



ALLUXIO

GET INSIGHTS FASTER WITH ALLUXIO AND INTEL

In today's data centers, bounded storage and compute resources on Hadoop* or Spark* nodes create challenges around data capacity, data silos, costs, performance, and efficiency. Alluxio's data orchestration platform, combined with 2nd Gen Intel® Xeon® Scalable processors and Intel® Optane™ persistent memory, simplifies data management and processing and significantly accelerates performance for today's big data and AI/ML workloads.

Alluxio and Intel's joint solution allows companies to unify on-premises and cloud data silos into a single, cloud-based data layer, increasing data accessibility and elasticity while virtually eliminating the need for copies—for less complexity, lower costs, and greater speed and agility.

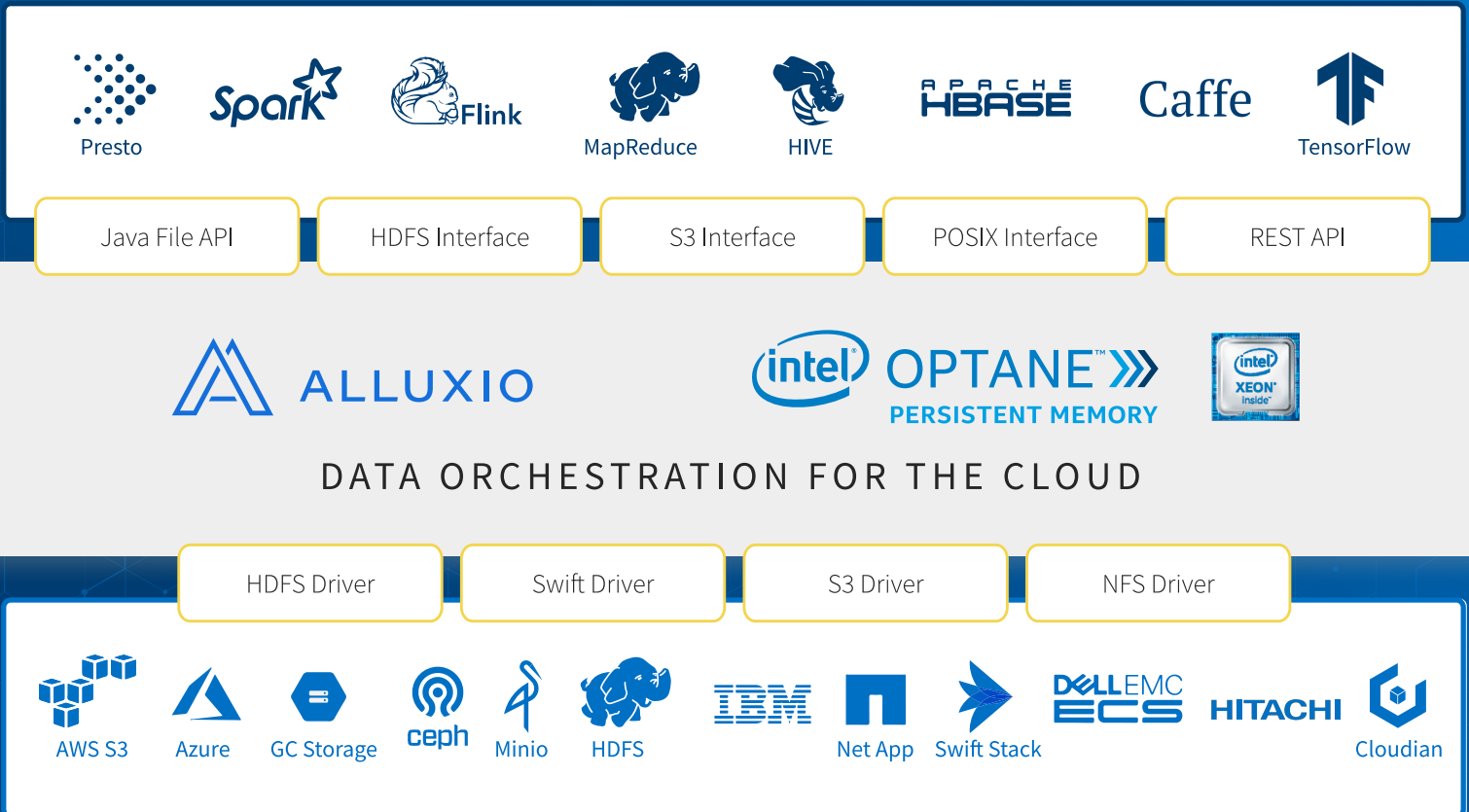
CONTACT US

1825 S Grant Street, Suite 600
San Mateo, CA 94402

alluxio.io/contact
sales@alluxio.com

Open Source Data Orchestration for Cloud Analytics

Alluxio creates a fast, shared data access layer between persistent storage systems and computation frameworks, dramatically speeding up access to files and objects on your storage systems for better data management and faster time to insights.



Alluxio's memory-centric data layer architecture allows organizations to:

Dramatically speed up user applications and computation frameworks.

With Alluxio, data is only read once from the storage system, so data can be served at memory speed when local, and at the computation cluster network speed when it is in Alluxio. This significantly accelerates access to data from slower connected storage. To achieve the best performance, Alluxio is recommended to be deployed alongside a cluster's computation framework.


Expand the set of workloads available to utilize its data.

Alluxio creates a unified storage layer between big data applications and traditional storage systems so applications and frameworks can transparently and quickly access data from any storage system through Alluxio, including multiple under storage systems with varied data sources.

Alluxio + Intel: Better Together


2nd Gen Intel® Xeon® Scalable processors and Intel® Optane™ persistent memory provide Alluxio with an in-memory acceleration layer, significantly boosting storage capacity and performance for rapid data processing.

2.13x
Speedup¹



over local HDFS for 4TB parquet format data, on
Decision Support workload with Intel Optane
persistent memory

1.92x
Speedup¹



over DRAM Cache for 4TB parquet format data, on
Decision Support workload with Intel Optane
persistent memory

“Together with Intel, we plan to disrupt the advanced analytics and AI status quo with an in-memory data accelerator layer to accelerate intermediate data access and ease data bottlenecks that many of our customers are highlighting as key challenges with their increasing big data requirements.”

**Rowan Scranage, VP of Business
Development at Alluxio.**

Why Intel Optane Persistent Memory?

Intel Optane persistent memory redefines traditional architectures, offering a large and persistent memory tier at affordable cost. With breakthrough performance levels in memory intensive workloads, virtual machine density, and fast storage capacity, Intel Optane persistent memory—combined 2nd Gen Intel Xeon Scalable processors—accelerates IT transformation to support the demands of the data era, with faster-than-ever-before analytics, cloud services, and next-generation communication services. The larger capacity of Intel PMem enables companies to consolidate workloads onto fewer nodes for better utilization and TCO.

In addition, Intel Optane persistent memory helps Alluxio improve big data application performance by caching your data locally on compute nodes.

Why choose Alluxio + Intel?

- **Simplified data access**

Data is accessible through one unified namespace, regardless of where it resides. Once data from on-premises Hadoop* clusters are in Alluxio, the same data can be accessible in many different ways using many different APIs including the HDFS API, S3 API, POSIX API and others. Rather than communicating to each individual storage system, applications can delegate this responsibility by connecting through Alluxio, which will handle the different underlying storage systems. This means all existing applications built for analytical and AI workloads can run directly on this data without any changes to the application itself.

- **Data locality**

Data is local to compute, giving you memory-speed access for your big data and AI/ML workloads. With Alluxio, your on-premise data gets co-located within the same instance as the compute framework, allowing data engineers to get the data faster to be extracted and analyzed.

- **Data on-demand**

Data is as elastic as compute so you can abstract and independently scale compute and storage. Alluxio can be elastically scaled along with the analytics frameworks, including in container orchestrated environments. Data can also be replicated within the Alluxio cluster with each file being replicated as many times as needed. Each application and storage system can be accessed using the same interface and the data exists in less formats, requiring no code changes or no complex integration.

READY FOR FASTER INSIGHTS FROM YOUR AI/ML WORKLOADS?

Learn more and request a demo at

alluxio.io/contact | sales@alluxio.com



¹ Tested by Intel as on 12/06/2019 with the following configuration. Detailed white paper will follow.

Optane PMem cache: 2 socket Intel® Xeon® Gold 6240 Processor, 18 cores HT On Turbo ON Total Memory 192 GB (12 slots/ 16GB/ 2666 MHz), DCPMM 1TB (8 slots/ 128GB/ 2666 MHz), DCPMM firmware version: 01.02.00.5410, BIOS: SE5C620.86B.0X.02.0094.102720191711 (ucode:0x500002c), BKC version: ww08.2019, Fedora 29 (Server Edition), 4.20.6-200.fc29.x86_64, Storage for application: 11x 1TB HDD (ST1000NX0313) for Ceph OSD, Hadoop version: Hadoop 3.1.2, Alluxio version: Alluxio 2.0.0, Spark version: Spark 2.3.0, Hive version: Hive 3.1.1, Ceph version: Ceph 12.2.12

DRAM cache: 2 socket Intel® Xeon® Gold 6240 Processor, 18 cores HT On Turbo ON Total Memory 768 GB (24 slots/ 32GB/ 2666 MHz), BIOS: SE5C620.86B.0X.02.0094.102720191711 (ucode:0x500002c), Fedora 29 (Server Edition), 4.20.6-200.fc29.x86_64, Storage for application: 11x 1TB HDD (ST1000NX0313) for Ceph OSD, Hadoop version: Hadoop 3.1.2, Alluxio version: Alluxio 2.0.0, Spark version: Spark 2.3.0, Hive version: Hive 3.1.1, Ceph version: Ceph 12.2.12

Without cache: 2 socket Intel® Xeon® Gold 6240 Processor, 18 cores HT On Turbo ON Total Memory 192 GB (12 slots/ 16GB/ 2666 MHz), BIOS: SE5C620.86B.0X.02.0094.102720191711 (ucode:0x500002c), Fedora 29 (Server Edition), 4.20.6-200.fc29.x86_64, Storage for application: 11x 1TB HDD (ST1000NX0313) for Ceph OSD, Hadoop version: Hadoop 3.1.2, Alluxio version: Alluxio 2.0.0, Spark version: Spark 2.3.0, Hive version: Hive 3.1.1, Ceph version: Ceph 12.2.12

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors.

Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit www.intel.com/benchmarks.

Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates. See backup for configuration details. No product or component can be absolutely secure.

Your costs and results may vary.

Intel technologies may require enabled hardware, software or service activation.

©2020 Intel Corporation. Intel, the Intel logo, the Intel Inside logo, Intel Optane, and Xeon are trademarks of Intel Corporation or its subsidiaries.

* Other names and brands may be claimed as the property of others