

MAXIMIZING EFFICIENCY AND REDUCING S3 EGRESS COST WITH HYBRID CLOUD DATA ACCESS

COMPANY	Comcast
INDUSTRY	Telecom, Media, & Entertainment
LOCATION	United States
EMPLOYEES	180k+
TOPOLOGY	Hybrid Cloud
COMPUTE	Presto (on-prem)
STORAGE	Amazon S3 (cloud)

ALLUXIO BENEFITS

- Significantly reduce S3 egress cost
- Efficient data management without making data copies
- Access remote S3 storage with high and consistent performance
- Simplify data access model with security

Comcast is the second largest broadcasting and cable television company in the world by revenue and is America's largest home Internet service provider. The Data Experience Team (DX) at Comcast is responsible for data engineering and data governance for Comcast data platforms, focused on gathering, organizing, and making the data available for business analysis.

Comcast has implemented a hybrid cloud architecture for its data platform. The data platform leverages Presto as the query engine and Amazon S3 for storage. This hybrid cloud architecture offers a continual "build vs. buy" cost modeling, allowing the company to leverage its existing on-premises compute resources while also taking advantage of the benefits of the public cloud, such as on-demand provisioning and scaling and reliability. While hybrid cloud has many benefits, it also presents some challenges.

CHALLENGE: Data Access in Hybrid Cloud is Inefficient, Expensive and Complex

- **Remote data access introduces network latency.** Network latency can occur during remote data access due to the physical distance between data centers. This can impact time-critical workloads and make data access slower.
- **Frequently retrieving S3 incurs high egress costs.** Data retrieval from a public cloud incurs network egress costs. This significantly increased the long-term TCO of the data platform.
- **Copying data is resource-consuming and complex.** Because data access on large, remote data significantly impacts performance, the team had to copy data locally to on-premises local storage, which is resource-intensive, complex, and error-prone.
- **Lack of unified data access model for storage protocol.** There is a storage protocol mismatch for applications between on-premises and public cloud. The on-prem applications have to be re-engineered for compatibility with the cloud storage.

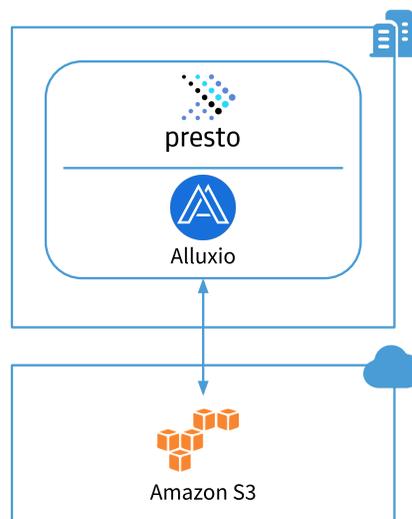
"Alluxio has proven to be a valuable solution in addressing the data access challenges of hybrid cloud for Comcast. It has provided us with faster data access, reduced egress costs, and streamlined data management, resulting in more efficient and effective data value creation for the organization."

-MICHAEL FAGAN, Distinguished Architect at Comcast

SOLUTION: Seamless and Secure Data Access in Hybrid Cloud

The DX team turned to Alluxio, which provides a new layer between storage and compute engines for a variety of different data-driven applications. Alluxio acts as a universal data plane for data access to the hybrid data lake when egressing data from AWS to on-premises.

- **Caching to eliminate data copies in hybrid cloud.** Alluxio provides caching for the Presto query engine running on-premises to access data from S3 in AWS. Alluxio caches the hot data and brings frequently accessed data to Presto. It eliminates the need to copy data from the cloud to local storage.
- **Flexible APIs to bridge compute and storage.** Alluxio translates data access requests from applications into underlying storage interfaces. On-prem applications can continue using HDFS API with no S3 reprogramming. Thus, applications no longer need to be tuned for specific storage types.
- **A universal security model with Apache Ranger integration.** Alluxio provides integration with Apache Ranger using a Ranger Plugin so Comcast can leverage on-prem security infrastructure and continue enforcing the same access control policies in the cloud.



RESULTS: Maximize Efficiency and Reduce S3 Egress Cost

- **Significantly reduce S3 egress cost.** By caching data, Alluxio helps the company avoid repeatedly fetching data directly from the cloud storage, significantly reducing the amount of data transferred over the network.
- **Efficient data management without making data copies.** Because Alluxio provides seamless data access in hybrid cloud, there is no need to copy data before it can be accessed. By implementing Alluxio, data management and operations are streamlined.
- **Access remote S3 storage with high and consistent performance.** By using Alluxio, Comcast was able to improve query execution times and better predict performance compared to directly accessing remote data.
- **Simplify data access model with security.** Alluxio provides a unified data access model, bringing simplicity to enforcing authorization and access controls.

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/>.