acceldata

Customer Case Study

How PhonePe/Walmart Scaled By 2000% and Saved $5 Million With Acceldata

PhonePe

Case Study: PhonePe

Problem:

Scaling and performance issues on open-source OLTP and OLAP.

Solution:

Acceldata Pulse helped PhonePe monitor HBase, Spark, and Kafka to distinguish between infrastructure issues and seasonal and campaign-based anomalies.

Results

- Scaled data infrastructure by 2000% – 70-1500 nodes – to support massive business expansion

- Delivered 99.97% availability across its Big Data infrastructure

- Saved $5m+ in annual software licensing costs

- Improved engineering productivity by minimizing daily emergencies

- Large HBase OLTP Cluster

- Hadoop, HBase, Hive, Spark, Ranger, Kafka

- 400M cash Transactions/ month

- Open-source, HDP binaries

PhonePe is a Walmart subsidiary that provides more than 350 million consumers across India with the ability to send and receive money, make payments at more than ten million physical and online retail stores, use ATMs and invest in mutual funds and other securities.

PhonePe's Challenge:

PhonePe uses a variety of open source data technologies, such as Apache Hbase, HDFS, Kafka, Spark, and Spark Streaming, to run their high-volume, real-time payments and cash transfer platform. With hundreds of millions of customers and millions of merchants on the system, PhonePe's Data Warehouse cluster must be highly performant, reliable and transparent, which includes the ability to accurately report on system and business performance to internal and external stakeholders 24/7.

Scaling to Meet The Needs of Growing Data Infrastructure

As PhonePe's business grew explosively in 2018-19, the company embarked on a massive data infrastructure expansion in terms of both scale and new technologies. The company needed to increase the size of its Hadoop infrastructure to support tens of millions of new consumers and millions of new merchants who were rapidly adopting the service, all while adding Hive LLAP, Spark 3.x and Druid to the platform, technologies that were needed to support new products and business requirements.

Even in the early stages of this infrastructure expansion, the technology team experienced tremendous pressure on system performance and reliability. Key engineers spent the majority of their time firefighting problems and searching for causes behind data application issues and infrastructure failure instead of focusing on increasing scale and new capabilities as required by the business.

PhonePe's Chief Reliability Officer, Burzin Engineer, quickly realized that his team needed tools to improve visibility into every aspect of the company's data operations. WIthout more advanced tools that matched the sophistication of his core open source technologies, PhonePe's critical data initiative would fail, jeopardizing the company's growth prospects and business success.

The Acceldata Solution:

After gaining an understanding of PhonePe's objectives and challenges with Burzin Engineer and the PhonePe team, Acceldata demonstrated how its Pulse data observability tool could provide real-time monitoring of Hbase, Hive, and Spark data pipelines.

Acceldata Began Delivering Value in 24 Hours

The PhonePe team implemented Pulse in less than a day and immediately began to identify problems with HBase region servers and tables that were under pressure. Pulse helped PhonePe distinguish between HBase cluster issues caused by hardware or poorly designed tables and anomalies resulting from seasonal and campaign-related surges.

PhonePe had previously tried to use open-source and other commercially available tools, like HBase Console and Ambari in addition to building single metric grafana dashboards, for root cause analysis but found that they were insufficient. HBase Console, for example, only provided aggregated information and required significant time and analysis from highly experienced data engineers before it delivered useful intelligence. In contrast, Pulse directs users to the problem's root cause quickly and clearly through automated alerts and easy-to-read dashboards. In many cases, Pulse even recommends fixes to solve the problem.

"Acceldata supports our hyper-growth and helps us manage one of the world's largest instant payment systems. PhonePe's biggest-ever data infrastructure initiative would never have been possible without Acceldata."

Burzin Engineer
Founder & Chief Reliability Engineer

Results:

In the first 18 months of using Acceldata Pulse, PhonePe has been able to realize these, among other, benefits:

- ⊙ Scale data infrastructure rapidly from 70 to more than 1500 Hadoop nodes; more than 2000% growth

- ⊙ Deliver 99.97% availability across its Hadoop infrastructure

- ⊙ Eliminate day-to-day engineering involvement and firefighting on outages and performance degradation issues

- ⊙ Support multi-cluster data and workload management with uniform configurations

- ⊙ Upgrade systems and migrate to new applications and nodes with no performance degradation

- ⊙ Reduce data warehouse costs by 65%, while eliminating the need for expensive commercial data warehousing licenses

Multidimensional Data Observability

Enterprises are overwhelmed with the challenges of observing, operating, and optimizing large-scale data systems.

Multidimensional data observability can simplify modern data pipelines by monitoring and correlating data workload events across application, data, and infrastructure layers to resolve issues that break production analytics and AI workloads.

The right data observability tools can significantly improve enterprise data system performance, cost, and agility.

Try Acceldata - The Only Multidimensional Data Observability Cloud