

THE DATA LIFECYCLE FROM BYTES TO AI

How to Build a Data Pipeline for Real-World AI

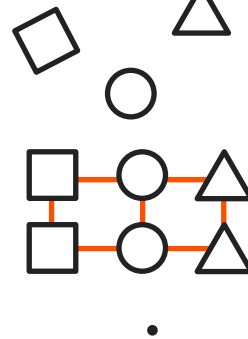


INGEST

From sensors, machines
& user generated

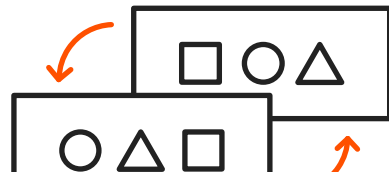
CLEAN & TRANSFORM

Label, anomaly detection,
ETL, prep, stage



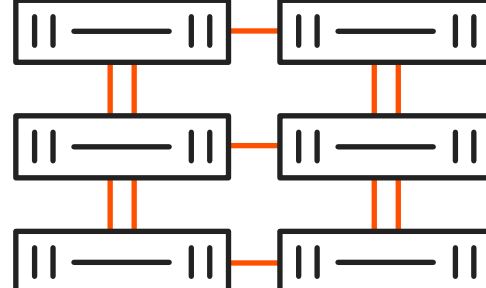
EXPLORE

Quickly iterate to converge on models



TRAIN

Run for hours to days
in production cluster
of GPU servers



INFRASTRUCTURE OPTIMIZED FOR AI DATA PIPELINE

Varying needs in AI data pipeline push
the limits of storage architecture.

	INGEST	CLEAN & TRANSFORM	EXPLORE	TRAIN
ACCESS PATTERN	sequential	sequential or random	random	random
ACCESS TYPE	write	read & write	read	read
FILE SIZE	small to large	small to large	small to large	mostly small
CONCURRENCY	depends on #of sources	high	low	high

WHY FLASHBLADE?

A centralized data hub increases the productivity of scientists and simplifies the pipeline for the data architect. FlashBlade™ is the industry's first data hub purpose-built for AI for the following reasons:



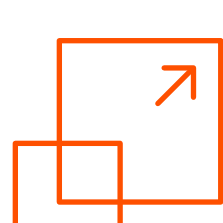
PERFORMANCE

With up to 75GB/s random read bandwidth, FlashBlade can support the entire pipeline at the same time.



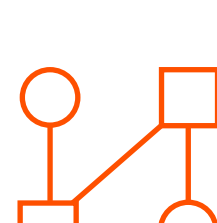
SMALL-FILE HANDLING

Read small files (50KB) at 50 GB/s with 75 blades for the most demanding training workloads.



SCALABILITY

Increase capacity and performance as training datasets grow, all without downtime.



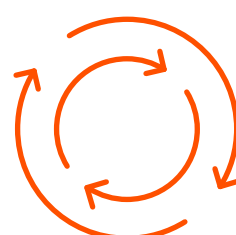
NATIVE OBJECT SUPPORT (S3)

Input data can be stored as either files or objects.



SIMPLE ADMINISTRATION

No need to tune performance for small or large files, sequential or random access.



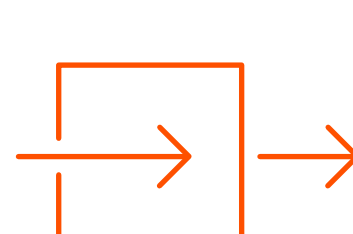
NON-DISRUPTIVE UPGRADE (NDU) EVERYTHING

Software upgrades and hardware expansion can happen anytime, even during production model training.



EASE OF MANAGEMENT

Pure1® cloud-based management keeps users focused on understanding data vs. administering storage.



BUILT FOR THE FUTURE

Purpose-built for flash to easily leverage new generations of NAND technology.

ITERATE FASTER!

In today's world, it's critical to have infrastructure that supports both massive data ingest and rapid analytics evolution. At Pure Storage, we built the ultimate data hub for AI, engineered to accelerate every stage of the data pipeline.

Visit purestorage.com/flashblade to learn how FlashBlade™ can help you transform big data into big intelligence.