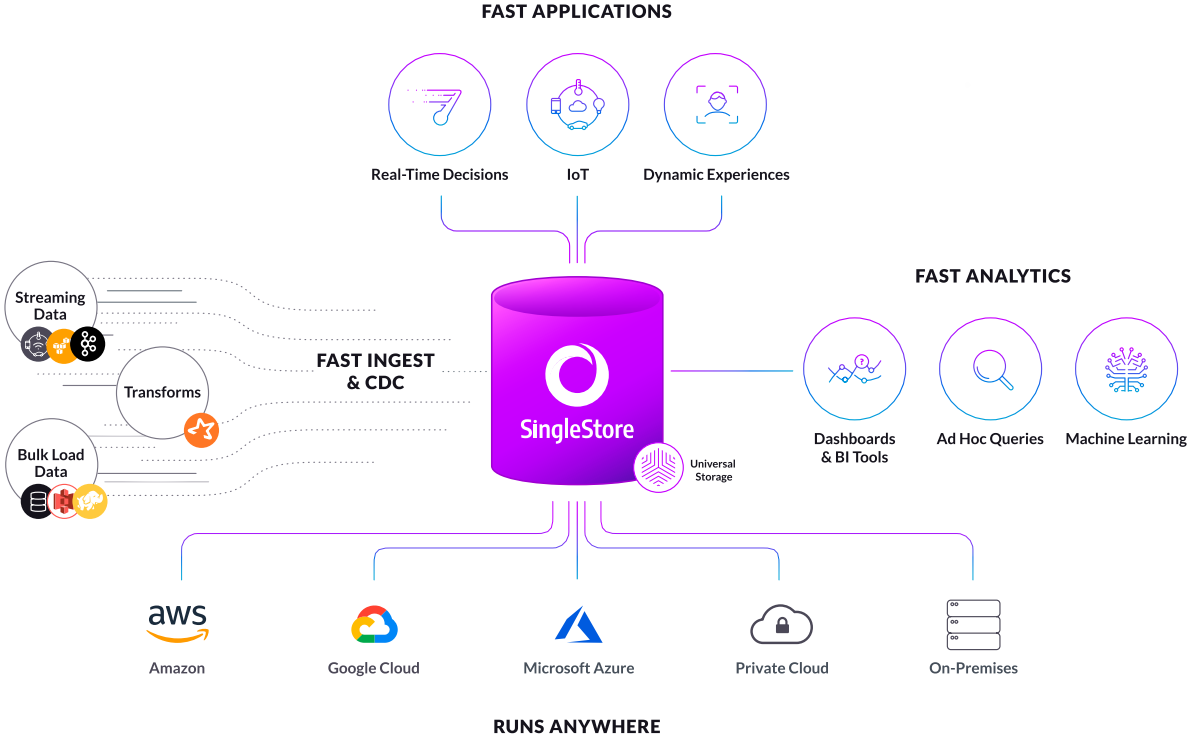
- Do you struggle with stale or slow-running dashboards or applications that don't reflect the most up-to-date information?
- Are you struggling with customer experience, performance issues, or escalating costs with your data warehouse environments?
- Are you trying to break down the barriers of slow batch processes or do you wish to accelerate your time-to-insights?
- Are you trying to move towards real-time or near-real-time insights or use cases?
- As you scale analytic systems to keep up with escalating data volumes and rising customer demands, do you have to approve large capital outlays to upgrade hardware and software infrastructure, or incur excessive usage charges from cloud providers?
- Do you face diminishing user-acceptance as people grow impatient with their inability to seize data-driven opportunities or keep up with burgeoning data processing demands?

If the answer is yes to any of these questions, it may be time to consider **augmenting** your data warehouse **with SingleStore.**

SingleStore Delivers

With **20x to 100x the performance at 1/3 the cost** compared to legacy infrastructure, SingleStore delivers the speed, scale, and agility in one powerfully simple, cloud-native, relational database, helping you to drive analytics and insights fast, and in the moment!

And with SingleStore Managed Service, the fully-managed, on-demand cloud database service you can get started in just a few clicks - on any cloud of your choice. Test drive now.



SingleStore

SingleStore

SingleStore Managed Service gives you the full capabilities of SingleStore on any public cloud without the operational overhead and complexity of managing it yourself.

Get Started Today
with $500 in Free Credits

About SingleStore

SingleStore offers a single unified database for your data-intensive applications. Its cloud-native, massively scalable architecture provides super fast ingest and query performance with high concurrency--the ideal architecture to power your data-intensive applications and dashboards.

SingleStore can ingest millions of events per second with ACID transactions while simultaneously analyzing billions of rows of data, all with the familiarity and ease of using SQL. It can handle both OLTP and OLAP workloads in a single system, which fits with the direction of new applications that combine transactional and analytical requirements.

With 20x to 100x the performance at one third the cost of traditional databases, SingleStore delivers speed, scale, and agility in one powerfully simple, cloud-native, relational database, helping you to drive analytics and insights fast.