



# **DDN LEVERAGES YEARS OF HPC KNOWLEDGE TO FUSE INGREDIENTS OF AI THANKS TO A3I INTO TRADITIONAL HPC WORKLOADS**

6<sup>th</sup> March 2019

Gautier Soubrane  
[gsoubrane@ddn.com](mailto:gsoubrane@ddn.com)



# A3I

ACCELERATED, ANY-SCALE AI SOLUTIONS FROM DDN

**Extract** a Lot More Answers per Hour From  
Your Data **at Any Scale**

# DDN LEADS ARTIFICIAL INTELLIGENCE AND DEEP LEARNING **AT SCALE**



**AUTONOMOUS  
VEHICLES**



**LIFE SCIENCES  
RESEARCH**



**PRECISION  
HEALTHCARE**



**SMART MANUFACTURING  
ROBOTICS**



**FINANCIAL SERVICES  
FRAUD DETECTION**



**SECURITY &  
INTELLIGENCE**



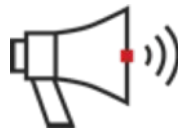
**PREDICTIVE  
ANALYTICS**



**SUPPLY CHAIN  
LOGISTICS**



**AUGMENTED  
REALITY**



**PERSONALIZED  
MARKETING**



**CONSUMER  
RETAIL**



**LANGUAGE  
PROCESSING**

# DDN A<sup>3</sup>I SOLUTIONS FOR END-TO-END AI ENABLEMENT

---

## Accelerate AI performance

Achieve full GPU saturation for greatest AI productivity

## Streamline DL workflows

Engage all workloads concurrently and continuously

## Maximize efficiency

Profit from greatest technical and economic benefits

## Scale seamlessly

Increase performance, capacity and capability at any scale

## Ensure data availability

Trust in robust data protection, integrity and access

## Deploy and manage easily

Fully-integrated and optimized solutions from DDN

# DDN COMPLETE AI DATA LIFECYCLE SOLUTIONS

**PICK YOUR COMPUTE  
NODES, YOUR AI WORKFLOW  
CONCURRENTLY,  
CONTINUOUSLY,  
IN-PLACE**

DDN Parallel AI solutions are all-round performers able to saturate GPUs for training at the same time as inference and data movements and preparation

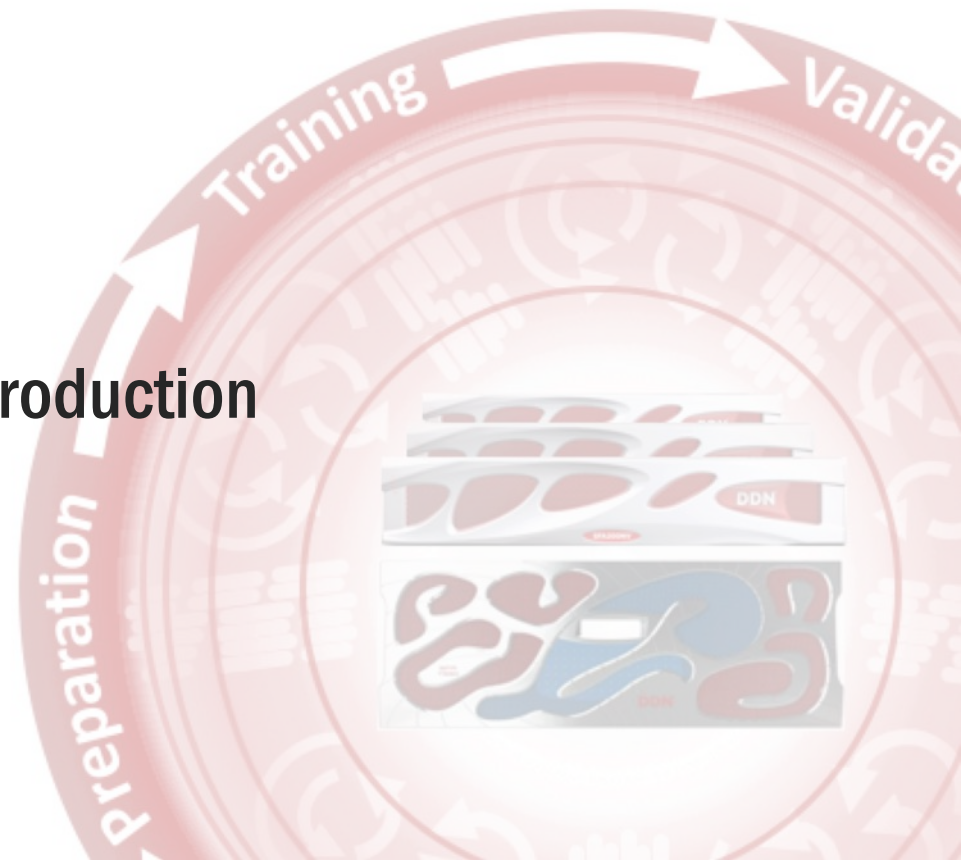




# DDN ENABLES AND ACCELERATES DEEP LEARNING

**Data is the new source code**

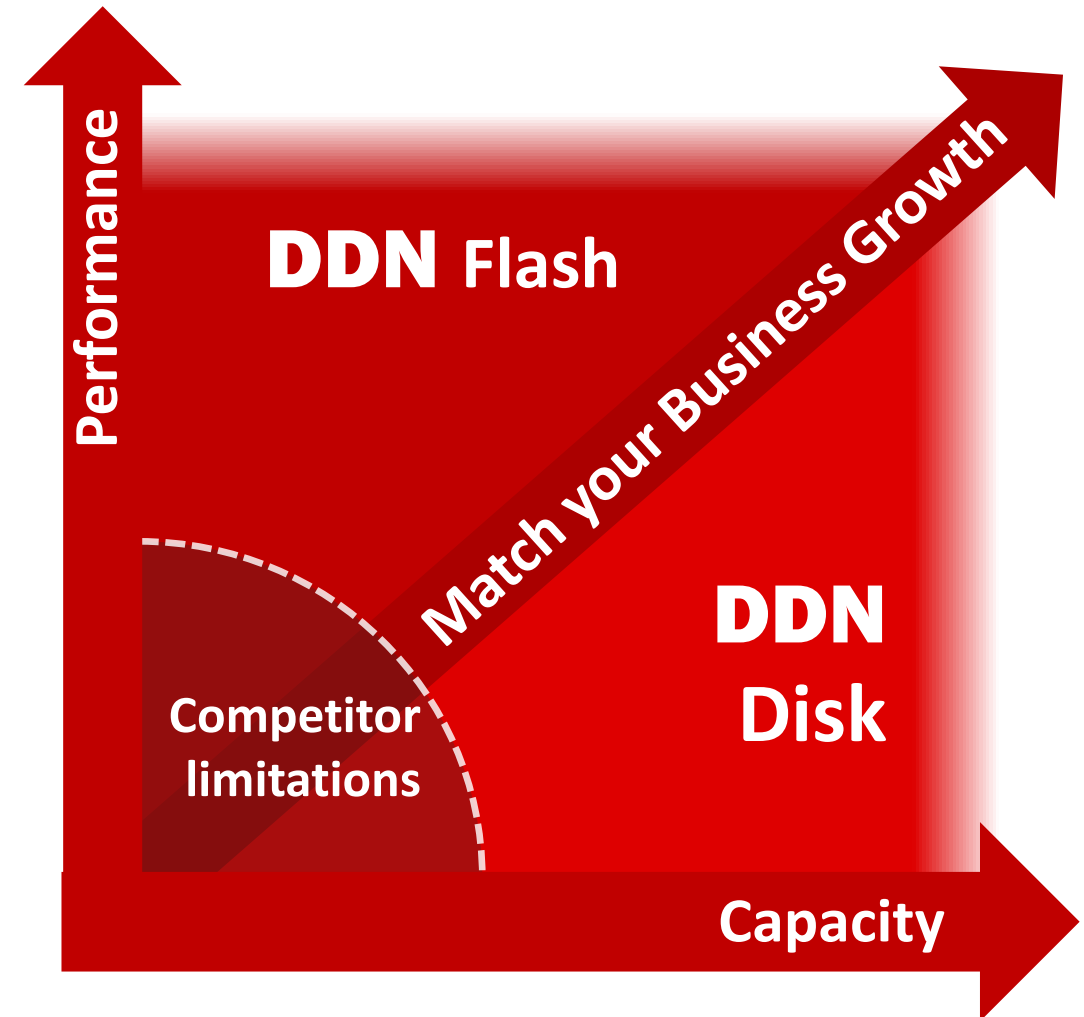
- ▶ **Collect and access** large amounts of data from a variety of sources
- ▶ **Easy data discovery**
- ▶ **Rapid iteration** of model training and refreshing
- ▶ **Quick transition** of new and refreshed models to production
- ▶ **At scale, concurrently, continuously**



# DDN FLEXIBLE SCALING

## Match your business growth with flexible all flash or hybrid scale-out

With data volumes continuing to rise, AI is getting smarter and learning faster, enabling organizations such as yours to leverage more-advanced Machine Learning (ML) and Deep Learning (DL) solutions.



# A<sup>3</sup>I

ACCELERATED ANY-SCALE AI SOLUTIONS FROM DDN

## Optimized and integrated data solutions for your AI frameworks

---

DDN brings you to production faster by creating a fully optimized and integrated data platform for AI frameworks, managing data efficiently whatever your choice of AI framework





# DDN PARALLEL FILE STORAGE APPLIANCES FOR AI AND DL



## AI200

All NVME parallel file storage appliance  
Optimized for the most intensive workloads  
30TB, 60TB, 120TB in 2RU



## AI7990

Hybrid parallel file storage appliance  
Optimized for capacity, intermix with flash  
1PB in 4RU

**20GB/s, 350K IOPS • 100Gb Ethernet or EDR InfiniBand • Start with 1, Scale out and mix**

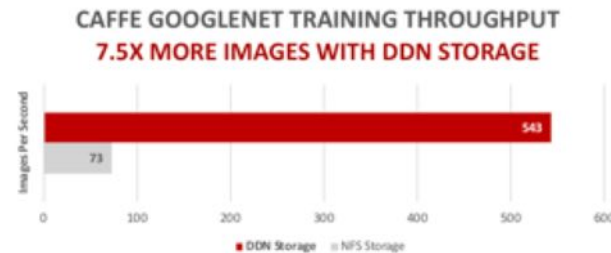
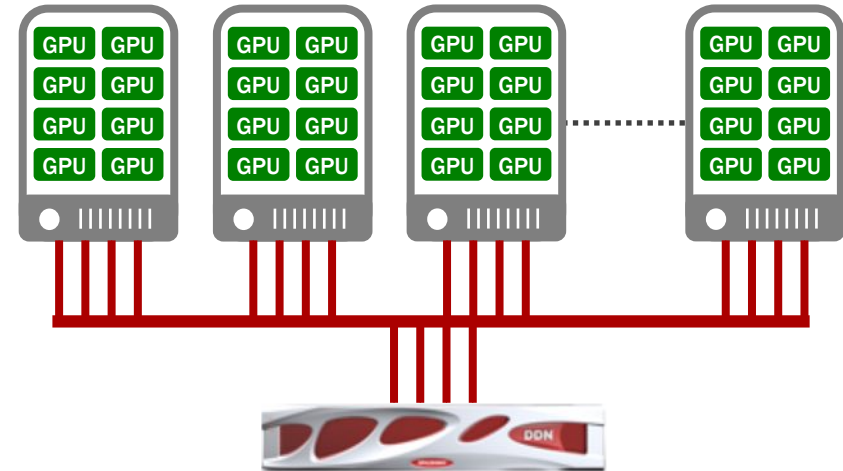
# DDN SHARED PARALLEL ARCHITECTURE AND PROTOCOL

## Superior performance, scalability and reliability

GPU saturation is critical for achieving the full potential of AI and DL.

The DDN parallel architecture and protocol delivers high-throughput, low-latency, massive concurrency to applications.

For DL frameworks, provides increased performance and faster processing than NFS.



# DDN AI SOLUTIONS FOR AUTONOMOUS VEHICLES

**DDN enables the largest autonomous vehicle programs in the world**



## FLEET OPERATIONS

Vehicle data generation  
internal and external sensors

Fleet data collection  
and processing

## TRAINING

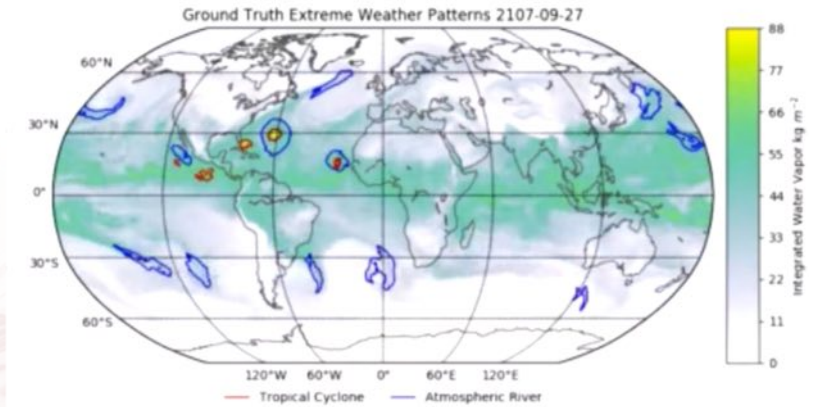
Data curation and indexing  
Model development, selection  
Discovery

## SIMULATION

Safety  
Resimulation  
Regression testing

# DDN AI SOLUTIONS CLIMATE AND WEATHER SIMULATION

**Traditional HPC simulation leveraging  
AI/ML approached**



## FUSION

Ground sensor data collection  
Segmenting, pattern labelling

## TRAINING

Data curation and indexing  
Model development, selection  
Discovery

## SIMULATION

HPC processing  
Interpolation  
Regression testing

# **DDN AI SOLUTIONS FOR LIFE SCIENCES AND HEALTHCARE**

**DDN accelerates life sciences and healthcare hundred fold with GPU-enabled genome and image analytics.**



## **CORE RESEARCH**

Genomics and  
microscopy

Biostatistics  
and bioinformatics

## **DRUG DISCOVERY**

Pharmacology and  
pharmacotherapy

Clinical trials

Post-market drug safety

## **PATIENT CARE**

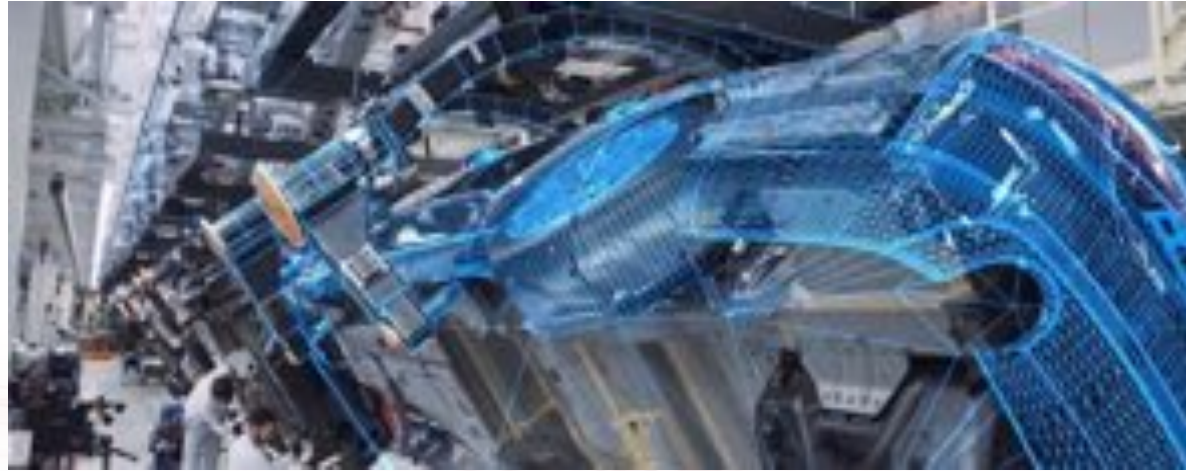
Diagnostics and  
precision medicine

Predictive health

Insurance



# **DDN AI SOLUTIONS FOR SMART MANUFACTURING**



## **ROBOTICS**

Collaborative teams  
Self-training and repair  
Agility and adaptability

## **QUALITY CONTROL**

Industrial inspection  
Better anomaly detection  
and fewer false positives  
Failure analysis and reporting

## **PLANT EFFICIENCY**

Raw materials ordering  
just in time delivery  
Operations scheduling  
Predictive maintenance



# DDN AI SOLUTIONS FOR FINANCIAL SERVICES AND ANALYTICS



# ANALYTICS

## Market prediction

## Contract intelligence

## Organizational performance

# BANKING

## Investment and asset management

# Creditworthiness

## Customer engagement

# SECURITY

## Risk management

## Fraud detection

## Regulatory compliance

# DDN AI SOLUTIONS FOR CONSUMER EXPERIENCES



## RETAIL

Cashier-free checkout  
Personalized services  
Inventory and logistics

## ADVERTISING

Personalized marketing  
Marketing forecasting

## ENTERTAINMENT

Experience innovation: AR,VR  
Recommendation engines  
Content protection

# HPC simulation accelerated

## 1. Macro simulation

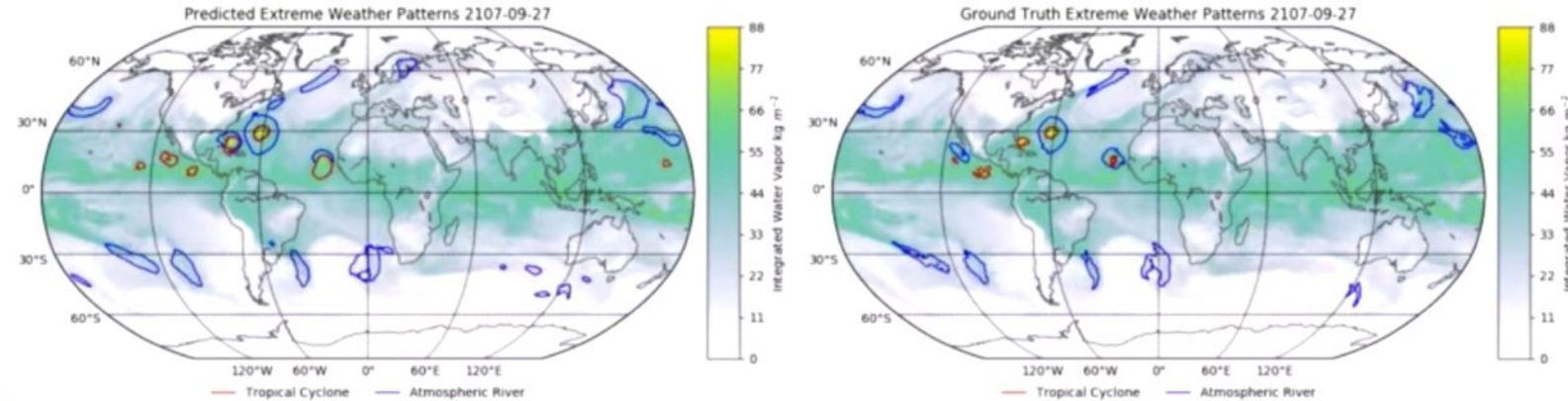
Using traditional methods

## 2. Interpolation

Inference spatial interpolation

## 3. Iteration loop

Ground truth confrontation and minimum error heuristic



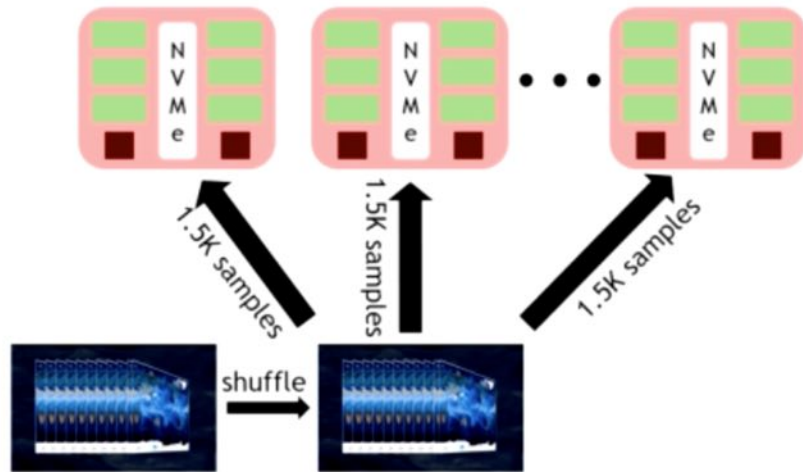
Courtesy from Michael Houston SC18 conference  
NVIDIA: Exascale Deep Learning for Climate Analytics

- Harness the ever growing computation needs
- Allow faster time to insight
- Remains still highly demanding on data access
- Smart staging and closely connected hot data storage required



# Reaching scale remains a challenge

Dataset Size	Required BW (27K GPUs)	GPFS/LUSTRE	BurstBuffer	NVMe or DRAM
20 TB (~63K samples)	3.8 TB/s	~400 GB/s	~2 TB/s	~26 TB/s



- 250 training samples/GPU (~15 GB), sample w/ replacement
- Each file will be read at most once from FS
- Files shared between nodes via MPI (mpi4py)

NVIDIA

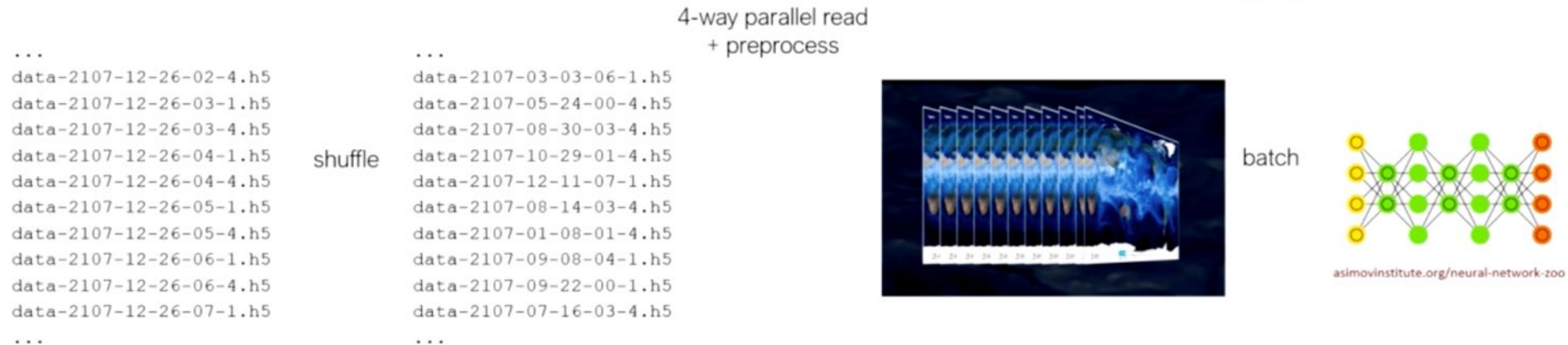
Courtesy from Michael Houston SC18 conference  
NVIDIA: Exascale Deep Learning for Climate Analytics

- Using local cache is a smart approach yet requires tradeoff on the ability to update and synchronize parameters
- Current R&D focus on a smart LPCC Lustre Persistent Client Cache - Enabling client side cache

# Reaching scale remains a challenge

- DDN approach to remove limits to scaling greatly improve performances
- Higher compute density is a real opportunity for DDN and its ability to feed data locally
- Staging the data you need to process is key (pseudo random brings efficiency)
- The relevance of IO benchmark is extremely relevant

Eg: <https://news.developer.nvidia.com/nvidia-dgx-2-accelerates-weather-forecasting/>



# A3I™

WITH NVIDIA® DGX-1™







**THANK YOU**