SingleStore

EBOOK

3 Key Attributes of a Modern Application That's Ready for Real Time

Table of Contents

Introduction

Real-time applications and analytics. As a consumer, you benefit from both. Probably multiple times a day.

Every time you

- find a new show based on a recommendation that your streaming service has given
- use your credit card at a new store and get a fraud alert from your bank verifying that you're the one making that purchase
- play a game online and expect an immediate response and tailored experience
- use a ride-sharing application to get from one place to the other

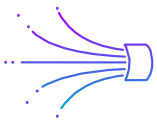
...your experience is powered by real-time applications and analytics.

Behind the scenes, these daily touchpoints online and in the marketplace require a powerful database architecture to work as they should.

But let's be real — as a consumer, you don't care how the experience is powered. You just want it to work. If the application takes too long to load or gives you recommendations that don't fit with your taste, you're less likely to use it next time. You'll just take your business elsewhere.

You know what you expect when you're using an application. So, how can you meet your own standards when you're the one providing the application or service?

In this eBook, we'll explore three key features that your database architecture needs to have to use live data that will make your application competitive: ultra-fast ingest, high concurrency with low latency and AI/ML capabilities.



Ultra-Fast Ingest

The Challenge

If only one user is using your application, you might not have a need for speedy ingest (you might have bigger business problems, but that's a subject for another eBook).

But if you're a large or growing business, your data inputs likely are growing exponentially. And your database architecture must be able to keep up. Data needs to be continuously ingested from diverse sources **as it's generated**, and be immediately available for indexing and querying. Batch data loading or ETL isn't good enough.

Stale data does you no good in today's marketplace.

Would **you** use an application that sends you recommendations based on your old preferences, or alerts you to potential fraud hours or days after the fact?

The Solution

Clearly, fast ingest is critical. But it can be tough. It's no secret that loading data into a database can be an arduous, time-consuming process.

But SingleStoreDB Pipelines makes it easy.

Pipelines are built into SingleStoreDB, and are a major differentiator from other database vendors. They allow users to continuously extract, optionally transform and load data in parallel at ultrafast speeds by using:

- Parallel loading, using the full power of a compute workspace
- Support for real-time streaming from files, cloud object storage and Apache Kafka
- Real-time de-duplication of data
- Support for several data formats such as CSV, JSON, Avro and Parquet
- Support for several data sources including Amazon S3, Azure Blob storage, Google Cloud Storage and HDFS

With SingleStoreDB Pipelines, you can load massive amounts of data fast. We're talking 100-billion-rows-in-10-minutes kind of fast.

Here's How it Works

SingleStoreDB tested this claim. You can see a step-by-step process here, but we've laid out the key steps below:

1. Create a Pipeline using the SingleStoreDB Portal or the command line. Specify name, directory, location of the bucket folder and table name where data will be loaded, specifying the data format delimiter.
2. Start the Pipeline with a straightforward command.
3. Monitor the status of the Pipeline by checking the real-time performance from SingleStoreDB Portal. You can verify a Pipeline status in SingleStoreDB Portal by clicking Ingest > Pipelines located on the left-hand navigation, selecting your workspace, then checking the "State" column.
4. Verify the loaded data. The SQL statement generated will return the total numbers of rows loaded and the duration in minutes.

In this example, our engineer was able to **query 100 billion rows of data** and **return 10 rows of aggregated results in 10.93 minutes**.

Think your application could use that kind of power?

High Concurrency

The Challenge

Let's say thousands of people are using your application at once. And at the same time, your team is querying the database to inform strategy and next steps. If your database can't handle this high concurrency, just the thought of this might stress you out.

Because the hard truth is that your customers won't wait in line to use your application. They want it to work when they need it—and if it doesn't, they'll take their business elsewhere.

If your application isn't available, you've lost customers — maybe for good.

To be competitive, your application must be able to handle tens of thousands of users and queries, and serve up results quickly.

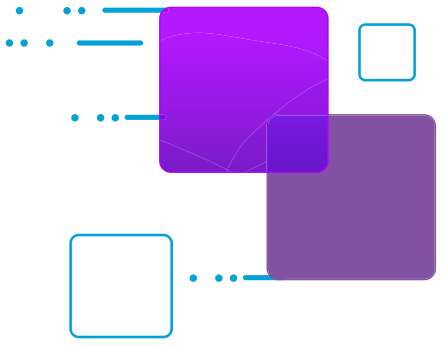
The Solution

SingleStoreDB delivers low-latency performance even when many people are accessing the application at the same time. In fact, our solution supports customer-facing applications with 40,000+ concurrent users.

How? Seekable columnstores. These allow highly concurrent read/write access, as well as subsegment access, a method for reading small parts of columnstore columns independently and efficiently.

To run selective queries fast, SingleStoreDB incorporates hash indexes on columnstores, which can be created on any individual column in a columnstore table. In addition, SingleStoreDB columnstore tables are broken into one-million-row segments, and within each segment, columns are stored independently into column segments. Subsegment access speeds up access to a row in a columnstore by allowing the system to calculate the location of the data for that row, reading only the portion of the column segment that is needed to materialize that row.

This very technical explanation boils down to a simple point: With less information to sift through, the database can answer queries faster and has more capacity to handle multiple users and queries at once.



Here's How it Works

Let's take a look at a real-life example from a household name: General Electric.

GE's vision for the future of business and technology was hampered by an overly complex data infrastructure that made it impossible to support every team's needs. The company sought a solution that would enable it to:

- Bring data together from disparate systems
- Support citizen development and self-service to empower business users to obtain their own answers and resolve their own issues without having to create a bottleneck by bringing every request to IT and data management teams
- Drive large-scale transformation programs in pursuit of multi-million-dollar benefits

Piece of cake, right?

With SingleStoreDB, it was.

SingleStoreDB enables users to unify transactions and analytics together in a single data platform that ingests millions of events per second while simultaneously analyzing billions of rows of data.

Today, SingleStoreDB works with streaming, real-time and historical data from across GE to surface the insights that matter most to teams. It also readily integrates with many BI tools and built-in data pipelines. This makes it easy for GE to implement ad hoc analysis in response to evolving business and market conditions.

With SingleStoreDB, GE reduced the complexity of its data fabric, gave business users the self-serve, real-time data they needed and reaped massive cost savings. They're seeing:

40,000+
concurrent users

100+
use cases solved with a single platform

4M+
transactions per minute

40X
cost savings

With SingleStoreDB, they've got things under control.

AI/ML Ready

The Challenge

The final feature that helps make your application competitive in today's marketplace is the ability to incorporate Artificial Intelligence and Machine Learning (AI/ML). AI/ML can lower costs, increase scale, speed up the delivery of results and help you achieve critical business goals. Here are just two of the benefits that AI/ML can deliver:

Let's face it — if you haven't already started incorporating AI/ML, you might be behind the curve.

Real-time analytics: This approach leverages one of the greatest advantages of AI: the ability to process large sets of data and interpret the results in real time. Businesses can benefit from making the right decisions based on real-time insights, helping the company remain competitive in the market.

Chart-Line Predictive analytics: AI/ML technologies are capable of handling large sets of data, identifying the patterns in the data set and making future predictions. For example, predictive analytics can help forecast which customer base has the highest probability of buying your product, then send targeted campaigns, discounts, coupons, etc. to the appropriate sets of people to expand the revenue pipeline. ML can also help in delivering highly efficient predictive maintenance plans for machines and connected devices.

The Solution

But AI and ML can be difficult to incorporate. Think it's worth it?

It absolutely is, if you have SingleStoreDB in your corner.

The data infrastructure behind AI/ML applications requires a high-performance platform to complete the calculations involved in the model, usually across a mix of streaming and historical data. The good news: SingleStoreDB was built to accomplish just that; a cloud-native, operational database built for speed and scale.

As a distributed, highly scalable SQL database, SingleStoreDB delivers maximum performance for both transactional (OLTP) and analytical (OLAP) workloads, using familiar relational data structures. SingleStoreDB also comprises an ultra-fast ingest and query platform that enables real-time model scoring on both streaming and historical data.

With real-time access to fresh data, the ability to stream data from multiple sources and low-latency response times that deliver instantaneous user experiences, SingleStoreDB is a great solution for operationalizing your ML models in production (check out the winner for this category in the SingleStore Hackathon). The SQL compatibility of SingleStoreDB makes it a solid choice as a platform for developers operationalizing ML models as part of user facing or internal applications that allow users to benefit from ML predictions.

The following key capabilities make SingleStoreDB a great solution to operationalize and scale AI/ML applications.

Fast, scalable data exploration

SingleStoreDB centralizes all operational data with built-in pipelines, while performing ultra-fast queries to help in identifying new models.

Fast analysis and scoring

SingleStoreDB is capable of discovering anomalies or predicting events as they happen by combining real-time and historical sources, and querying them with standard SQL.

Proven compatibility

SingleStoreDB leverages the familiarity of ANSI SQL with full data persistence to drive sophisticated analytics while seamlessly working with existing business intelligence, data integration and middleware tools.

Unlimited scale

SingleStoreDB has a shared-nothing architecture for scale-out and scale-up, using standard hardware. It can be deployed on-premises, in the cloud, as a service or as a hybrid cloud.

Here's How it Works

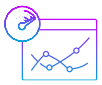
EPIGEN

Epigen Technology uses software to innovate in many fast-changing areas of technology, including cybersecurity, analytics, machine learning, AI and the emerging area of cognitive computing. Epigen's clients in areas such as the military, law enforcement and air transportation need facial recognition that works under increasingly challenging circumstances, including working at scale and using video frame captures — not just static images. To support this, Epigen is developing a scalable architecture for facial recognition, which is a solid use case of AI. A SingleStoreDB-based data platform will serve as input to a cognitive computing application for facial recognition.

With this facial recognition application, Epigen is creating a reference architecture for cognitive computing. Data is initially read into an Amazon S3 database in the cloud. Epigen then uses a SingleStoreDB Pipeline to extract the data from S3 at top speed, rapidly transforming and loading it into SingleStoreDB. From there, the data goes to the GPU array.

This design goes beyond many of today's AI and machine learning tools to make decisions, and offers support for decision-making by humans. For instance, a narrowly targeted AI tool might be able to show that someone boarding an airplane is not the same person as shown on their passport photo. With cognitive computing, the live photo would then be matched against law enforcement databases, with possible matches identified, and each possible match assigned a probability. A tool like Graphistry would then be used to present a complete picture of the information uncovered to security personnel for them to act on.

Epigen's facial recognition platform takes full advantage of several SingleStoreDB capabilities to solve this business problem using AI, with significant cost savings:



Rapid data ingest: Rapidly bringing in photographic data from Amazon S3 and, in the future, from other sources.

Fast CPU-based processing: SingleStoreDB's ability to handle the "transform" part of an ETL operation within a pipeline is crucial to rapid processing, as is SingleStoreDB's Pipeline to stored procedures capability when more complex processing is required.

Fast GPU-based processing: A GPU (Graphics Processing Unit) is a processor designed to handle graphics operations. This includes both 2D and 3D calculations, though GPUs were originally designed for rendering 3D graphics. The GPU is used by the application and in ML training. SingleStoreDB does not run on the GPU. Instead, it helps in rapid transfer of processed data into and out of the GPUs, which is crucial to the performance of the entire system.

Highly scalable: SingleStoreDB was proven to be a scalable solution at every point in this process, delivering a 4-5x increase in performance, higher total throughput and the ability to flow data into the GPU.

Conclusion

In today's hyper-competitive, high-intensity marketplace, you need your application to provide the best possible experience for your customers. This means that it needs to be available all the time, understand (and anticipate) users' preferences and give them a custom experience. Otherwise, they'll take their business elsewhere.

SingleStoreDB is the real-time, distributed SQL database that's fast, unified and resilient — empowering you to respond to your customers' needs. With ultra-fast ingest, high concurrency and the ability to help you incorporate AI and ML easily, it's the database architecture for today and the future.

Sounds pretty great, right?

Ready to see what SingleStoreDB can do for you?

Visit www.singlestore.com/cloud-trial/ to sign up for a free trial of SingleStoreDB Cloud — delivering the full capability of a real-time distributed SQL database without the operational overhead and complexity of managing it yourself.