

**CUSTOMER**

Lenovo

**INDUSTRY**

Mobile Communication

**USE CASE**

Analytics

Remote Data Access

**APPLICATION STACK**

Spark SQL + Hive +  
Alluxio + HDFS

**BENEFIT HIGHLIGHTS**

- 4x faster model processing time
- 10x faster data load time
- 95% reduction in cost of compute in the cloud
- Data is encrypted and secure
- No changes to applications or existing infrastructure

## Lenovo Analyzes Petabytes of Smartphone Data from Multiple Locations and Eliminates ETL with Alluxio

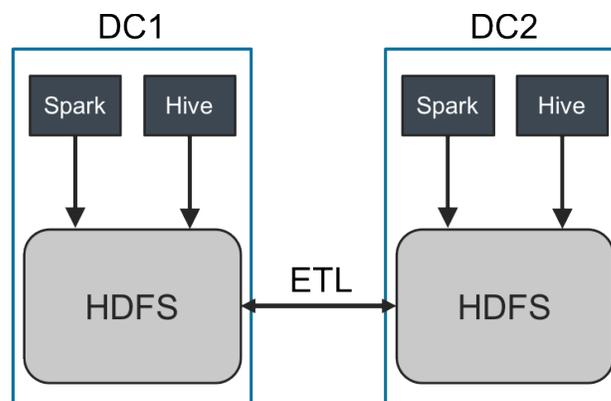
Lenovo is the world’s largest personal computer vendor and one of the world’s largest smartphone vendors. The company has invested extensively in global information technology infrastructure, including multiple data centers worldwide collecting petabytes of smartphone data. Analyzing data located in multiple data centers worldwide is critical for Lenovo to understand and improve the usability and reliability of their products. With Alluxio, Lenovo unified data from multiple data centers and eliminated the ETL process while lowering storage cost due to multiple data copies.

### The Challenge

Smartphone data is processed on the Lenovo Enterprise Analytic Platform. Previously this required a time-consuming and error-prone ETL process to transfer the data from multiple locations to a single data center for analysis. Lenovo uses big data technologies like HDFS to store the data and Hive metastore to store the metadata associated with the structured data. Analytics is performed using Hive and Spark SQL to gain insight into user behavior, popular applications, log analysis and more. The volume of data and number of geographic locations presented multiple challenges:

- High storage cost due to duplication of data
- Bandwidth and performance limitations transferring data from multiple locations
- Regulations preventing the transfer of certain data and excluding it from analysis

The following diagram shows the initial Lenovo infrastructure.

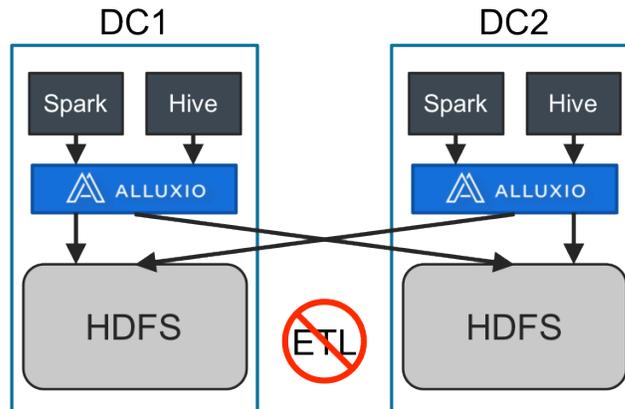


### The Solution

Lenovo addressed the technical challenges by using Alluxio as the data management layer for all smartphone data collected worldwide. The HDFS data stores were connected to Alluxio, providing seamless access for multiple applications through the global namespace. No changes to the application were

required. With this architecture, Lenovo performs advanced analytics involving cross data center data synchronization, joins and unions. Alluxio also temporarily stores the data in memory accelerating performance. Alluxio fits within existing security frameworks and enforces the policies in place, ensuring regulatory and compliance requirements from different countries and jurisdictions are met.

The following diagram shows the Lenovo infrastructure with Alluxio



## The Results

With this new architecture, Lenovo now has the infrastructure that allows them to analyze their worldwide data without the need for error prone, time consuming and costly ETL or the need for data duplication. Alluxio maintains the latest copy of the data in memory, or fetches it from HDFS for new requests, so data freshness is assured.

With Alluxio, the Enterprise Analytics Platform now stores data locally in memory from remote HDFS locations and provides transparent access for analytics applications. Alluxio presents the same API to the applications that they were already using. This allowed Lenovo to achieve the benefits without disrupting the existing stack or changing applications.

## Looking Forward

With Alluxio integrated in the data processing stack, Lenovo is now able to access and transform massive amounts of mobile data into valuable insights. This meets the business objectives of improved product quality and customer satisfaction at the lowest possible cost for their analytics platform.

## Stay Connected

**Twitter:** @Alluxio

**LinkedIn:** [linkedin.com/company/alluxio-inc/](https://www.linkedin.com/company/alluxio-inc/)

**Meetup:** [meetup.com/Alluxio/](https://www.meetup.com/Alluxio/)

**Slack:** [alluxio.io/slack](https://alluxio.io/slack)

## About Alluxio

Proven at global web scale in production for modern data services, Alluxio is the developer of open source data orchestration software for the cloud. Alluxio moves data closer to big data and machine learning compute frameworks in any cloud across clusters, regions, clouds and countries, providing memory-speed data access to files and objects. Intelligent data tiering and data management deliver consistent high performance to customers in financial services, high tech, retail and telecommunications. Alluxio is in production use today at seven out of the top ten internet companies. Venture-backed by Andreessen Horowitz and Seven Seas Partners, Alluxio was founded at UC Berkeley's AMPLab by the creators of the Tachyon open source project. For more information, contact [info@alluxio.com](mailto:info@alluxio.com) or follow us on LinkedIn, or Twitter.



1825 S. Grant Street | Ste 600  
San Mateo, California 94402

[www.alluxio.io](http://www.alluxio.io)