

ACHIEVING HYBRID AND MULTI-CLOUD ARCHITECTURE WITH APPLICATION PORTABILITY

COMPANY	Fortune 50 Technology Company
LOCATION	SF Bay Area, CA
EMPLOYEES	10k+
TOPOLOGY	Hybrid and Multi-cloud
COMPUTE	Spark, Trino, Hive
STORAGE	AWS S3, on-prem HDFS & object stores

ALLUXIO BENEFITS

- Gain agility for hybrid and multi-cloud
- Access data anywhere without app changes
- Reduce S3 egress cost

This Fortune 50 company serves over 1 billion users worldwide and ingests and retains petabytes of data in both on-premises and cloud managed by the data platform team.

Modernizing the data platform with agility and onboarding more teams is critical to this tech giant. A future-proof data infrastructure will accelerate the time-to-insights to enhance customer experience, boost operational efficiency, and more.

CHALLENGE: Unable to Bridge On-premises and Public Cloud

Before adopting Alluxio, on-premises Spark and Trino applications could not directly run on S3 data since they could only use HDFS APIs. The data platform team had to copy the S3 data to HDFS and came across the following challenges:

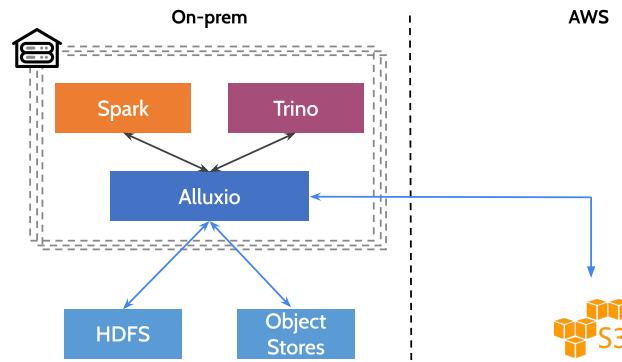
- **Huge S3 Egress Costs:** Data scientists constantly need to retrieve data in S3 buckets, resulting in high egress costs.
- **Long Time to Insight:** Data is not immediately available, leading to poor user experience and long time to insight.
- **Cloud Journey is Hindered:** Because data resides in both HDFS and S3, applications must be reprogrammed for data access. This hinders the adoption of hybrid and multi-cloud.

SOLUTION: Achieving Hybrid and Multi-Cloud with Application Portability

From a strategic viewpoint, the tech giant wants the agility to deploy their applications and compute capacity to any environment based on cost and operational overhead. As the company was looking for a solution to achieve hybrid and multi-cloud architecture, it turned to Alluxio.

Alluxio provides a data layer between storage and compute engines. This data layer completely virtualizes all the storage systems including on-premises HDFS, object stores, and S3 in AWS. Spark and Trino applications do not need to care about the location of data anymore.

The company deployed a standalone Alluxio cluster on-premises. The Alluxio cluster provides a single interface and governance for applications to access data anywhere.



RESULTS: Significant Business and Technical Benefits Achieved

Business Values

- **Gain Agility for Hybrid and Multi-Cloud:** With Alluxio, the applications have a single access to all the data. During migration, the data platform continues serving data science teams with minimum impact on their applications.
- **Reduce S3 Egress Cost:** There is no need to manually copy data. Alluxio presents the data to the applications and avoids repeatedly fetching remote data, thus significantly reducing S3 egress cost.
- **Shorten Time-to-insight:** Data is now immediately available to users without waiting for being copied over. Data scientists enjoy data availability, leading to shorter time-to-insights.

Technical Benefits

- **Access Data Anywhere with Zero Application Reprogramming:** Because Alluxio translates data access requests from applications into underlying storage interfaces, Spark and Trino applications continue to use HDFS API with no S3 reprogramming.
- **Standardize on A Common Data Stack Across Heterogeneous Environments:** Alluxio enables a standardized data stack with unified access to data. The platform is multi-cloud ready and applications have gained portability.
- **Achieve a Future-proof Architecture for Compute and Storage Innovations:** The data platform has decoupled data management from elastic compute resources and is ready to accommodate the future state of compute and storage technology.

LOOKING AHEAD: The Next-Generation Data Platform

When looking forward, the company wants a balance of reduced operational complexity and cost optimization. The next-generation data platform should be able to spin up compute on-premises or in cloud services from different vendors with access to data anywhere. Alluxio is expected to help the company achieve this vision in the future.

WHY ALLUXIO

- | | |
|---|---|
|  Gain agility for hybrid and multi-cloud |  Access data anywhere without app change |
|  Reduce S3 egress costs |  Standardize on a common data stack |
|  Shorten time-to-insights |  Achieve a future-proof architecture |

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 is in production use today at eight out of the top ten internet companies. Venture-backed by Andreessen Horowitz, Seven Seas Partners, Volcanic Ventures, and Hillhouse Capital. Alluxio was founded at UC Berkeley's AMPLab by the creators of the Tachyon open source project.

For more information about Alluxio, go to: <https://www.alluxio.io/>.