



STARTREE BRIEF: PINOT POWERS MULTIPLE USE CASES

Apache Pinot - Versatility For Real-Time Analytics Use Cases

Few solutions can match Apache Pinot for the high throughput, low latency, and high accuracy required to support all your real-time analytics use cases at scale.

As enterprises across industries seek to unlock the advantages of real-time analytics, they have a diverse set of use cases in mind. They may be pursuing faster decision-making, for example, around operational intelligence, fraud detection, and customer incentives, or they may be looking to enable personalization as a means of driving growth and improving retention. They may want to leverage real-time business metrics, or go deeper with ad-hoc analytics. Or they may want access to all of the above and more.

A diverse set of real-time use cases presents daunting architectural requirements for an online analytics processing (OLAP) platform. To begin with, the solution must be able to ingest data from a wide variety of sources in real time and immediately make that data available for querying. Very high throughput and low query latency are mandatory, but the solution must also be capable of maintaining high query accuracy — via data deduplication techniques and real-time updates.

Purpose-Built for Real-Time Analytics

Apache Pinot's flexible, powerful architecture is uniquely suited to real-time analytics workloads. Capable of powering upwards of 100k queries per second with a p99th latency of 100ms, Pinot delivers high query accuracy and high performance regardless of query complexity. These and other native integrations and capabilities make Pinot ideal for all of the following real-time analytics use cases — and many more.

1. User-Facing Analytics

A critical use case for any business, user-facing analytics provides a more personalized experience for customers via real-time insights. Good examples include feedback on a LinkedIn user's profile and posts, or real-time analytics on sales, ordering, and inventory for an Uber restaurant manager. Successful user-facing analytics can drive engagement and growth, while broken SLAs can result in churn or negatively impact revenue. At scale, user-facing analytics can produce 10s or 100s of thousands of queries per second (QPS) and requires a 100ms p99th query latency SLA.

2. Personalization

A subset of user-facing analytics, personalized recommendations, or actions based on user activity seeks to deepen the product experience for each user by presenting product features and benefits in real-time. Think of a personalized LinkedIn news feed, for instance, or an Uber driver receiving real-time financial incentives based on predicted needs. As with user-facing analytics, personalization demands high throughput and low query latency.

3. Business Metrics

Real-time business metrics enable critical enterprise functions such as operational intelligence, anomaly and fraud detection, financial planning, and more. As an example, a company like LinkedIn needs to track a metric like “page views” in real time in order to detect and resolve operational issues as fast as possible. For the analytics platform, this use case demands not just consistently high QPS and low latency, but also a high degree of query accuracy — and that means the platform must be able to handle duplicates or upserts across a variety of data sources.

4. Anomaly Detection and Root-Cause Analysis

Beyond business metrics, organizations need to be able to detect anomalies on large time-series datasets instantly. If LinkedIn’s “page views” metric shows an unexpected drop week over week, teams need to understand which dimensions were primarily responsible for causing the anomaly so they can take corrective action immediately. Anomaly detection and root-cause analysis require an OLAP platform capable of performing high QPS temporal scans and “group by” queries.

5. Visualization

For all the above insights, enterprises need an effective means of visualizing data. From simple dashboard charts to complex geospatial representation, clustering, trend analysis, and more, data visualization requires the OLAP platform to provide seamless integrations to standard visualization solutions such as Apache Superset and Grafana.

6. Ad-Hoc Analytics

Organizations also want access to real-time data exploration in order to perform issue debugging and pattern detection on streaming events. In practice, this might look like a data scientist at Uber identifying order delays in real time, or joining specific real-time events with offline data sets. These tasks require the underlying OLAP platform not only to remain performant with complex queries, but to support ANSI SQL compatible queries — not something typically built-in with OLAP solutions.

7. Log Analytics / Text Search

A less common but important use case, real-time text search queries on application log data present challenges because the log data can be very large and often includes unstructured (e.g., JSON format) data. However, this use case can be vital: organizations need to be able to perform real-time, regex-style text searches on logs in order to triage production issues, and certain applications require aggregation queries with text search predicates as part of their core business logic. For debugging use cases, QPS is typically low, but can go high for user-facing apps.

Want to learn more? [Talk to us](#) about leveraging real-time analytics in your organization.