

CUSTOMER

TalkingData

INDUSTRY

Data Intelligence Solutions

USE CASE

Analytics

Data Unification

BENEFIT HIGHLIGHTS

- Unified heterogeneous storage in cloud, on-premise, and hybrid environments
- Self-service data access for analysts and data scientists
- Eliminated ETL and manual IT data management
- Enabled best of breed solutions from application frameworks and storage systems of their choosing

TalkingData Unifies Terabytes of Data for Analytics Across Disparate Data Sources

TalkingData, China's largest data broker, provides data intelligence solutions and processes over 20 terabytes of data and more than one billion session requests per day. TalkingData deployed Alluxio to unify disparate cloud, on-premise, and hybrid data sources for a range of analytics applications. The architecture provides self-service data access for data scientists and engineers, eliminating the need for ETL or manual IT assistance.

The Challenge

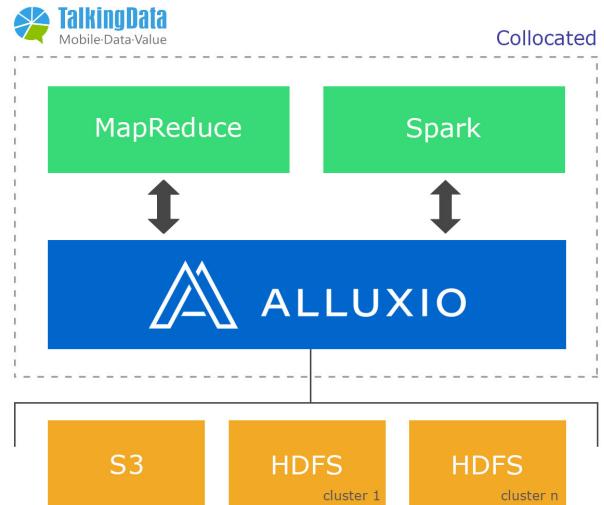
TalkingData is China's largest data broker, reaching more than 600 million smart devices on a monthly basis and processing over 20 terabytes of data and more than one billion session requests every day. TalkingData products are powered by its massive proprietary data set and provide services to over 120,000 mobile applications and 100,000 application developers. The company serves a wide range of clients in both Internet and traditional industries, including leading enterprises in the financial services, real estate, retail, travel, and government sectors. TalkingData gathers data from a wide variety of data sources, and the ability to access, store, and process that data is mission critical. Data can be stored at the customer (on-premise or in the cloud) as well as in the company's own various storage systems. Previously, for a department to perform analytics or process data it required a request to the central IT department who would determine where the data is located and then transfer it from the various locations to a centralized storage system. This created a complex environment that data scientists, analysts, and IT needed to manage on a daily basis.

It became clear that TalkingData needed a solution with the following characteristics:

- **Single platform to manage data access across disparate sources.** Data is collected from hundreds of millions of devices and directly from customers then stored in a wide variety of storage systems including HDFS, AWS S3, and Ceph.
- **High Performance.** The speed with which data can be transformed into value directly impacts the top line and provides a competitive advantage. Avoiding ETL would remove a time consuming step.
- **Integrated with best of breed big data technologies.** Serving a broad customer base across a wide range of industries requires the ability to match the right application framework to the required task.
- **Flexible deployment on-premise, in the cloud, or hybrid.** Data needs to be accessed seamlessly regardless of the physical location.
- **Scalable.** The volume of data is growing beyond terabyte-scale, and with the rise of IoT increasing exponentially; cost-effective scalability is critical.

The Solution

TalkingData deployed Alluxio as a single platform to manage all data across disparate data sources on-premise and in the cloud. Alluxio removes the complexity of the environment by abstracting the different data sources and providing a unified interface. Applications simply interact with Alluxio, and Alluxio manages data access to different storage systems on behalf of the applications. Alluxio effectively democratizes data access, allowing data scientists and analysts in various business units to accomplish their goals without needing to consider where the data is located or having to go to central IT or the engineering team to transfer or prepare the data.



Looking Forward

With Alluxio, TalkingData now has a single platform that manages data across disparate data sources at memory speed while providing a flexible deployment model. Additionally, Alluxio decouples storage from compute which enables the company to scale each resource independently. With Alluxio, data scientists and analysts can work with the application frameworks and storage systems of their choosing without any of the complexity. For TalkingData that is a game changer. The goal for TalkingData is to build a smart data marketplace where companies in any industry can purchase the exact type of data they need and directly get value from the data without having to process it because the data has already been processed and analyzed. Alluxio is a key enabling technology to achieving this goal. Alluxio is currently deployed on hundreds of nodes in production with plans to expand to more use cases, products, and deploy at a larger scale.

Stay Connected

Twitter: @Alluxio

LinkedIn: linkedin.com/company/alluxio-inc-/

Meetup: meetup.com/Alluxio/

Slack: 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 or follow us on LinkedIn, or Twitter.



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

www.alluxio.io