

acceldata

Customer Case Study

How True Corporation Managed 400% Data Cluster Growth With Data Observability

true



Case Study: True Corporation

Problem:

Pervasive data system performance and scalability issues left 50% of ingested data unprocessed.

Solution:

Acceldata tools isolated bottlenecks, automated performance improvements, and distinguished between mandatory and unnecessary data to ensure 8PB+ data lake could reliably support all critical enterprise analytics requirements.

Results

- Reduced annual software licensing costs by \$2M+
- Improved existing system capacity and saved additional \$1M+ in projected capex.
- Eliminated unplanned outages and Sev 1 issues five consecutive months and running.
- Optimized HDFS storage cost by -2PB or 25%.

Company Overview

- ↳ One of Thailand's largest communications companies
- ↳ Operations in Thailand, Indonesia, and the Philippines
- ↳ Develops solutions for digital media, data analytics, and IoT
- ↳ Founded 1990

Data Environment

- ↳ 100+ nodes, 8+PB data lake cluster
- ↳ Uses Hortonworks Data Platform (Apache Hadoop, Hive, and Spark), Apache Ranger, Kafka~69K Streaming messages/second
- ↳ 500M + user impressions/month

True Corporation is one of the core businesses of True Group, an integrated telecommunications and digital service provider, based in Thailand. The company is a leader in the development and integration of artificial intelligence, big data, blockchain, cloud, Internet of Things (IoT), and robotics solutions. As the company expanded its business operations across Southeast Asia, they built a unique ecosystem of digital platforms and solutions to address the needs of consumers, merchants, and enterprises. The data operations team at True Corporation quickly discovered that they needed visibility into their data pipelines so they could scale growth and ensure the reliability of their data.

True Corporation's Challenge:

True Corporation runs a large environment of Hadoop clusters that power several aspects of the business including:

- **Measurement of Quality of Experience (QoE):** This is a measure of customer satisfaction. QoE can be service-specific, or a measure of overall satisfaction across all services (video, voice, text).
- **Abuse detection:** SIM boxes can be used fraudulently and in violation of the fair use policy of the SIM cards issued by the telecom operators
- **Monetization and upselling opportunities:** Accumulating customer behavior data across the network enables the Analytics

team to identify preferences that can be turned into product recommendations and opportunities for increased monetization.

As data volume grows exponentially, so do the operational challenges of managing this data and maintaining optimal analytics performance. In response, True Corporation's Analytics team deployed hundreds of nodes/servers across multiple clusters to address storage and processing requirements of the volume, velocity, and variety of data.

The company was in the process of setting up a new cluster with larger nodes, better specs, and more processing resources as more data was generated on the network. As part of a major migration process, the team had to onboard applications from the old cluster to the new one.

Once the flow of execution was changed to the new cluster, the team discovered that despite having more capacity, processing slowed dramatically. As a result, new business processes could not be supported.

The Analytics team couldn't identify issues with the hardware, resource manager configuration, or Spark application code. Performance issues in complex, interconnected data systems are difficult to isolate. Without multidimensional data observability, True Corporation couldn't resolve its data flow issues or achieve the pace and scale of data operations required to support analytics-based business goals.

The Acceldata Solution:

The data flow issue could have been attributed to one or multiple layers of the processing environment which would suggest a correlation between Yarn Metrics, Spark Metrics, Infrastructure, User, and Concurrency. The first step, therefore, was to figure out where the issues existed.

After evaluating True Corporation's environment, the Acceldata team identified specific points within data pipelines that were preventing a normal flow of data into new clusters. To address this, Acceldata Pulse was implemented which immediately enabled end-to-end observability into the True Corporation Analytics environment, along with additional data-related information in the Pulse dashboard that enabled the True Corporation team to improve data-related decision making.



"Acceldata's tools fixed our analytics pipeline issues, improved visibility into our data systems and recommended ways to scale and optimize our systems to meet future requirements. They helped True Corporation transition to open-source technologies, allowing us to reduce licensing costs, while delivering mission-critical analytics across the enterprise."

Wanlapa Linlawan

Head of Analytics Platform

Pulse categorically showed the path to figuring out the right OS parameters, the right Yarn parameters, the right OS parameters, fixed the calendar days on which the jobs should be run, the time of day that streaming jobs should be ideally executed.

After these improvements to the configuration, code changes, and setup were implemented, Pulse was then put to the task of continuous observation, alerting, and remediation of issues across Hadoop clusters and their data sources. Pulse delivered operational alerts which were domain-specific and related to use cases specific to True Corporation. These include cases where massive amounts of streaming data operate at high velocity and require an alignment in capacity in order to effectively process all of the data. Because these alerts are specific in what they address, and are unique to True Corporation's needs, the Analytics team has dramatically reduced the alert noise that used to deliver false positives and other erroneous, or useless, information.



Acceldata Pulse enables IT professionals and data engineers to gain greater visibility across hybrid data environments by connecting data sources, defining metrics, observing thresholds, and monitoring results from a customizable dashboard.

Results:

Within the first year of using Acceldata Pulse, True Corporation has been able to realize these, among other, benefits:

- ▶ Data environment increased from 35 to 200+ nodes.
- ▶ Clusters processing more than 2x data.
- ▶ Improved stability has allowed the Analytics team to support multiple lines of business.
- ▶ Eliminated engineering involvement in daily operational issues freeing resources to focus on expanding business use cases.
- ▶ Replaced expensive commercial solutions with open source technologies

Using Acceldata Pulse has enabled True Corporation to focus on key business use cases while optimizing the underlying data infrastructure.

The level of automation and cross-sectional visibility provided by Acceldata has allowed True Corporation to keep costs under control, move away from the predatory pricing of other OEMs which provide remote Hadoop support, and enabled the company to be self-reliant with their technology investments.

As a result, True Corporation has moved away from expensive commercial Hadoop Support, is capable of managing and maintaining its cluster environment, and has saved over \$1 million USD.

Learn More

Discover more about Acceldata at www.acceldata.io