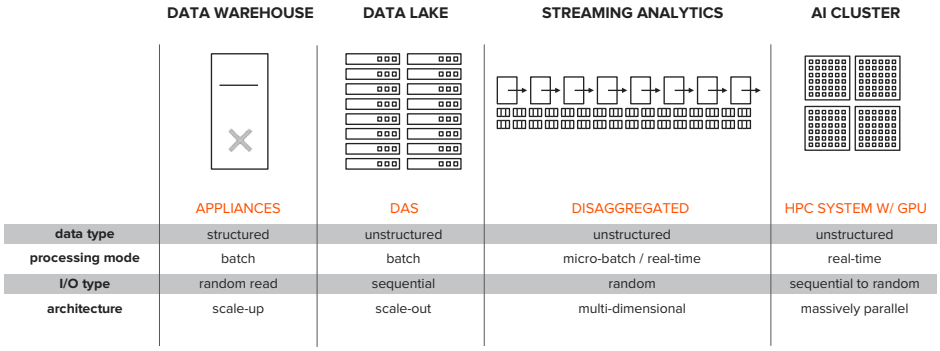
# DATA HUB: A MODERN STORAGE ARCHITECTURE


## SUMMARY

For today's enterprises, data is the fuel driving insight and innovation. Yet most data is stored in silos, out of reach of analytics and AI applications. The storage industry is largely to blame – for offering legacy solutions like data lakes and data silos that lock data away. It's time for the storage industry to step up and deliver a fresh architecture for the modern era.


## DATA LAKE & OTHER SILOS HOLD DATA BACK

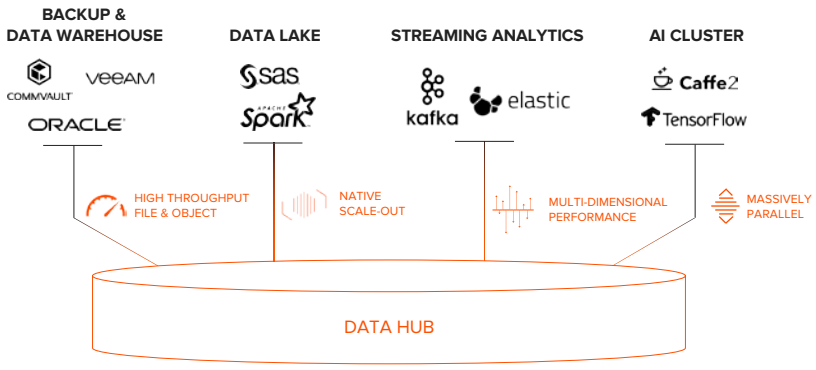
Data is fragmented. Some of it is stored in data warehouses, some is lost in data lakes. The unification of data is broken and the velocity of data crippled. So why is it so hard for storage systems to unify data on a single platform? The problem is that each application has different requirements for its data – thus the proliferation of data silos. It's time to rethink storage.



| | DATA WAREHOUSE | DATA LAKE | STREAMING ANALYTICS | AI CLUSTER |
|---|---|---|---|---|
| | APPLIANCES | DAS | DISAGGREGATED | HPC SYSTEM W/ GPU |
| **data type** | structured | unstructured | unstructured | unstructured |
| **processing mode** | batch | batch | micro-batch / real-time | real-time |
| **I/O type** | random read | sequential | random | sequential to random |
| **architecture** | scale-up | scale-out | multi-dimensional | massively parallel |

## A NEW ARCHITECTURE TO UNIFY AND SHARE DATA

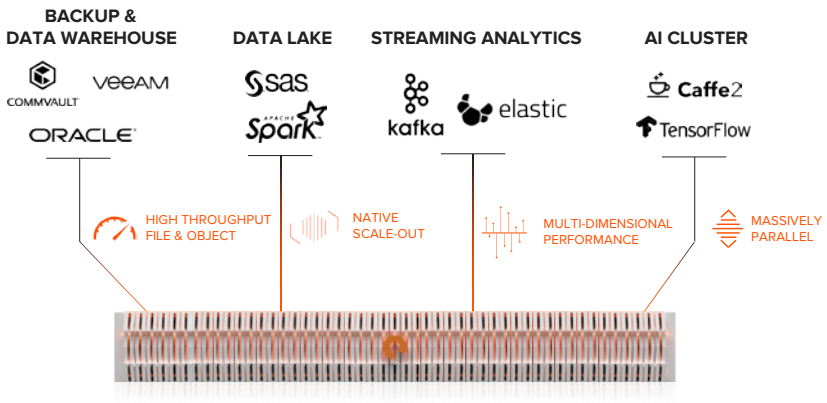
Data hub is a data-centric architecture for storage that powers analytics and AI. It enables enterprises to consolidate data silos and share data in today's rapidly-evolving, data-first world. Data hub takes the key strengths of each silo and integrates them into a single unified platform that includes four must-have qualities: high throughput file & object, native scale-out, multi-dimensional performance, and massively parallel architecture.



## FLASHBLADE: INDUSTRY'S FIRST DATA HUB

FlashBlade is built from the ground-up to consolidate all data-intensive applications by unifying file and object on a single scale-out platform. FlashBlade delivers on the four key features required for a data hub:

- **Natively built file and object protocols** to deliver the highest throughput for legacy and cloud-native applications
- **Scale-out architecture** to distribute work intelligently and deliver linear performance
- **Multi-dimensional performance** for any I/O to eliminate data bottlenecks
- **Massively parallel architecture** scaling to thousands of clients and billions of objects




## ISN'T DATA LAKE GOOD ENOUGH FOR ANALYTICS?

Data lake and data hub are vastly different at their core. Data lake is designed to store data as efficiently as possible. Data hub is designed to share and deliver data. The former is engineered with legacy technologies like DAS-based storage. The latter is a modern system based on an all-flash, cloud-like disaggregated architecture. In the new era of analytics, where data is needed on demand, data hub is the right platform for enterprises.

| | DAS DATA LAKE | DATA HUB |
|---|---|---|
| PURPOSE | STORE DATA | SHARE DATA |
| DATA ACCESS | LARGE, SEQUENTIAL | SMALL TO LARGE, RANDOM TO SEQUENTIAL |
| WORKLOAD | BATCH | REAL-TIME & BATCH |
| INFRASTRUCTURE | STATIC | ON-DEMAND |
| HARDWARE | PHYSICAL | VMs, CONTAINERS, PHYSICAL |
| STORAGE MEDIUM | DISK | FLASH |
| MEASURE OF VALUE | $/GB | BETTER MODELS, FASTER INSIGHT |