USE CASE

Unified Data Access

On-prem Compute Acceleration

APPLICATION STACK

Spark + Drill + Hadoop + Custom AI & Analytics Alluxio

HDFS + NFS + Object Storage

CHALLENGES

- Operational complexity with access to multiple data stores
- Compatibility with disparate compute services

BENEFITS

- Centralized data access for all data
- Accelerated perfor-mance with in-memory caching
- Hardware flexibility with tiering of storage

Aunalytics Leverages Alluxio as a "one-stop-shop" for Data I/O

Alluxio is a leading data orchestration platform that offers a compute agnostic, storage agnostic, and cloud agnostic solution for big data and machine learning applications. Aunalytics is a data platform company delivering Insights-as-a-Service to answer enterprise and mid-sized companies' most important IT and business questions.

Aunalytics Daybreak™ industry intelligent data mart combined with the power of its data platform provides industry-specific data models with built-in queries and AI to ensure access to timely, accurate data and answers to critical business questions. Aunalytics leverages internally developed analytics software on top of open source compute technologies to deliver enriched data marts to clients in a question-and-answer format. The Daybreak product is usable by any business personnel irrespective of tech-prowess and leverages Natural Language Processing to deconstruct real-world questions into derived data queries.

Challenge: Digital Transformation gap in search of a single solution for the cloud

As a startup in 2012, Aunalytics was faced with the adoption of very complex analytics environments that Fortune 500 companies were struggling to implement and adopt. The Aunalytics team sought to eliminate the knowledge gap and personnel deficiencies that the middle-market clients experienced when trying to address what we now call the Digital Transformation gap. The team was successful in doing so but also generated operational overhead as an early adopter of these technologies.

In 2020, Aunalytics implemented the next generation of their compute platform which sought to completely segregate storage, compute, and delivery for ease of development and scale, with an adaptive analytics software (Aunsight™) which is capable of submitting workloads to many compute services sourcing data from many storage resources.

The Aunalytics team faced the option of either

1. Centralizing on a well-known storage environment such as NFS, iSCSI, etc. which would be fraught with performance, consistency, and concurrency issues, OR
2. Adopting a many/any storage environment with a single point of access for those systems.

As an alternative to a new stack with Alluxio, Aunalytics evaluated cloud computing platforms which bundle storage and could be deployed in a private cloud. Instead the team adopted Alluxio, and sought to centralize the data I/O footprint on top of one technology to simplify integrations with newer compute engines going forward.
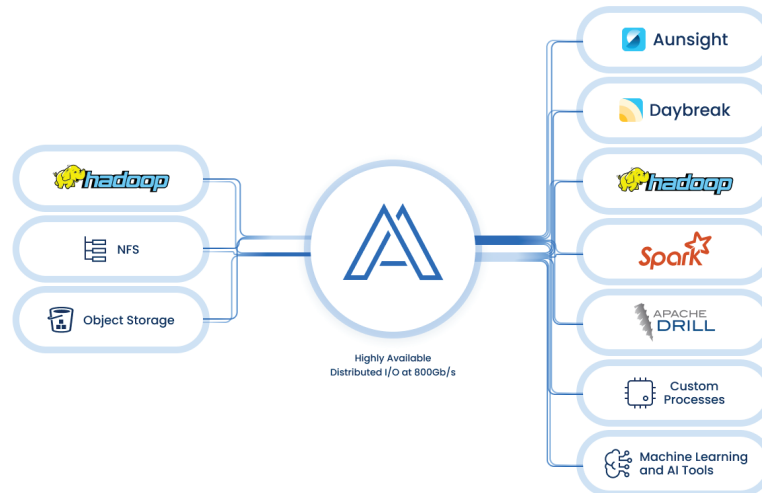
Why Alluxio

Alluxio presents Aunalytics compute environments with a single point of access for all stored data regardless of its volume, velocity, or storage I/O performance capabilities. This allows the team's developers to focus on analytics integrations irrespective of storage environment limitations or compatibility. Alluxio is the primary data access and write mechanism for the batch compute environ-ment, dynamic query environment, and primary product Daybreak.

Alluxio has allowed the Aunalytics team to leverage existing legacy Hadoop storage, new private-cloud storage, and future bulk object storage for cost-efficient scalable storage flexibility over time. Aunalytics avoids the complexity of supporting legacy environments as the platform continues to evolve.

Integrating Alluxio

Alluxio is deployed as part of the new Aunalytics data environment. While Alluxio enables the team to do many new things, it is not implemented as a supplemental system. Alluxio allows the team to build the Aunalytics next-gen platform with the flexibility the team desires where other options would have forced some combination of adoption and sacrifice.

Highly Available
Distributed I/O at 800Gb/s

Realized Benefits

Alluxio delivered a 'one-stop-shop' for data I/O. It enabled the Aunalytics team to transition away from large monolithic analytics environments such as Hadoop while still leveraging the hadoop storage ecosystem during migrations. It also has enabled the team to perform 'dead-drop' handoffs of data between analytics systems where data can be written to Alluxio by one system and read by a completely separate incompatible system a moment later.

Sharing data across systems has decreased the data movement/copy requirements by up to 90% post-migration and improved the compute and delivery times by up to 30%.

Next Steps

Alluxio fundamentally enables separation of storage and compute. It brings speed and agility to big data and AI workloads, and reduces costs with the elimination of data duplication by enabling users to move to newer storage solutions like object stores. The team continues to leverage open source compute projects adjacent to the Daybreak product, supported by Aunsight™ and Alluxio, in the AI workspace to deliver collective products. Aunalytics does the heavy technical lifting of combining Alluxio and other data management technologies for its clients, to provide an out of the box analytics solution with built-in data management ready for business users.

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

@Alluxio

linkedin.com/company/alluxio-inc-/

alluxio.io/slack

meetup.com/Alluxio/