



Architect of an Open World™

## Data in the world of FLOPS

Jacek L. Madajczyk, Extreme Computing Manager  
Central and Eastern Europe

**LIBERATE IT**

# Agenda

- HPC Storage World – a case study
- HPC Storage Ecosystem – a Bull's perspective
  - High Performance Storage system
  - Parallel File System
  - HSM
- Trends in HPC Storage
- Q/A

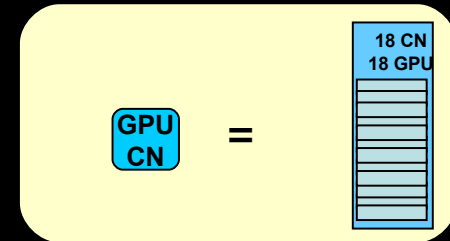
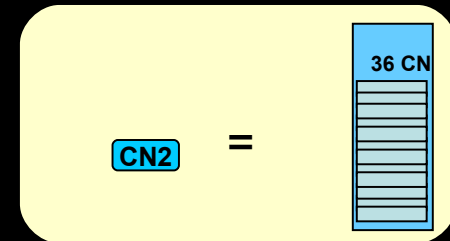
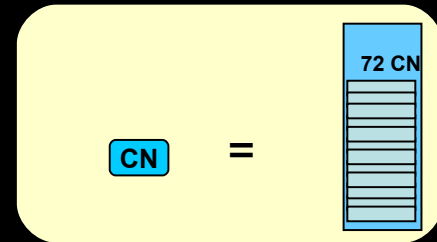
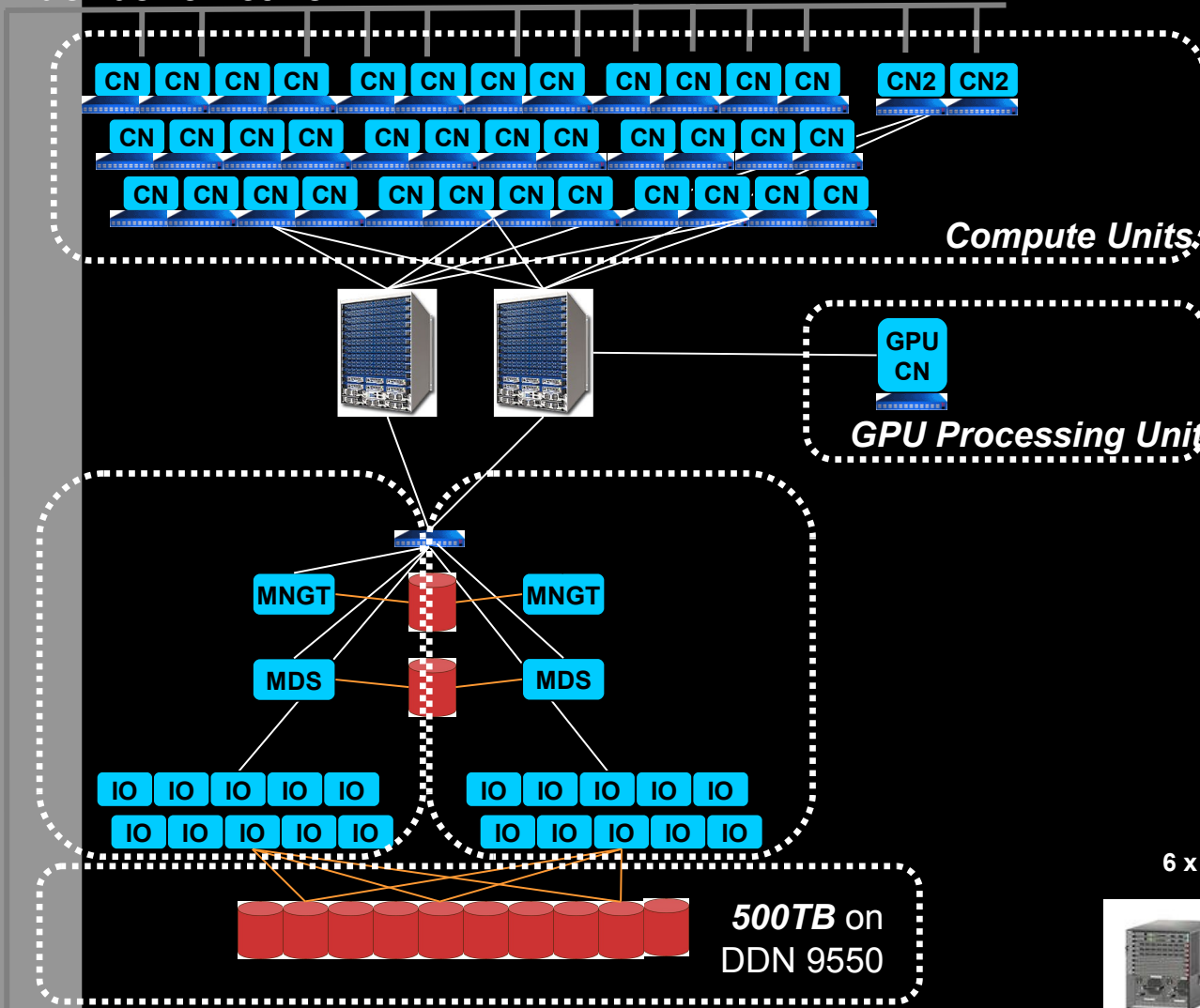
# *HPC Storage World – a case study*

**LIBERATE IT**



# GENCI-CCRT Cluster Global Architecture

## Backbone network



6 x 10Gb Ethernet links  
For Backbone



3x CISCO 6509



*HPC Storage Ecosystem – a Bull's perspective*

**LIBERATE IT**



# Which Storage for a Parallel File System

- **A performant storage subsystem**

  - A good storage subsystem with sequential IOs in Read and Write operations ( 1 GB/s to 10 GB/s)**

  - A good ratio MB/s per Terabyte ( max performance for a minimal number of disks)**

  - The storage subsystems have to be big enough and modular to be coherent with the parallel filesystem ( reducing the complexity).**

- **A dense storage subsystem**

  - A good ratio performance / m2 used (density of the controllers, dense disk drawers )**

  - A controlled and quantified power consumption**

- **A reliable storage subsystem**

  - A storage subsystem without « SPOF »**

  - Redundant Controllers, batteries, saved cache memory**

  - RAID disk groups , data control mechanism, Hot spare disks**

# *Storage and Parallel File Systems*

**LIBERATE IT**



# High Capacity and Access Mode

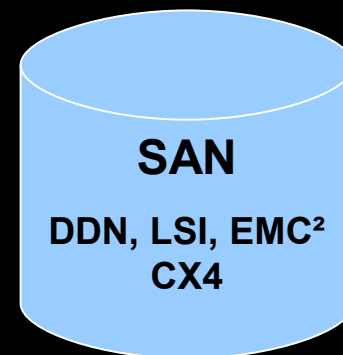
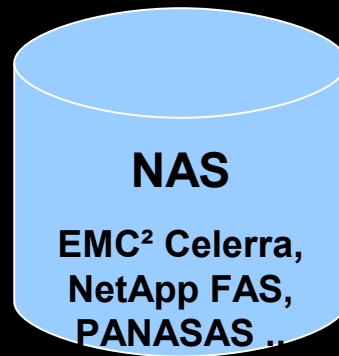
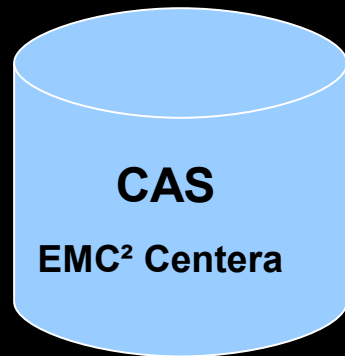
Performances



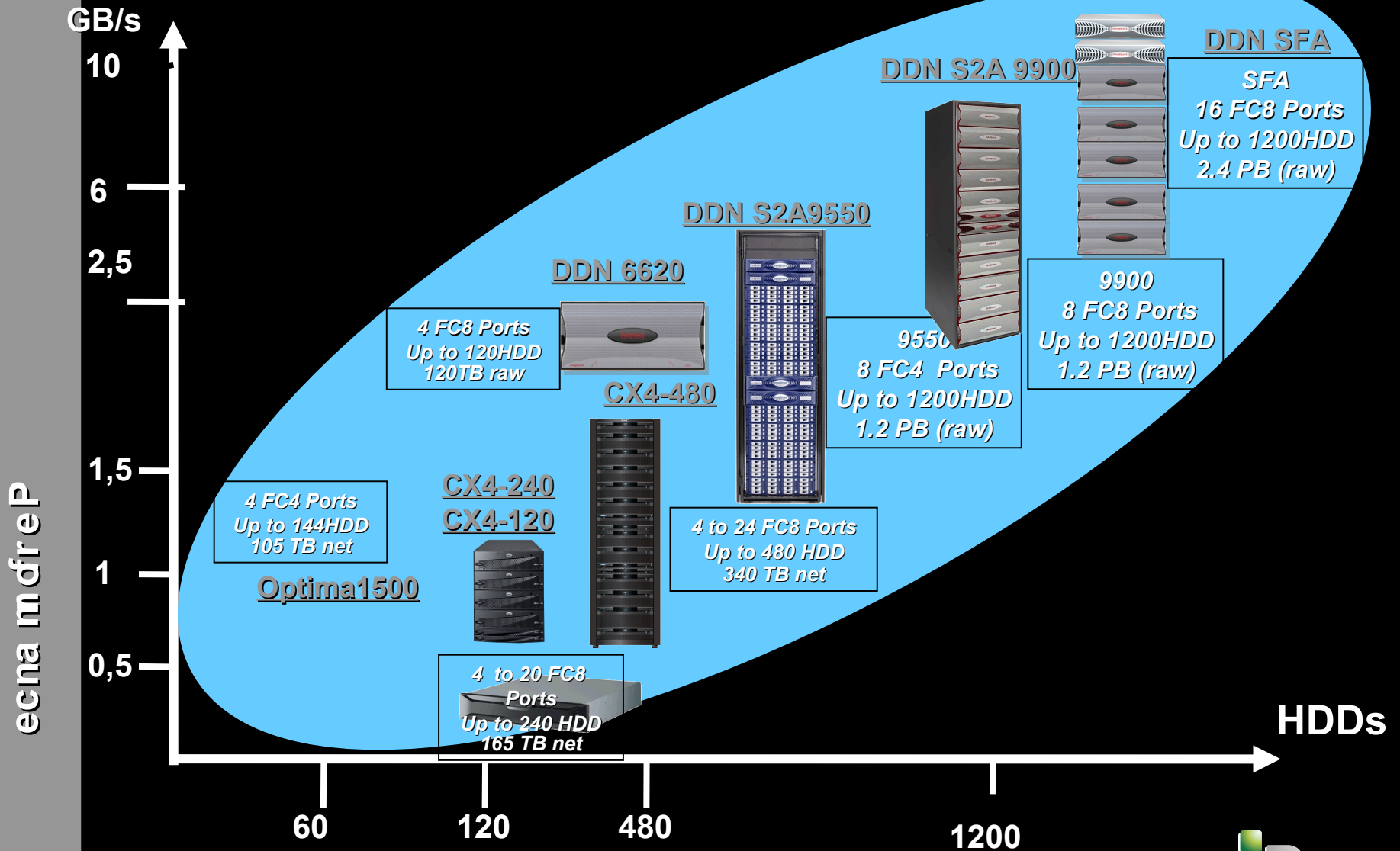
**CAS Storage**  
API upload,  
download

**File mode  
Storage**  
NFS, CIFS

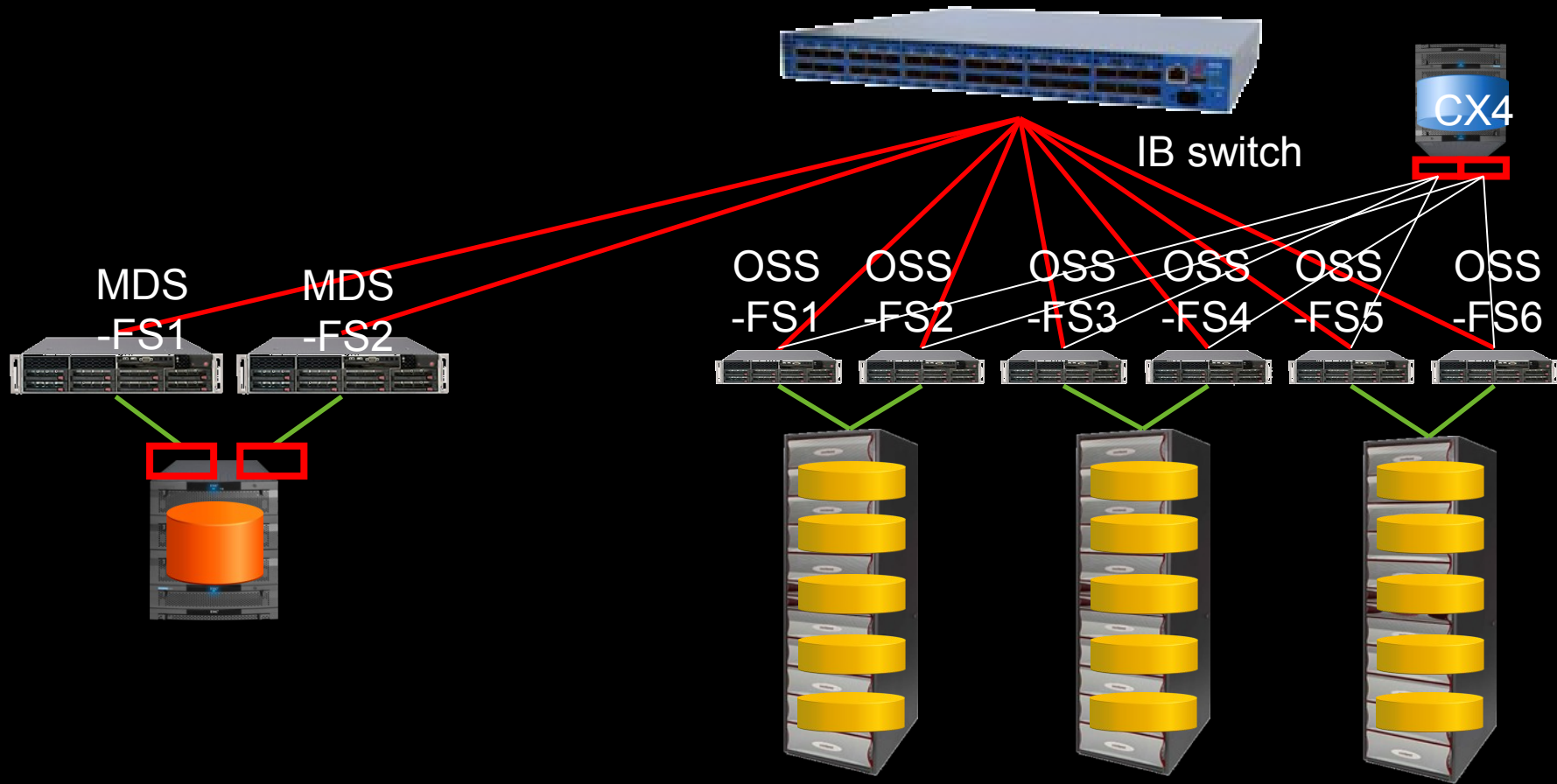
**Block mode  
Storage**  
Parallel File System  
(Lustre, GPFS...)



# Bull HPC Storage



# EMC CX4 and DDN 9900 High Availability system



Add 1 CX4 for the journal storage for every bunch of 3 IO cells (DDN9900)

# Worldwide references in a variety of sectors

Educ/Research	Industry	Aerospace/Defence	Other

... and many others



## High capacity storage systems features

IOs for high capacity storage systems require a high sequential throughput thus an optimal use of the storage hardware.  
The high capacity storage systems are characterized by:

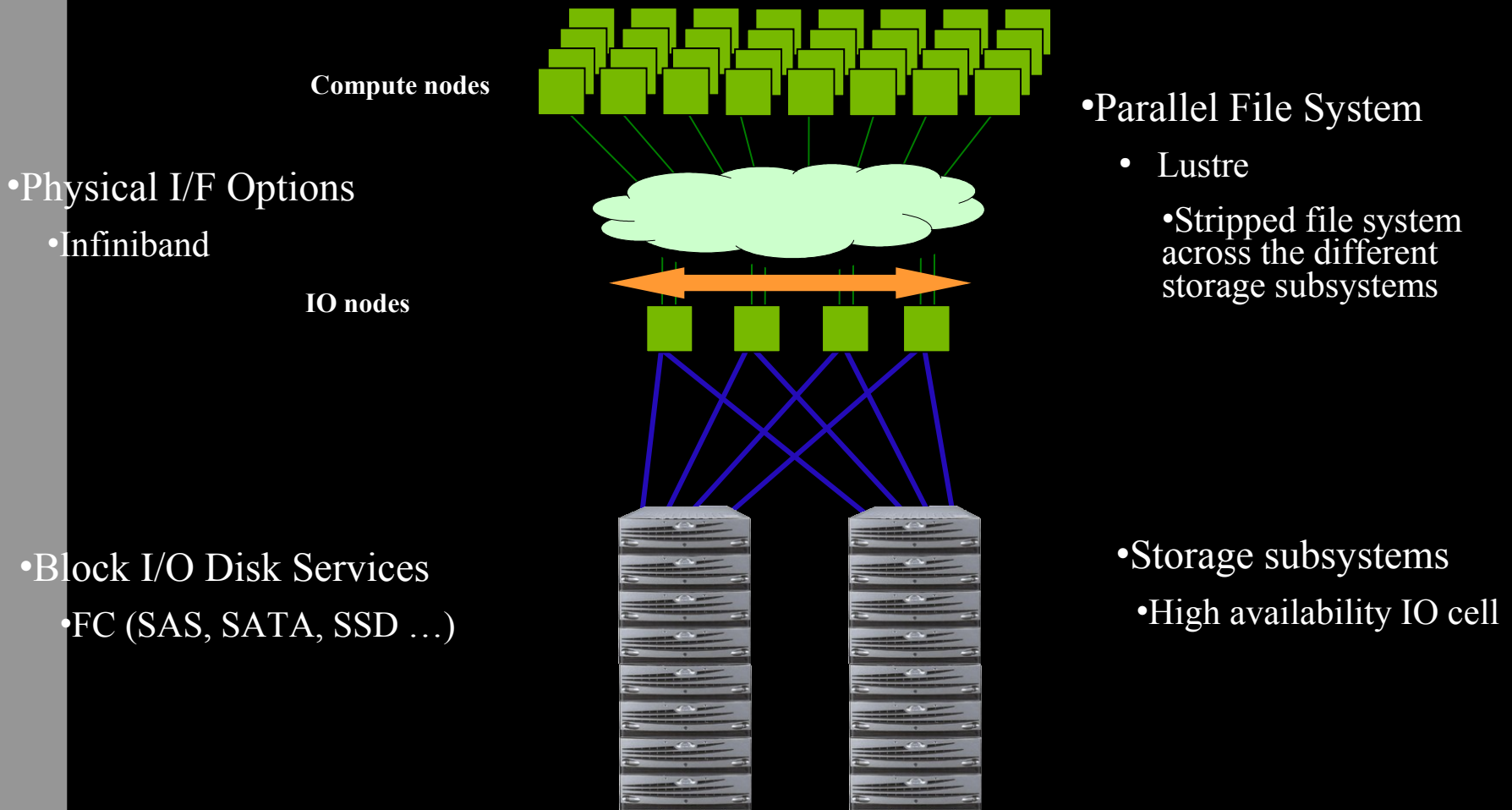
- High performance storage array or multiple performant storage arrays
- Large disk space ( terabytes to Petabytes) with a global namespace for end users
- A performant data communication infrastructure ( sometimes complex)

The parallel filesystems fulfill these requirements

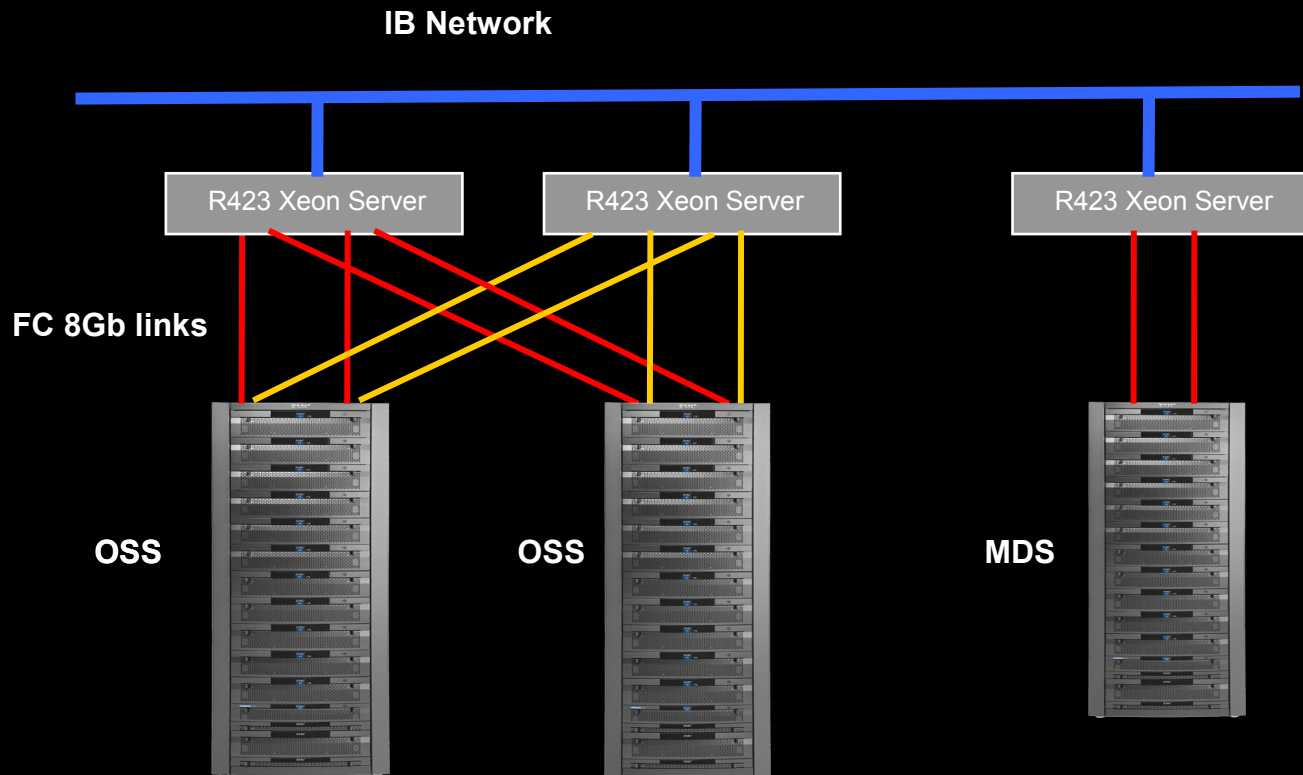
# Parallel File Systems

- Performance scaling linearly with the addition of storage subsystems under the same filesystem.
- File system performance close to the raw maximal performance of the storage subsystems (very few loss through the software layers).
- Ability to use different storage subsystems under a single filesystem
- Optimized management of multiple clients accessing the data.

# Cluster Architecture with a Parallel File System



# Lustre I/O cell



OSS: Lustre Object storage server

MDS: Lustre MetaData server

# *Trends in HPC Storage*

**LIBERATE IT**

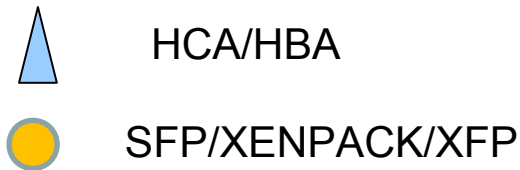
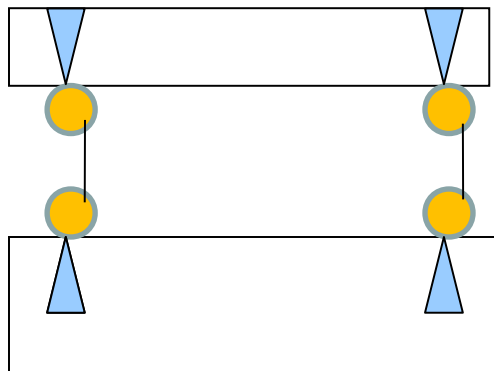


# Trends in HPC storage

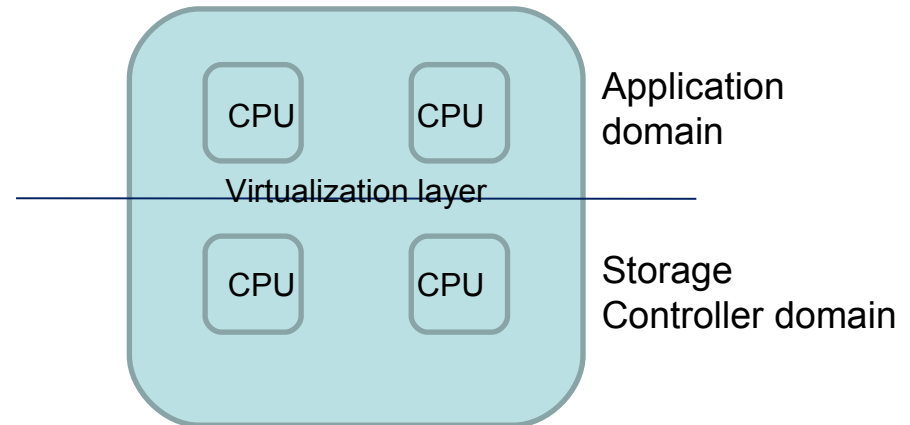
- New controllers architecture with the multi CPU and multi cores: A software Raid controller is running on one CPU and an Application server ( Lustre OSS, GPFS, NFS server ...) is running on another CPU.
- New level of cache with SSDs in the controller to boost the performances with IO burst

# Data Direct Networks Janus Appliance architecture

## Today



## Janus



- All I/O Node to Storage Controller interconnect components eliminated.
- Protocol overhead and conversions reduced
- Shared memory space
- High-speed internal connections

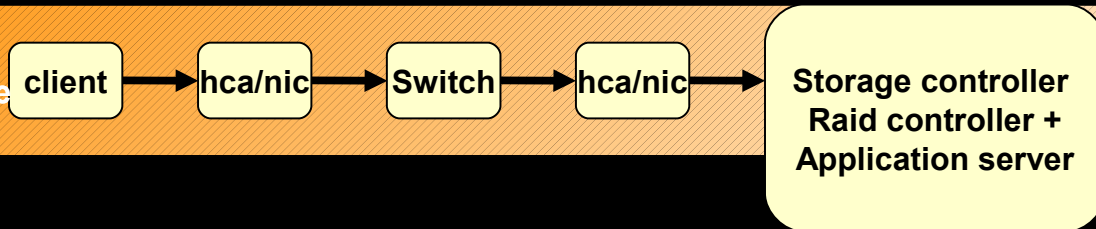
# IO paths with application server in the controller

Latency

Traditionnal



Application  
Server in the  
controller



**Less protocol conversion**  
**Reduction of the latency**  
**More IOPS**

# Trends in HPC storage – future HSM over Lustre project

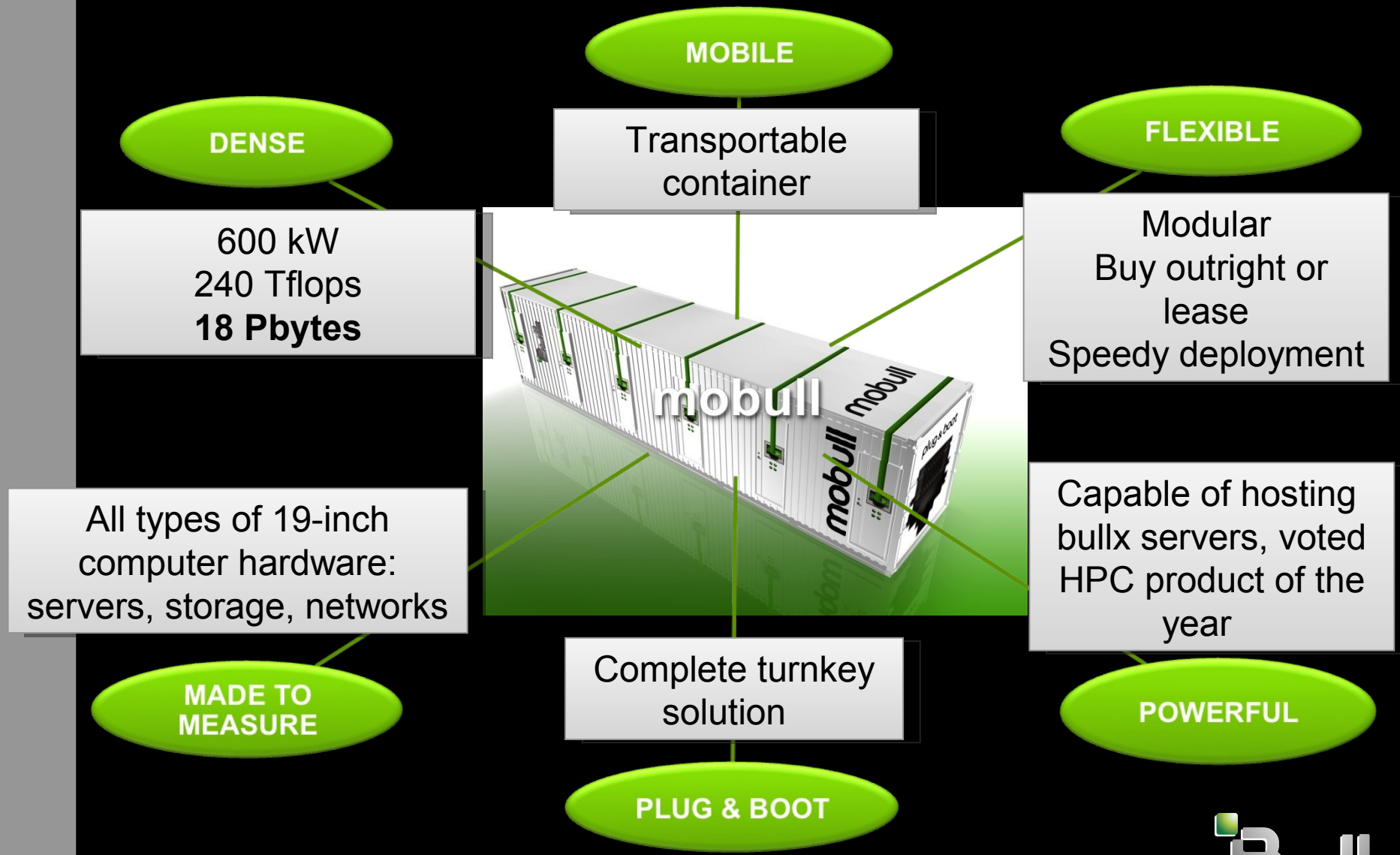
- Best of two worlds

**Lustre filesystem – size from some TB to few PB**

**External HSM – size from lots of TB to many PB**

- Add a high performance and scalable access to HSM
- Extend Lustre disk space in economical way
- Backup Lustre data in a resilient and stable archiving system

# Trends in HPC storage – Mobile Data Center





Architect of an Open World™

LIBERATE IT

