

Accelerating analytics & AI in Kubernetes with Alluxio Open Source Data Orchestration

KEY BENEFITS

- **Data locality on demand:**

Alluxio caches data close to compute for your big data analytics or machine learning workloads

- **Data sharing:**

Alluxio gives you high-speed data sharing across compute jobS

- **Data abstraction:**

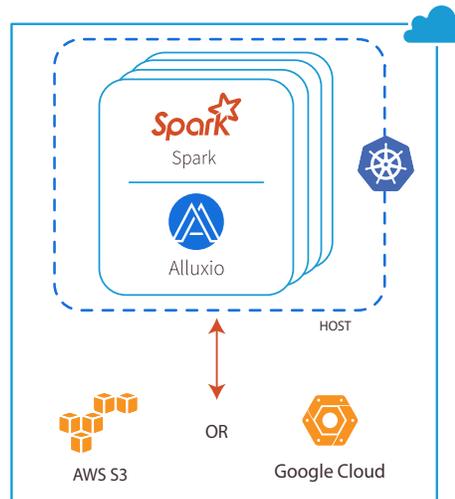
Alluxio enables data abstraction across data silos

Kubernetes is widely used across enterprises to orchestrate computation. While Kubernetes helps improve flexibility and portability for computation in public/hybrid cloud environments across infrastructure providers, running data-intensive workloads on it can be challenging.

For data-driven workloads in disaggregated stacks, there's no native data access layer within a Kubernetes cluster. For query engines and machine learning frameworks that are deployed within a Kubernetes cluster, any critical data sitting outside the cluster breaks locality. Alluxio can help.

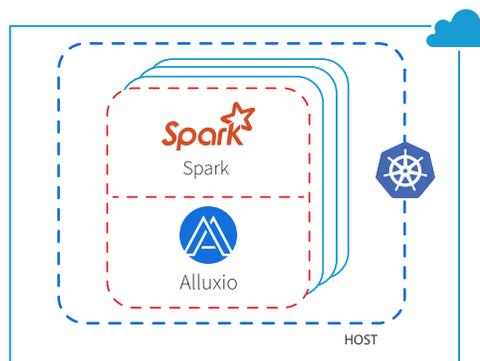
ELASTIC DATA FOR ELASTIC COMPUTE WITH TIGHT LOCALITY

With Alluxio in Kubernetes, get data locality on demand, data sharing across compute, and data abstraction across data silos. A full deployment may look something like this:

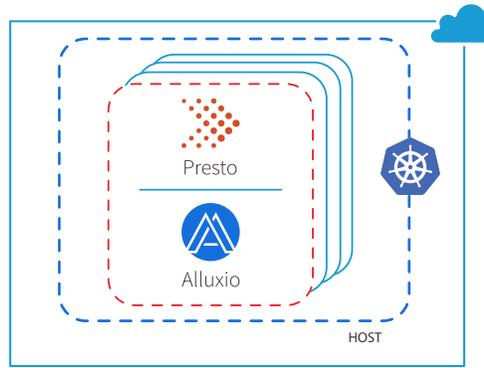


Depending what your use case is, there are a few ways to deploy Alluxio with compute in K8s.

Alluxio and the compute framework in different pods on the same host for use when your compute (like Spark) is short running and ephemeral:

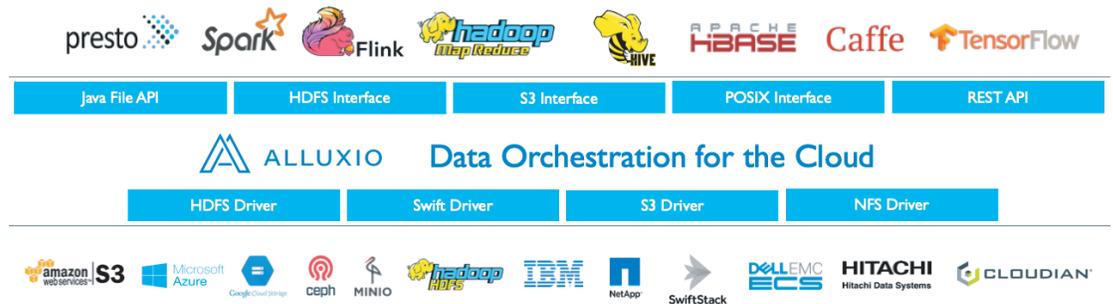


Alluxio and the compute framework in the same pod for use when compute (like Presto) is long running:



MEET ALLUXIO

Alluxio brings your data to compute, on demand. It's an open source data orchestration layer for big data and AI/ML workloads in the cloud, enabling data locality, data accessibility, and data elasticity.



RESOURCES

Online tech talk: Accelerating Spark with Kubernetes

<https://www.alluxio.io/resources/presentations/tech-talk-accelerating-spark-with-kubernetes/>

Docs: Deploying Alluxio on Kubernetes

<https://docs.alluxio.io/ee/user/stable/en/deploy/Running-Alluxio-On-Kubernetes.html>

Tutorial: Get started with Alluxio/Presto Sandbox using Docker

<https://www.alluxio.io/alluxio-presto-sandbox-docker/>

About Alluxio

www.alluxio.io

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.



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



@Alluxio



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



alluxio.io/slack



meetup.com/Alluxio/