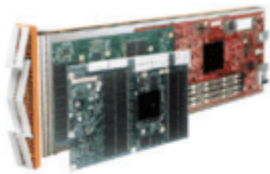


FlashBlade and Spark for 6x Faster Queries



BLADE

SCALE-OUT DIRECTFLASH + COMPUTE

Ultra-low latency, 8 & 52TB capacity options that can be hot-plugged into the system for expansion and performance



PURITY

SCALE-OUT STORAGE SOFTWARE

The heart of FlashBlade, implementing its scale-out storage capabilities, services and management



FABRIC

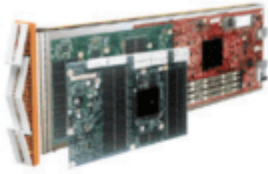
LOW-LATENCY, SOFTWARE-DEFINED NETWORKING

Includes a built in 40Gb Ethernet fabric providing a total network bandwidth of 320Gb/s for the chassis

Spark revolutionized large-scale data processing. The value it provides includes:

- 100x faster than Hadoop MapReduce
- Enabling applications to be written in Java, Scala, Python, or R
- Combine SQL, streaming, and complex analytics to be run on the same stack.
- Run it standalone, on Hadoop and Mesos with data stores from HDFS, Cassandra, HBase, and S3

[FlashBlade®](#) is a true cloud-scale, big-data storage platform that provides consumers with a platform to handle the concurrency required to accelerate typical big data processing workloads. These include:



BLADE

SCALE-OUT DIRECTFLASH + COMPUTE

Ultra-low latency, 8 & 52TB capacity options that can be hot-plugged into the system for expansion and performance

PURITY

SCALE-OUT STORAGE SOFTWARE

The heart of FlashBlade, implementing its scale-out storage capabilities, services and management

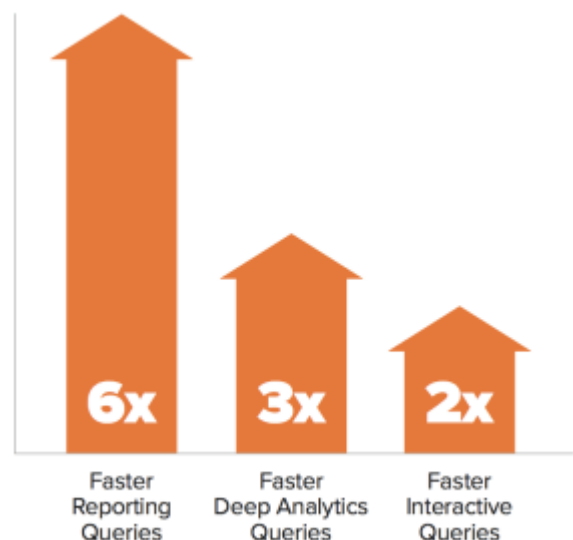
FABRIC

LOW-LATENCY, SOFTWARE-DEFINED NETWORKING

Includes a built in 40Gb Ethernet fabric providing a total network bandwidth of 320Gb/s for the chassis

- Simple and easy to deploy and manage
- Provides NFSv3 and S3/Object
- Centralized cloud management via Pure1
- 17GB/s at 1,000,000 operations per second
- Consistent linear scaling through metadata hyper-partitioning across blades
- Scale from 98TB to 1.6PB in 4RU with scalability to 10s of PB with multi-chassis configurations (coming soon)
- Linear scaling per blade without disruption or downtime
- Low latency software define networking

Pairing FlashBlade and Spark



- Ultra low consistent latency for all queries
- Consolidate data across multiple Hadoop clusters to leverage all compute on all capacity
- Ability to scale compute and storage separately
- 6x faster reporting queries
- 3x faster deep analytics queries
- 2x faster interactive queries

- Use any orchestration/file format (Mesos, Kubernetes, Parquet, Hadoop Yarn, and more)
- Rest API enabling integration into any tool or custom script (coming soon)

If you are running Spark, you should check out our [whitepaper](#) on FlashBlade and Spark.