

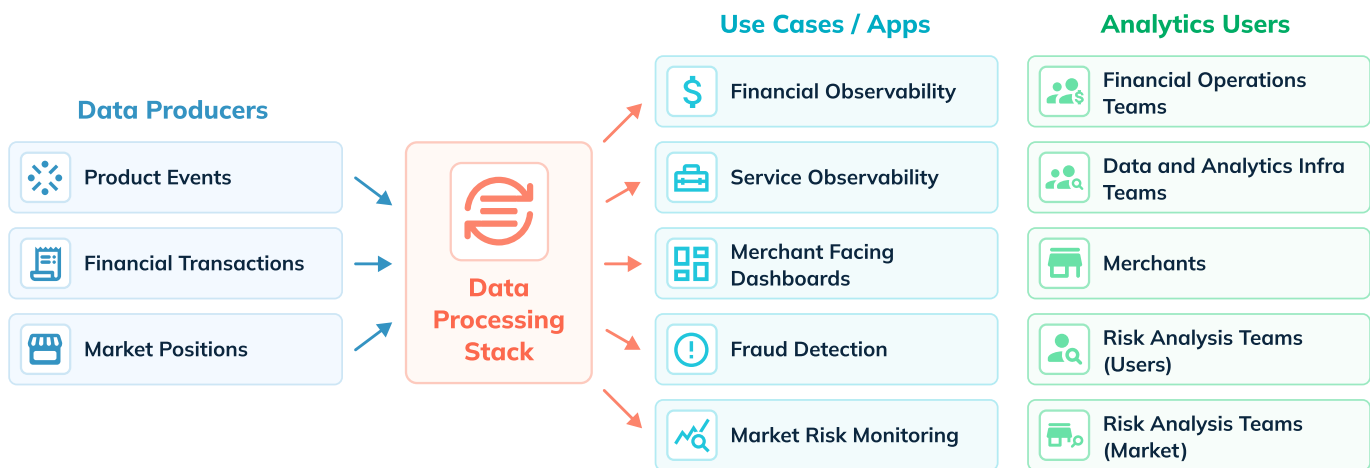


CASE STUDY

Delivering Analytics Infra at Scale to FinTech with StarTree Cloud Powered by Apache Pinot™

The world is becoming more digitally connected every day. Processes, applications, businesses, and people are able to collaborate like never before at a global scale. For growing FinTech firms, this means delivering financial services across many products, countries, currencies, banking institutions, and platforms. In this heterogeneous ecosystem, external users (e.g., merchants, customers) and internal teams (e.g., analysts, risk) using financial services expect fast, fresh, and accurate analytics from both real-time events and historical data. StarTree Cloud powered by Apache Pinot is at the heart of modern data architectures to deliver real-time analytics at scale, helping FinTech companies democratize the value of data.

FinTech Use Cases



1. Financial Observability

FinTech companies generate massive amounts of data - in the order of 100s of billions of transaction records to date, pushing boundaries of the analytics infra. A key requirement for internal finance operations teams is the ability to identify and triage transaction-related issues from the petabyte-scale data in real-time. For example, they need to identify incomplete payments across all merchants and perform interactive drill-down analysis to investigate these transactions. Having global financial observability using analytics across this massive dataset helps FinTech companies reduce risks and save on operating costs and overhead of doing this manually.

2. Service Observability

High service availability is crucial for FinTech companies to meet their SLAs and deliver a superior user experience. For data and analytics infra teams, fine-grained representation of how the billions of transactions flow through the service components and microservices becomes increasingly critical. Specifically, they need a way to trace financial events as they get processed through the system's various layers and identify real-time issues in any of the layers - such as network bottlenecks, delays, or failures. In addition, they also need the ability to root cause such issues by analyzing analytics of the various system parameters and fixing them quickly.

3. Merchant-Facing Dashboards

Merchant-specific dashboards empower FinTech merchant customers to track day-to-day growth, customer, and financial metrics. Merchants can analyze transaction volumes and slice and dice them across different dimensions such as credit/debit cards, net banking, digital wallets, and more. Other examples include dashboards showing gross volume, balance, payouts, and customer growth over a specified period of time -both real-time and historical. These dashboards can also be monetized by FinTech companies by charging a premium for access to these insights.

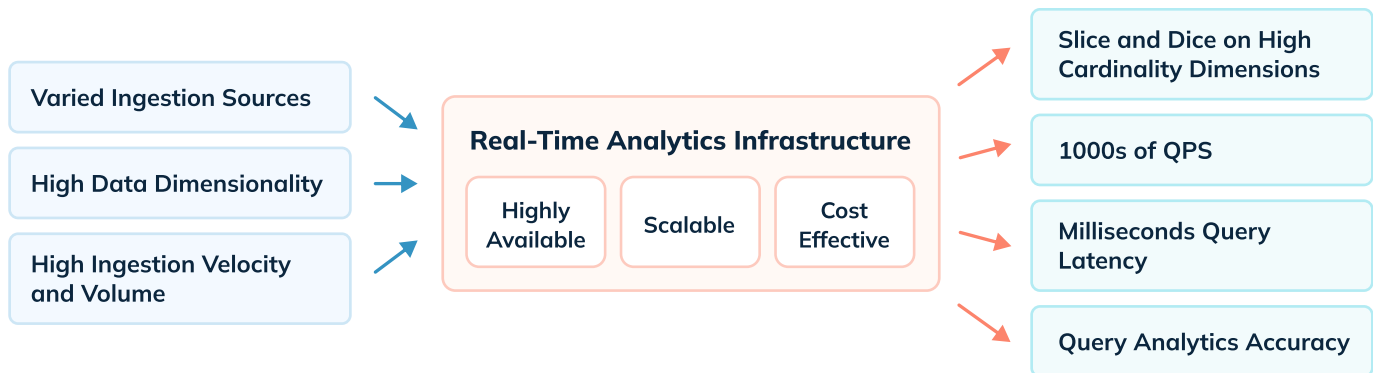
4. Fraud Detection

A common scenario across the FinTech industry is when companies generate features such as total payments processed or unique credit card merchants across different time buckets (minutes/hours/day/week/month). These features are utilized by algorithms built for real-time fraud detection. The underlying analytics infrastructure needs to be able to produce these features in a reliable manner with a high degree of accuracy in order to increase effectiveness and prevent fraud in real time.

5. Market risk monitoring

This helps FinTech customers measure risks associated with their portfolio and make appropriate decisions in real time. Analysts are looking to analyze data as it is generated and combine it with historical data to calculate overall risk and make decisions that impact the bottom line significantly. Only a few years ago, it was sufficient to calculate risk once a day after market close, but now it is imperative to do this calculation as often as possible to minimize risk. A key requirement for this use case is to be able to ingest data in real-time into a data store that not only supports updating the records (upserts) but also supports low latency queries involving complex aggregations.

Challenges



- Query performance: This relates both in terms of latency (query response) as well as throughput (queries per second). Even though it's hard to guarantee query SLA, the underlying analytics infra needs to deliver query results with sub-second latency even with varied query complexity. It needs to continue to deliver that performance as query volume rises per second.
- Number precision: When dealing with financial records, a system that delivers analytics must be able to handle arithmetic operations on fixed precision numbers used for denoting amounts. For instance, it must support 6-decimal accuracy for relatively large numbers.
- Data volume: Most organizations deal with massive amounts of data, and it isn't unusual for a FinTech company to have petabyte scale. It is essential to keep that

much data in a query-ready state for analytics without linearly increasing the infra footprint or cost to serve.

- **Ingestion:** As event streaming architectures are becoming mainstream, data needs to be ingested into the analytics infra from multiple and different sources (e.g., Confluent Cloud, AWS S3, Snowflake, etc.). Additionally, ingestion from these sources could require updates to the data pipeline to make data ready for analytics.
- **Data Freshness:** For some use cases, there is a need to query the data as soon as it's generated. In some cases, the freshness SLA is in the order of < 30 seconds. This is quite challenging, given the volume of real-time data generated.

The use cases discussed are all business critical and require addressing the above challenges with an analytics infra component that is highly available and reliable.

Traditional Approaches

Although traditional data warehousing solutions like Snowflake are great at crunching big data, they don't meet the requirements of the use cases above. It is very challenging to ingest data quickly into these systems and achieve a freshness SLA of < 30 seconds. These systems are not optimized for user-facing analytical use cases. Executing queries on the order of 10K+ or 100K+ queries per second is either impossible or prohibitively expensive on these systems. Similarly, it is hard to achieve sub-second query latency for OLAP queries of different shapes on warehouses. An OLAP datastore that is purpose-built to address these requirements is needed.

Solution

Leading FinTech companies like Stripe, WePay (a JPMorgan Chase company), Citi, and others rely on Apache Pinot to solve these challenges. Pinot can easily ingest data from a wide variety of data sources and make it available for querying in a high throughput, low latency, and cost-efficient manner. Here are ways in which Pinot makes that possible at FinTech companies.

Query Powerhouse

Apache Pinot was purpose-built for powering really high QPS OLAP queries (think of 100s of 1000s of queries per second) while still maintaining query latency in milliseconds.

- It has out-of-the-box support for a variety of indexing strategies that accelerate aggregation queries and minimize the number of data scans.
- Pinot's StarTree index is especially useful for powering user-facing analytics by intelligently pre-aggregating metrics without compromising on storage cost.
- Data in Pinot can be sorted, partitioned, and replicated, enhancing query

performance.

- Pinot supports BigDecimal as a primitive data type for storing fixed precision numbers. This is extremely useful for processing financial records, as mentioned above.

As a result, these companies have seen query latency drop dramatically. In one case, real-time reporting latency went from 30 minutes to 2 seconds while handling 200k events per minute. In another deployment, Apache Pinot reduced latency from 30 seconds to 300 milliseconds, even for accounts with billions of transactions.

Low Cost to Serve

As a distributed OLAP data store, Pinot architecture has different components, including broker, controller, server, minion, and zookeeper, that can each be scaled independently.

- It is designed to be a highly cost-efficient analytics engine with its columnar storage layout and various encoding techniques (fixed bit length/run-length).
- Users can configure Pinot clusters with different storage tiers on the servers, which allows for balancing query performance against cost of serving.
- Pinot has excellent support for multi-tenancy and is able to adhere to strict production availability and reliability SLAs in companies like LinkedIn and Uber, where Pinot SLA violation has a direct impact on the company revenue.

By choosing Pinot as the analytics infra component, one FinTech organization reduced its cost to serve analytics by roughly 50%, or over \$1M.

Simplified Ingestion

Apache Pinot has excellent support for ingesting data from a wide variety of sources, such as Streaming sources (Kafka, Kinesis, PubSub, Pulsar), Batch sources (S3, GCS, ADLS, HDFS), and OLTP changelog. The ingestion latency is typically in seconds, which means an event can be queried as soon as it is generated. It supports exactly once data consistency for real-time sources such as Kafka and has excellent support for bootstrapping and backfilling the data.

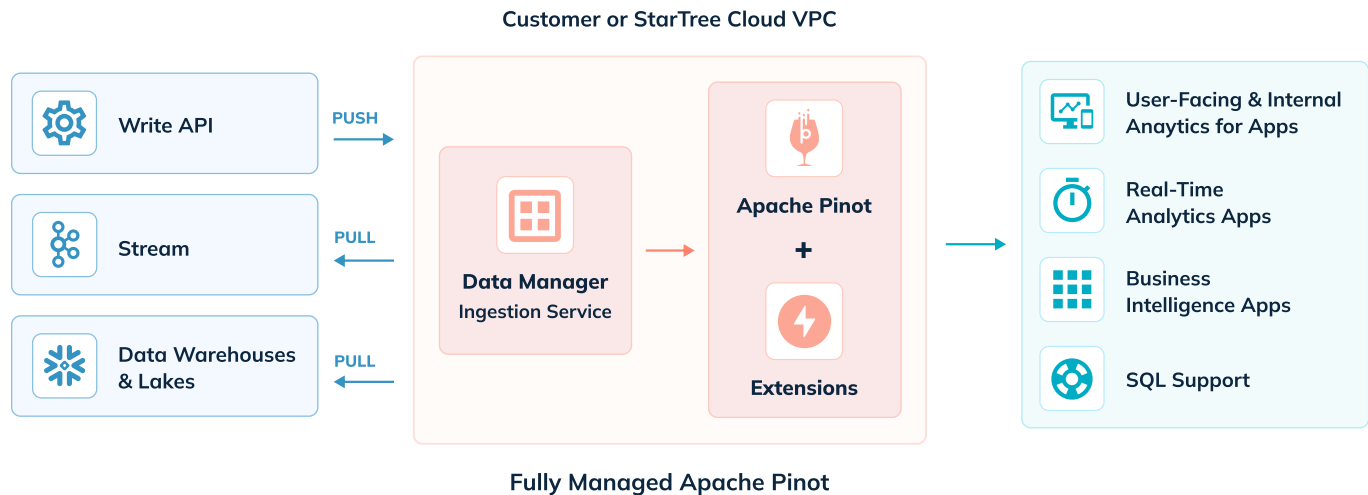
StarTree Cloud

StarTree Cloud offers two deployment options for a fully managed Apache Pinot service - Bring Your Own Cloud (BYOC) and SaaS, which give customers an option wherever they are on their data and real-time analytics journey.

The BYOC Edition of StarTree Cloud delivers secure, enterprise-ready Apache Pinot clusters with StarTree's services and apps – within the comfort of a customer's cloud account where the data resides – all deployed, configured, and managed remotely by StarTree. This enables companies to deliver analytics by securely keeping data in their cloud environment while also benefiting from having a managed service that

reduces the complexity often associated with creating and operating a data cluster within their cloud environments.

Self-Service Edition of StarTree Cloud enables developers worldwide to ingest data into Apache Pinot cluster in just a few clicks and integrate with their apps to deliver real-time analytics with millisecond response times and high throughput.



Get Started with StarTree Cloud Free Trial

We offer a 30-day StarTree Cloud free trial so that you can take it for a spin with only an email address. The trial includes a self-service provisioned Apache Pinot cluster that can support up to 50 QPS, StarTree Data Manager for no-code data ingestion, and a set of Pinot APIs to integrate with data apps and external BI tools.

Sign up for StarTree Self-Service Cloud [here](#) and join the StarTree community slack [here](#)!

Benefits of Pinot for FinTech Companies

- Superior Query Performance** (Low latency)
- Fresh, Accurate Data Analytics**
- Scalable Performance with Growing Datasets and Queries**
- Lower Costs from Greater Efficiency**
- Ingests from Varied Data Sources**

There are numerous competitive advantages to be gained by analyzing data instantly at scale. A FinTech company with a trillion rows of data can perform hundreds of thousands of queries per second without sacrificing precision. Investing in the analytics infrastructure to enable real-time calculations prepares FinTech companies for present success and future growth.