

HIGH PERFORMANCE DATA ACCESS LAYER FOR AI

Challenges Serving Data to AI/ML Pipelines



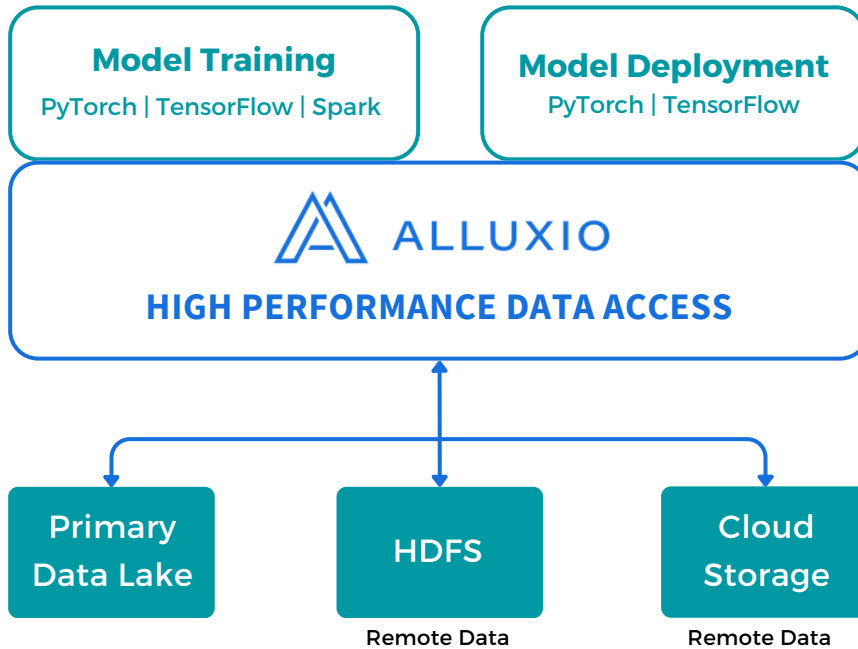
Low GPU utilization &
GPUs are expensive and scarce



Specialized hardware
is needed to feed GPUs



Extensive data engineering
to manage data copies



- Deploy GPUs anywhere
- End-to-end ML pipeline acceleration
- Run on standard low cost hardware
- Read and write data to single source of truth



>90% Increase GPU utilization from 20-30% to 90%+



Faster model training and fresher models



Increase **productivity** of data engineering teams

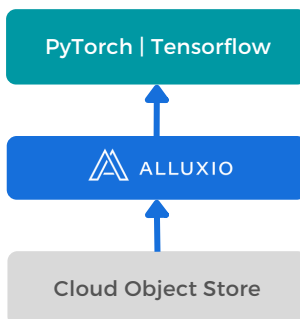


Lower TCO of AI infrastructure

AliPay: Large-Scale Computer Vision Model Training on Billions of Files

The world's largest mobile payment platform, serving over 1.3 billion users and 80 million merchants

Architecture with Alluxio



Results

- Achieve faster model training on billions of small files
- Increase GPU utilization and lower infrastructure costs
- Reduce engineering time spent managing data access and copies

“ We have tried various methods to improve our AI training but only Alluxio has made a significant difference. ”
— Chuanying Chen, Senior Software Engineer at AliPay

In Production at Leading Organizations

HIGH PERFORMANCE DATA ACCESS LAYER FOR ANALYTICS

Challenges of Modernizing Data Platform



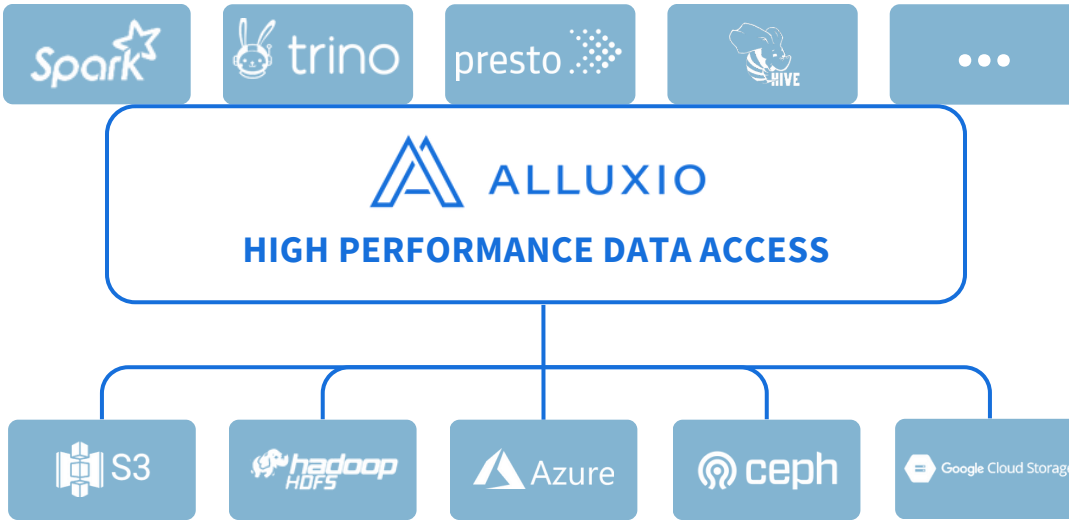
Data gravity and federated data access



Out-of-control cloud & network costs



Inconsistent and slow performance



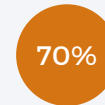
- Unified access across all data sources
- Data caching to accelerate queries
- Seamless migration to the cloud
- Onboard new applications with ease



Data access across data centers, regions or clouds without copy



Application portability with consistent experience across all environments

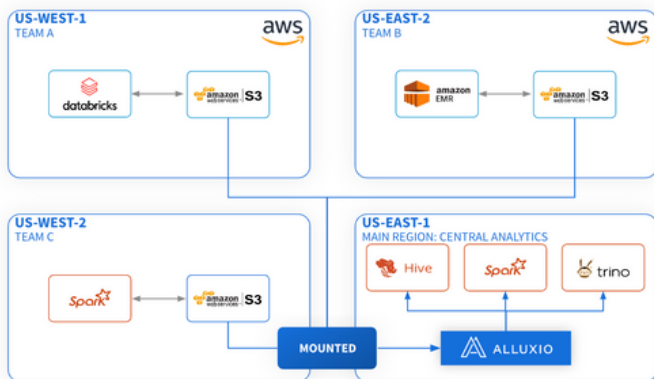


Up to 70% TCO savings, including cloud storage and operations cost

Expedia: Unify Data Lake Across Multiple Geographic Regions in the Cloud

American top online travel shopping company with a portfolio of 20+ brands

Architecture with Alluxio



Results

- Unify data silos without the need to copy or move data
- Enhanced user experience with consistent and high performance analytics, reducing time to insights
- 50% reduced S3 egress costs per query

“ With the introduction of Alluxio, we see better performance, increased manageability, and 50% lower costs.
 — Jian Li, Senior Software Engineer at Expedia Group ”

About Alluxio

Alluxio, a high performance data access layer for AI and analytics, makes it easy to manage your data and serve it from any storage to any compute engine in any environment — on-premise, in the cloud, or across clouds. Alluxio powers 7 out of the top 10 internet companies globally. To learn more, contact info@alluxio.com or visit www.alluxio.io.