

Ultra-fast SQL Analytics using PAS (Presto on Alluxio Stack)

Haoyuan (H.Y.) Li | Founder & CTO | Alluxio | haoyuan@alluxio.com

Bin Fan | Founding Engineers | Alluxio | binfan@alluxio.com



ALLUXIO
2019

Overview

History and Open Source Community



The Alluxio Story

–amplab
2013

Originated as Tachyon project, at the UC Berkley's AMP Lab by then Ph.D. student & now Alluxio CTO, Haoyuan (H.Y.) Li.

 ALLUXIO
2015

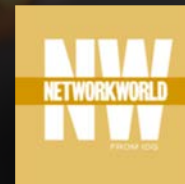
Open Source project established & company to commercialize Alluxio founded

ANDREESSEN
HOROWITZ

Goal: **Orchestrate Data for Analytics & ML in the Cloud** for data driven apps such as Big Data Analytics, ML and AI.

CRN

2018



2018



2019

Open Source Started From UC Berkeley AMPLab



1000+ contributors & growing



4000+ Git Stars

Join the conversation on
Slack
slackin.alluxio.io



Apache 2.0 Licensed



GitHub's Top 100 Most Valuable Repositories
Out of 96 Million



Companies Running Alluxio ([Learn More](#))

Financial Services



Retail & Entertainment



Data & Analytics Services



Technology



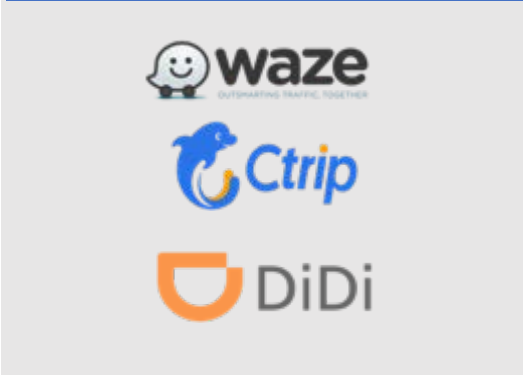
Consumer



Telco & Media



Travel & Transportation



Alluxio: a Data Orchestration System



Java File API

HDFS Interface

S3 Interface

FUSE Interface

REST API

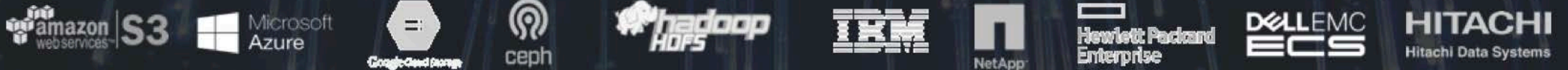
Data Orchestration for the Analytics & AI for the Cloud

HDFS Driver

Swift Driver

S3 Driver

NFS Driver



Data Orchestration Requirements

Global
Namespace

Intelligent
Caching

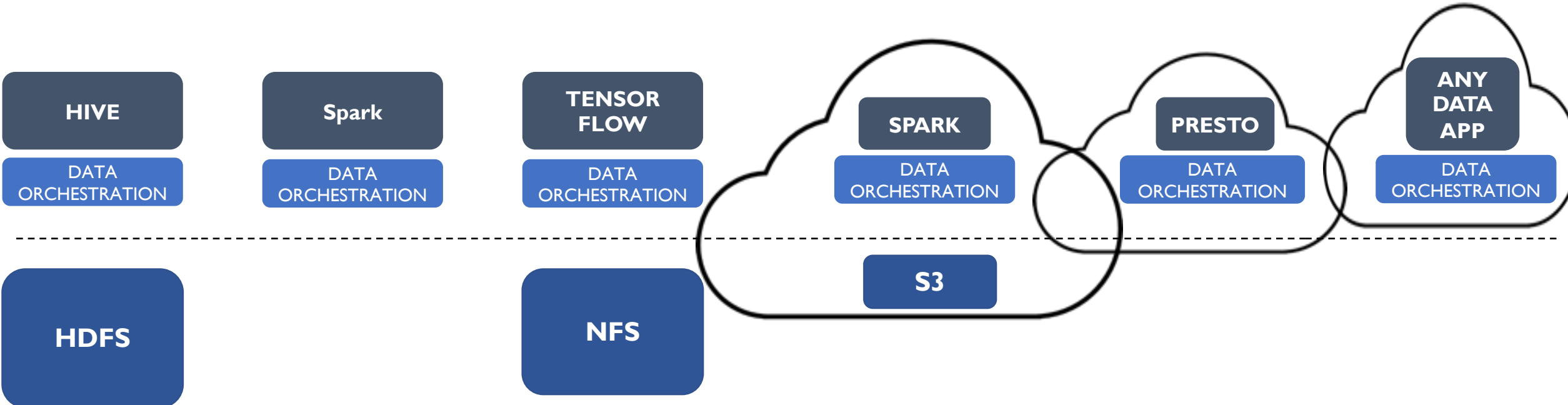
Data
Management

Structured
Data Catalog

Data
Transformation

Data Platform using a Data Orchestration Approach

COMPUTE SPREAD ACROSS MANY DIFFERENT FRAMEWORKS



DATA IN DISPARATE STORAGE SYSTEMS



Presto on Alluxio Today

Architecture, Benefit, Production Cases

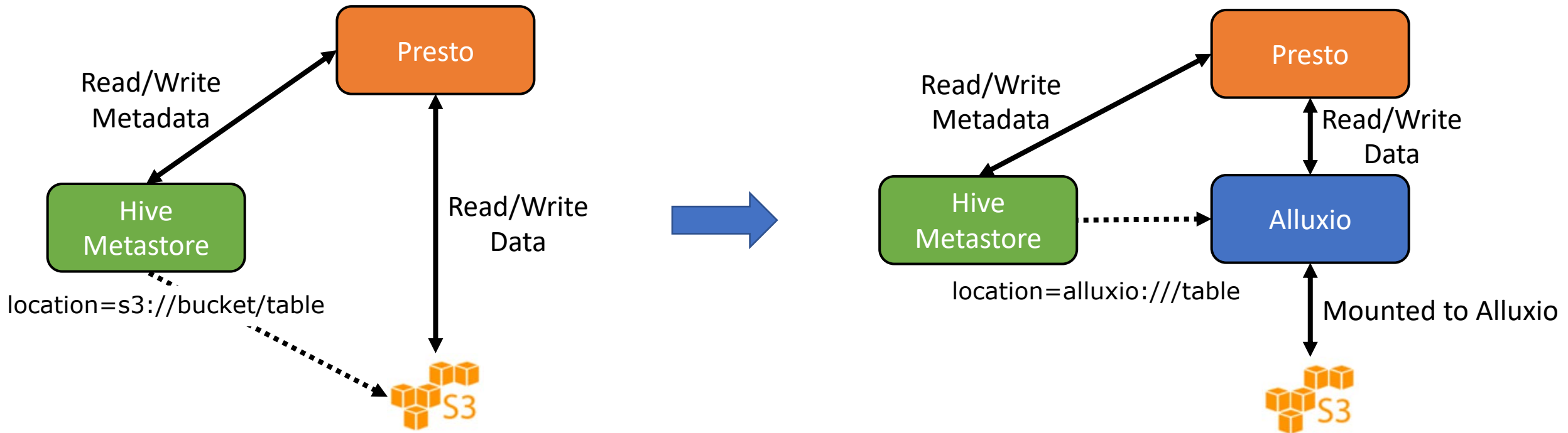
ALLUXIO
2019

Why Presto on Alluxio

- Distributed Data Caching on Demand
 - Lower query latency
 - More consistent performance
 - More concurrency
 - Less data transfer
- Deeper Presto Alluxio Integration
 - New Alluxio catalog service
 - New Alluxio transformation service

Now available as Developer Preview in v2.1

How Presto Works with Alluxio



How to Use Alluxio in Presto CLI

Create A Table on Alluxio

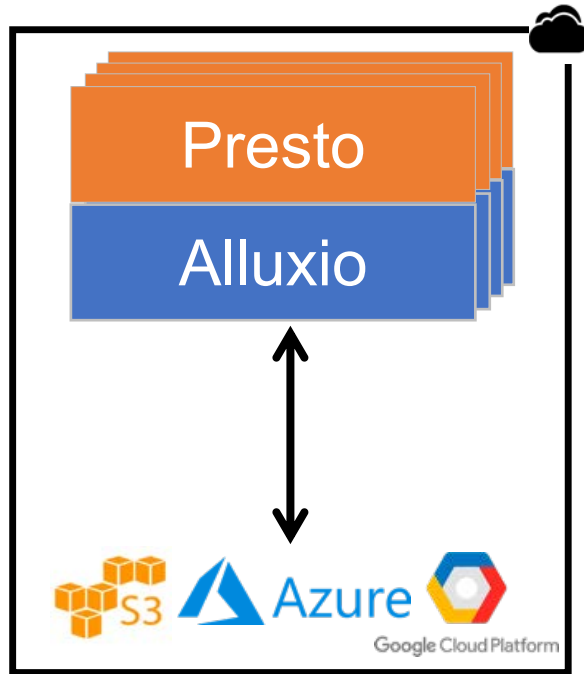
```
> CREATE TABLE alluxio_table (id varchar)
  WITH (external_location = 'alluxio:///table');
```

Read A Table from Alluxio

```
> SELECT * FROM alluxio_table
```

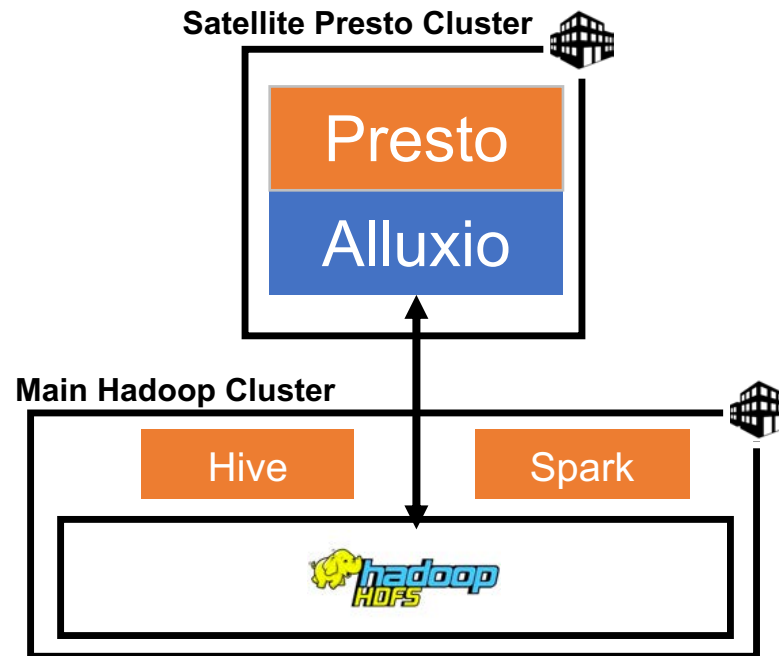
Common Use Cases

Accelerate query performance
as cloud storage caching



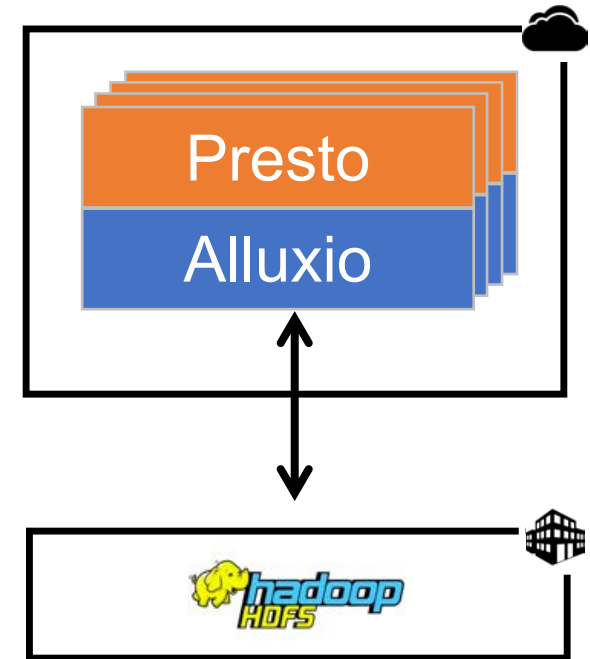
Read more
[Case Study from Ryte](#)

On-premise satellite
compute clusters across data centers



Read more
[Case Study from JD.com](#)

Zero-copy burst workloads in
hybrid cloud environments



Read more
[Case Study from WalmartLab](#)



Now available as Developer Preview in v2.1

Alluxio Structured Data Management

Deeper Integration with SQL Engines like Presto

ALLUXIO
2019

Storage Systems

SQL Frameworks

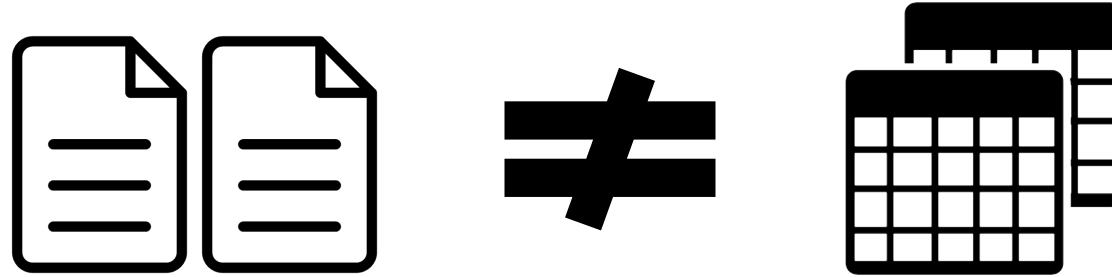
Files/Objects

Directories

Raw Bytes

Cost-efficiency

Durability



Impedance Mismatch



Further Expand Benefits!

Tables

Schemas

Rows/Columns

Compute-optimized

Computation

Benefits of Alluxio Data Orchestration

Caching

Unified Interface/Namespace

Schema-Aware Optimizations

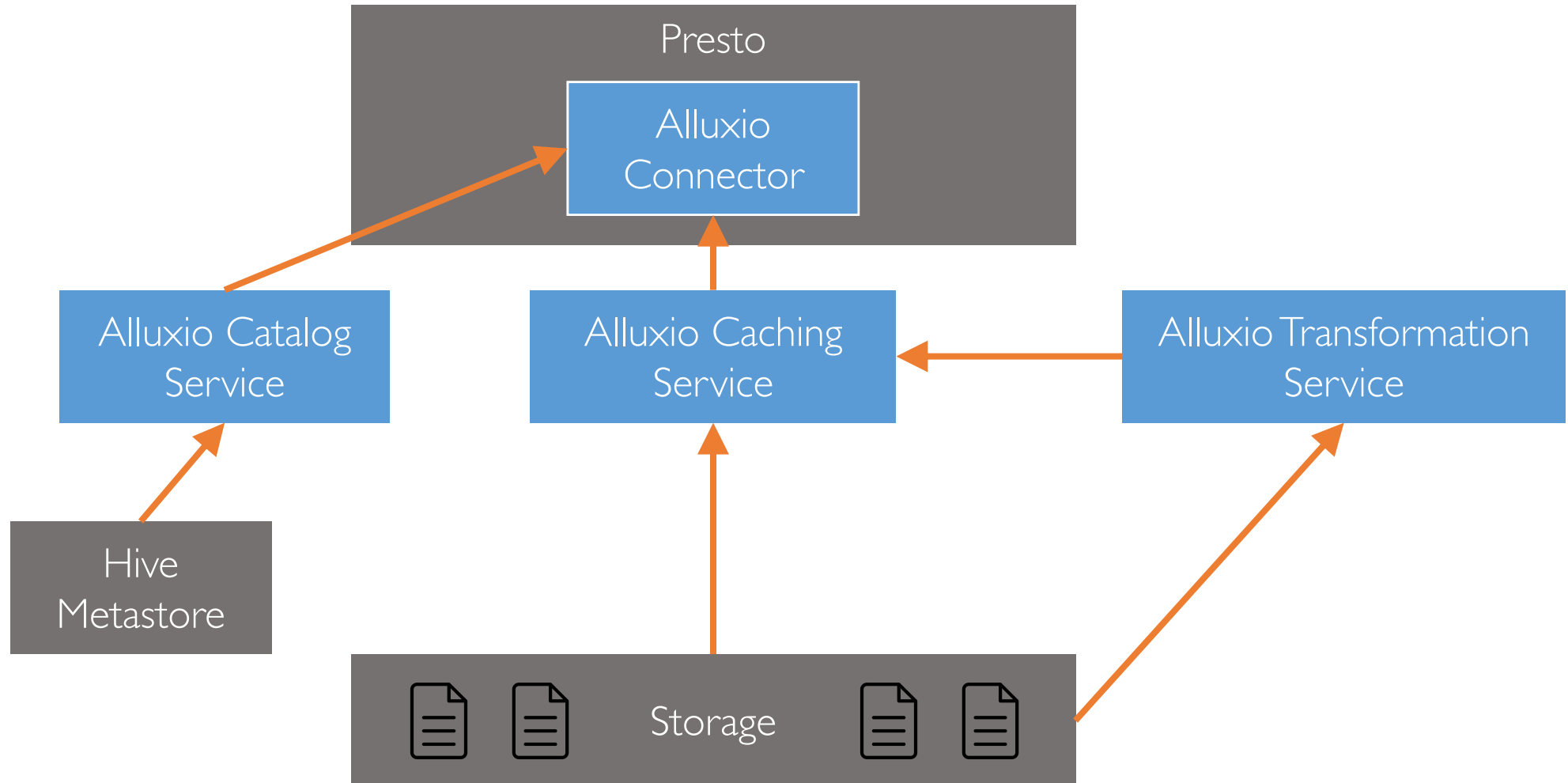
Compute-Optimized Formats

Physical Data Independence

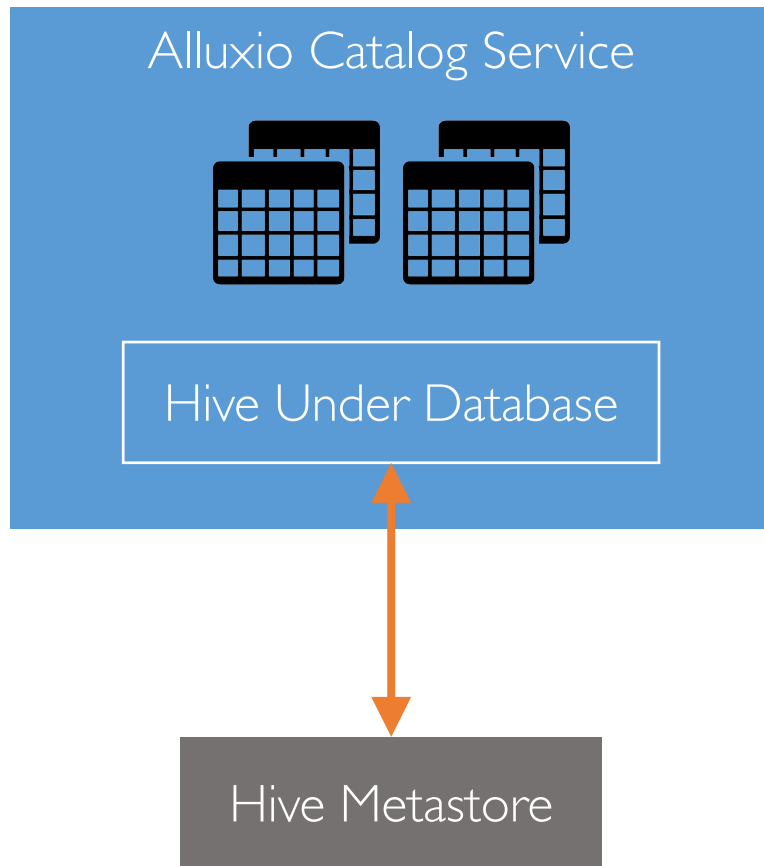
Storage
Systems

SQL
Frameworks

Alluxio Structured Data Management Preview



Alluxio Catalog Service



Functionality

Manages metadata for structured data

Abstracts other database catalogs as Under Database (UDB)

Benefits

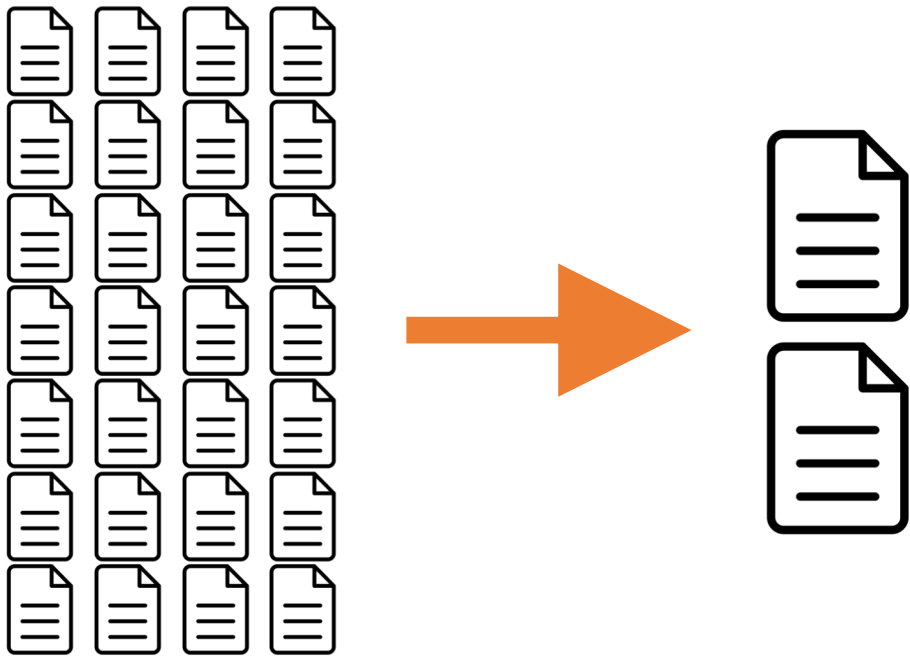
Schema-aware optimizations

Simple deployment

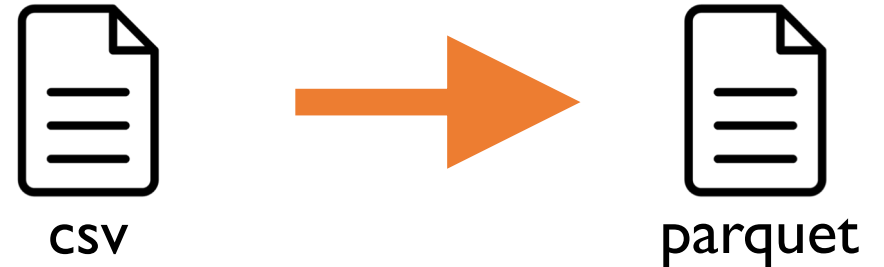
Transformation Service

Transform data to be compute-optimized
independent of the storage format

Coalesce



Format Conversion



Join Our User Community



Join Slack channel
alluxio.io/slack



Join meetup groups near you
alluxio-open-source-community/



Wechat Public Account

Next Step

- Check out more tutorials <https://www.alluxio.io/presto/>
- Additional Reads:
 - Starburst Presto + Alluxio = better together
<https://www.starburstdata.com/technical-blog/starburst-presto-alluxio-better-together/>
 - Top 5 performance tips running Presto with Alluxio
<https://www.alluxio.io/blog/top-5-performance-tuning-tips-for-running-presto-on-alluxio-1>
 - Presto + Alluxio + Hive Metastore on your Laptop in 10 min
<https://www.alluxio.io/blog/tutorial-presto-alluxio-hive-metastore-on-your-laptop-in-10-min/>