

CASE STUDY

How PhonePe scaled by 2000% and saved \$5 million with Acceldata



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 10 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, Hive, Kafka, Spark, and HDFS, 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. This includes the ability to accurately report on system and business performance to internal and external stakeholders 24/7.

Problem:

Scaling and performance issues on open-source Hadoop environments.

Scaling Data Infrastructure to Meet the Needs of a Fast Growing Business

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.

Solution:

Acceldata's data observability platform helped PhonePe gain deep insights into their HBase, Spark, and Kafka environments to uncover and remediate infrastructure issues that were masked by normal seasonal and campaign-based activity.

- 350M users
- 400M cash transactions per month
- 1500+ nodes, 28+ HDP clusters, 20+ PB
- Open-source Hadoop, HBase, Hive, Spark, Kafka, and Ranger

Results:

-  Scaled from 70 to 1500+ nodes
-  Improved availability for Hadoop data lakes to 99.97%
-  Saved \$5M/year in software licensing costs
-  Reduced processing time by 55% Freed top developers from support & maintenance activity
-  Zero Sev-1 issues for 18 months and counting

Key engineers spent the majority of their time firefighting instead of focusing on increasing scale and adding new capabilities 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 to simplify the management and evolution of their data environment, 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 data observability platform could provide real-time insights into Hbase, Hive, Kafka and Spark.

Acceldata Began Delivering Value in 24 Hours

The PhonePe team implemented Acceldata in less than a day and immediately began to identify problems with HBase region servers and tables that were under pressure. Acceldata 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 information. In contrast, Acceldata directs users to the problem's root cause quickly and clearly through automated alerts and easy-to-read dashboards.

“

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, PhonePe was 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 firefighting and performance issues, freeing up engineers to focus on innovation
- Simplify and reduce support and administration time and effort
- Upgrade, migrate, and refactor systems and workloads with no performance degradation
- Reduce data warehouse costs by 65% by eliminating the need for commercial licenses

Multi-Layer Data Observability:

Enterprises are frequently challenged with managing and optimizing complex, large-scale data environments.

Multilayer data observability correlates information across infrastructure, platform, processing and data layers to identify and alert on trouble spots, bottlenecks, and inefficiencies. Analytics and recommendations simplify remediation and administration. In addition, Acceldata provides an extensible library of auto-actions to make systems self-healing and self-tuning. The right data observability tools can significantly improve the reliability, performance, scale, and cost of enterprise data environments.