ЛАБОРАТОРИЯ
ИНФОРМАЦИОННЫХ
ТЕХНОЛОГИЙ
имени М.Г. Мещерякова

Distributed system for processing and data storage for experiments at the Complex NICA

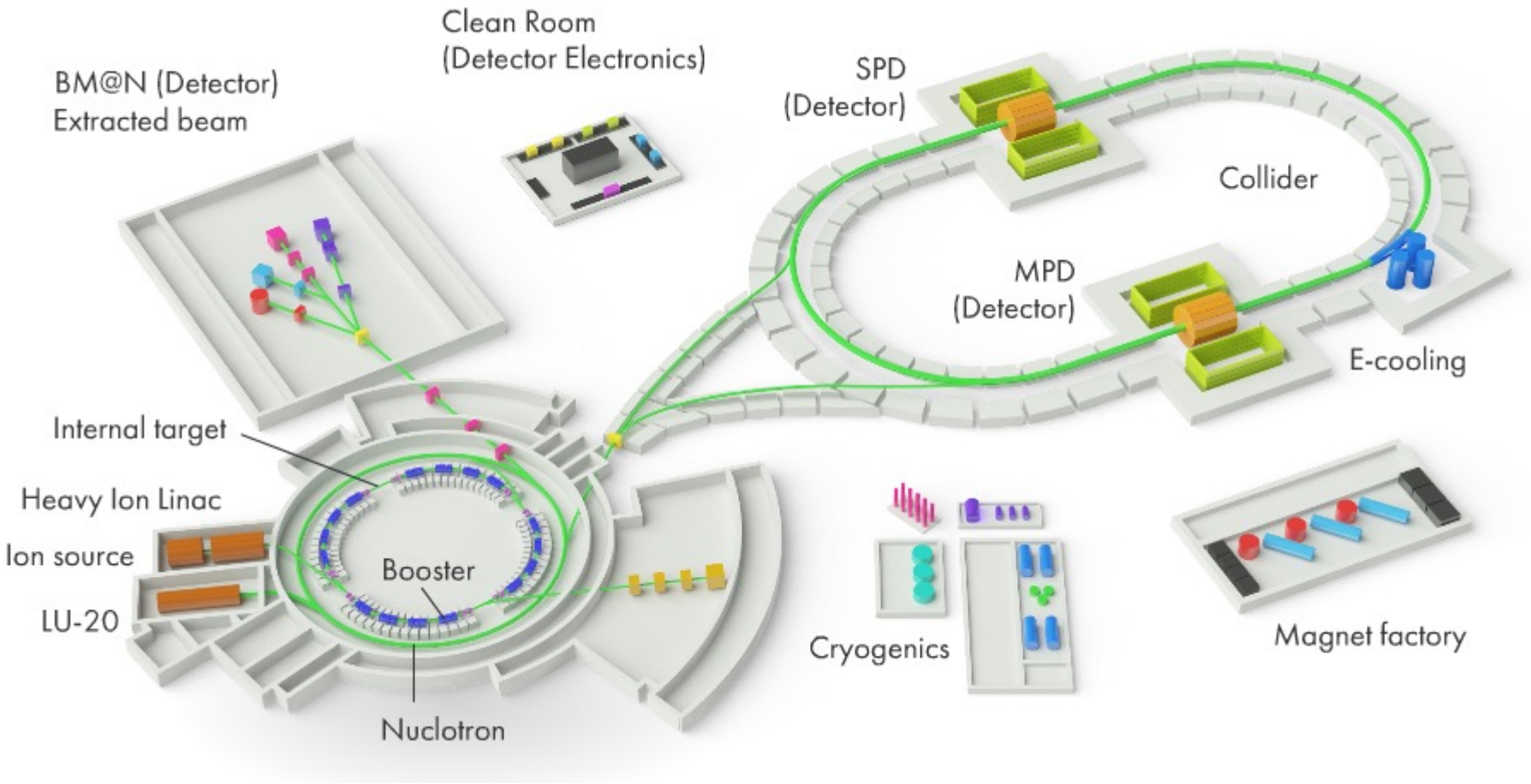
Belyakov D.V., Dolbilov A.G., Kokorev A.A., Lyubimova M.A., Matveev M.A., Moshkin A.A., Podgainy D.V., Rogachevsky O.V., Slepov I.P., Zuev M.I.

Joint Institute for Nuclear Research,
Dubna, Russia

Russian Supercomputing Days 2023
International Scientific Conference

September 25, 2023

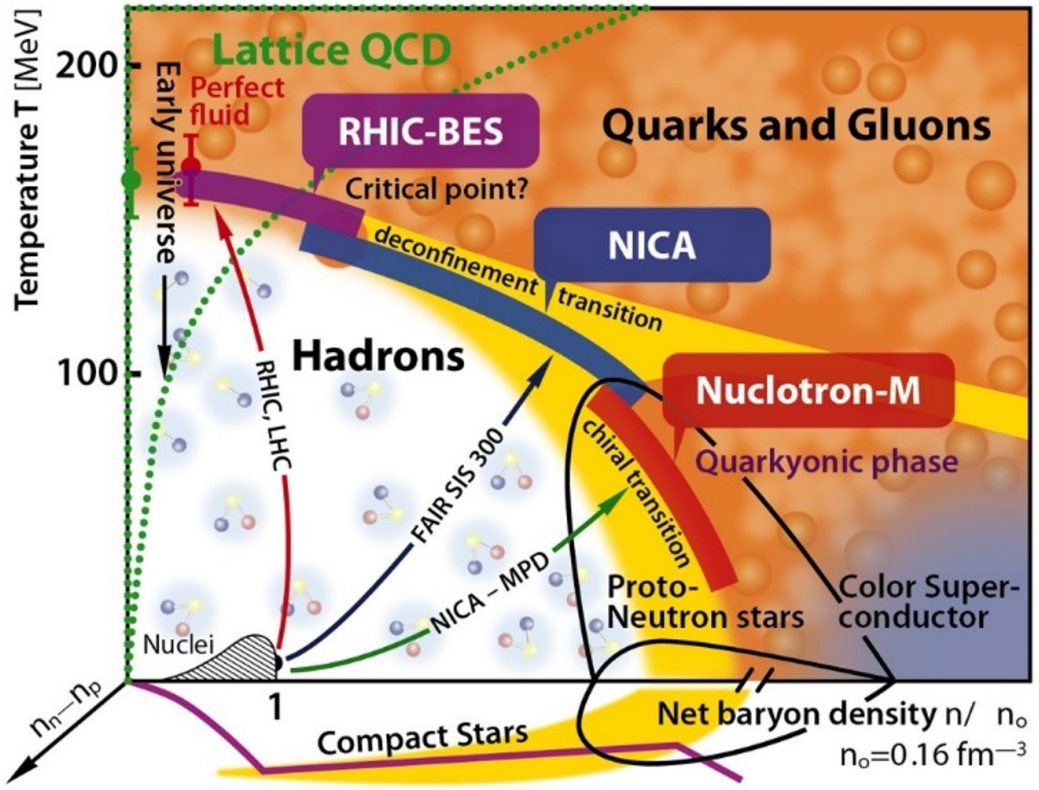
Complex «NICA» (Nuclotron-based Ion Collider fAcility)



The Complex NICA is being created at Joint Institute for Nuclear Research to study the properties of superdense baryonic matter.

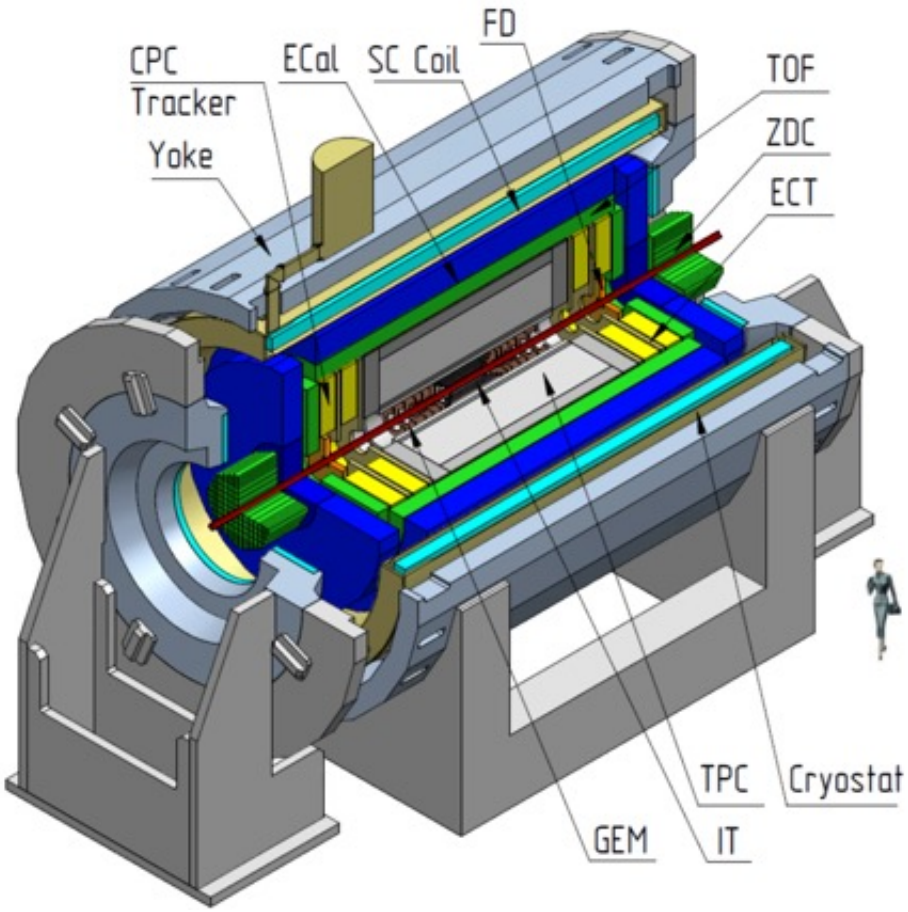
Complex «NICA»
Phase Diagram of Hadron Matter

The most important fundamental problems in this area of physics are:



- The nature and properties of Strong interactions between the elementary components of the Standard Model of particle physics (quarks and gluons)
- Search for signs of phase transition between hadronic matter and QGP
- Search for new phases of baryonic matter
- Study of the basic properties of vacuum of the Strong interactions and QCD symmetries

Complex «NICA»
Multi-Purpose Detector (MPD)



Data Acquisition System

Raw DATA

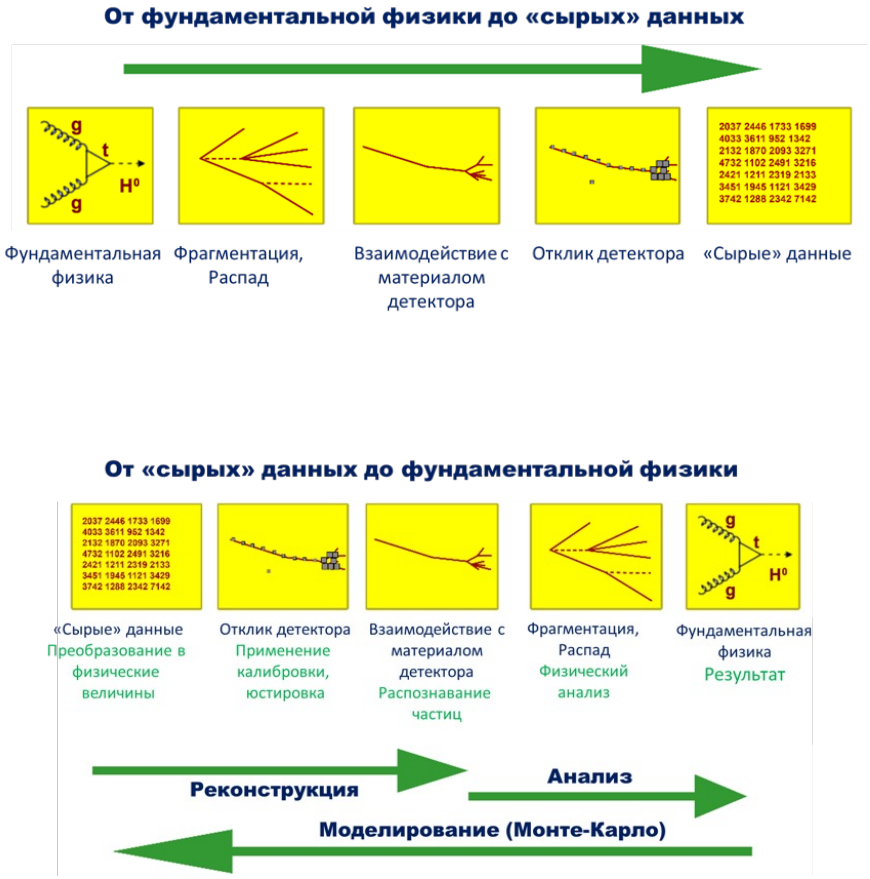
↔ Control

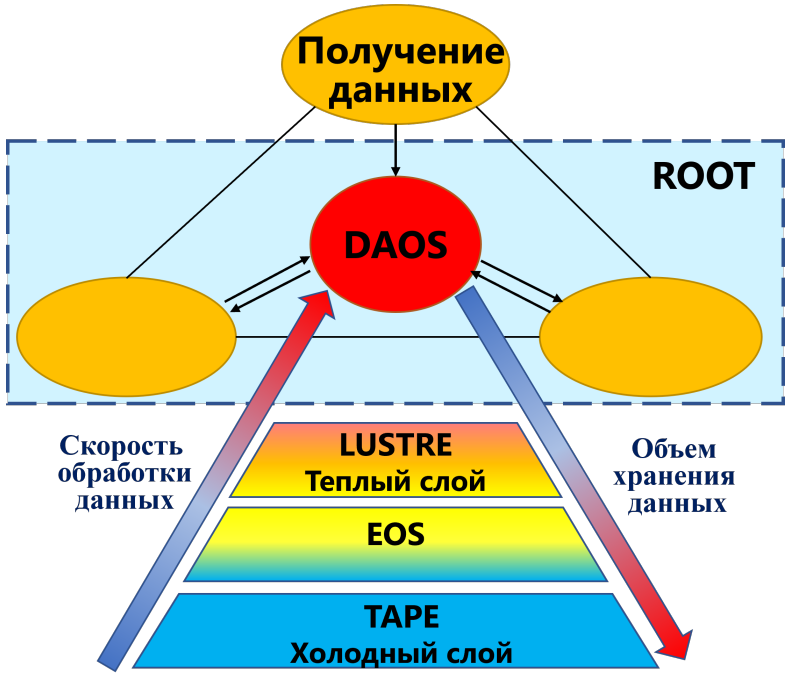
← Trigger

← Timing

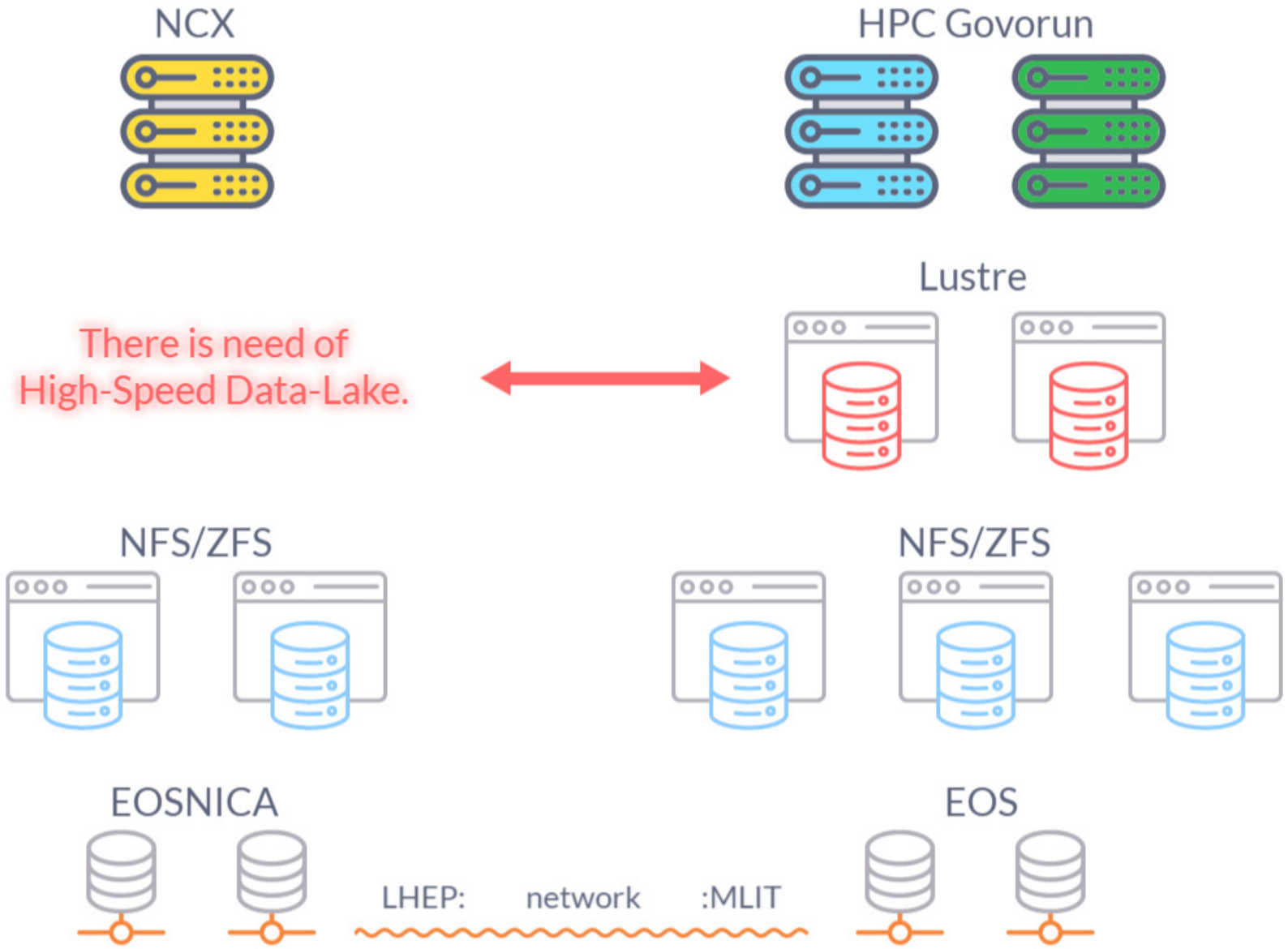
MPD Stage-1 DAQ parameters	
Beam	Au-Au 9 GeV
Trigger rate	7 kHz
Event size	1400 KB
Raw data rate	9.8 GB / s
Data taking time	8 months / year
Beam available	50% of time
Annual raw data size	38 PB
Compression factor	1:5 – 1:30
Annual storage size	1 – 8 PB

Stages of receiving, processing and storing data in experimental high energy physics

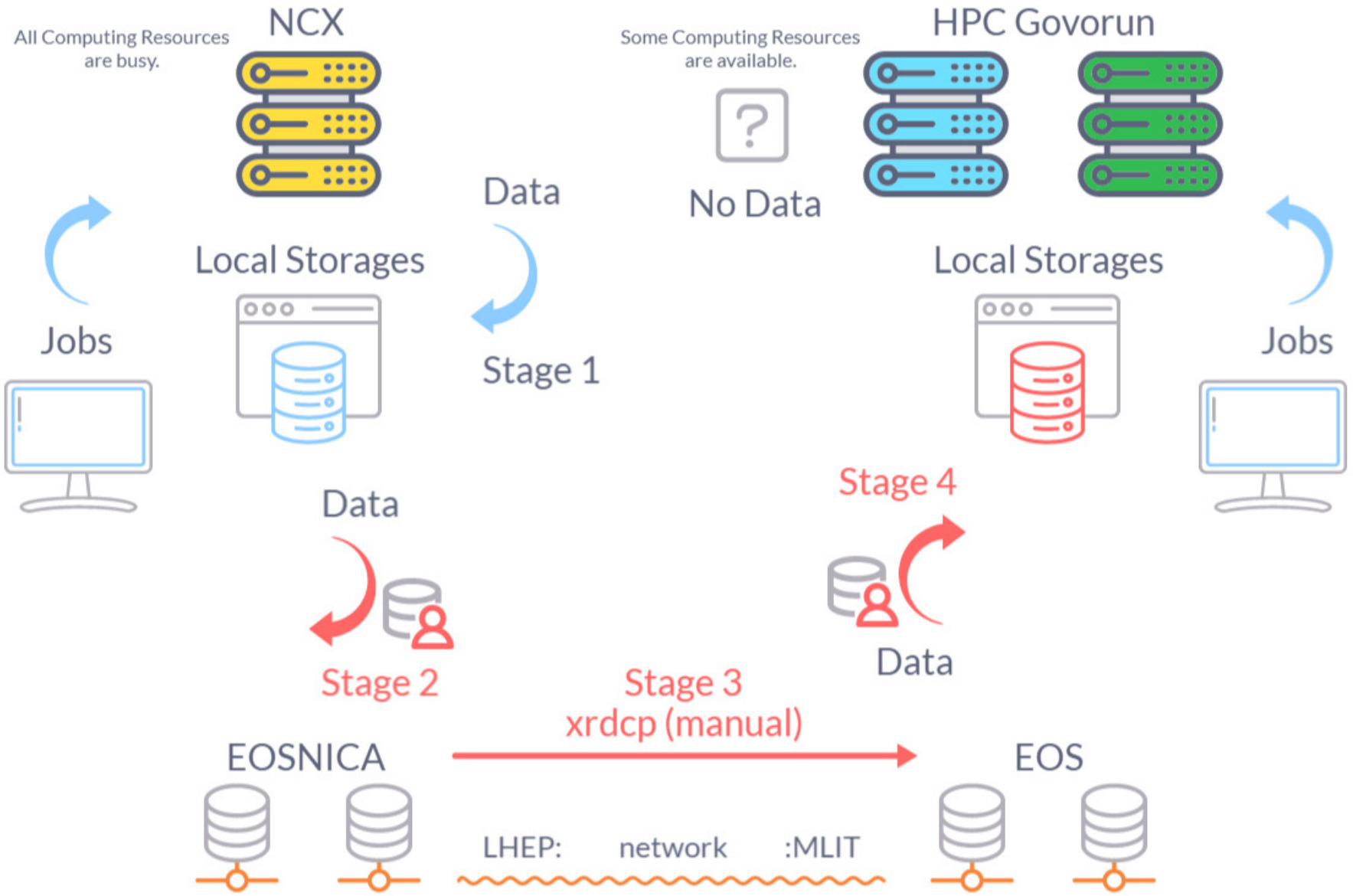




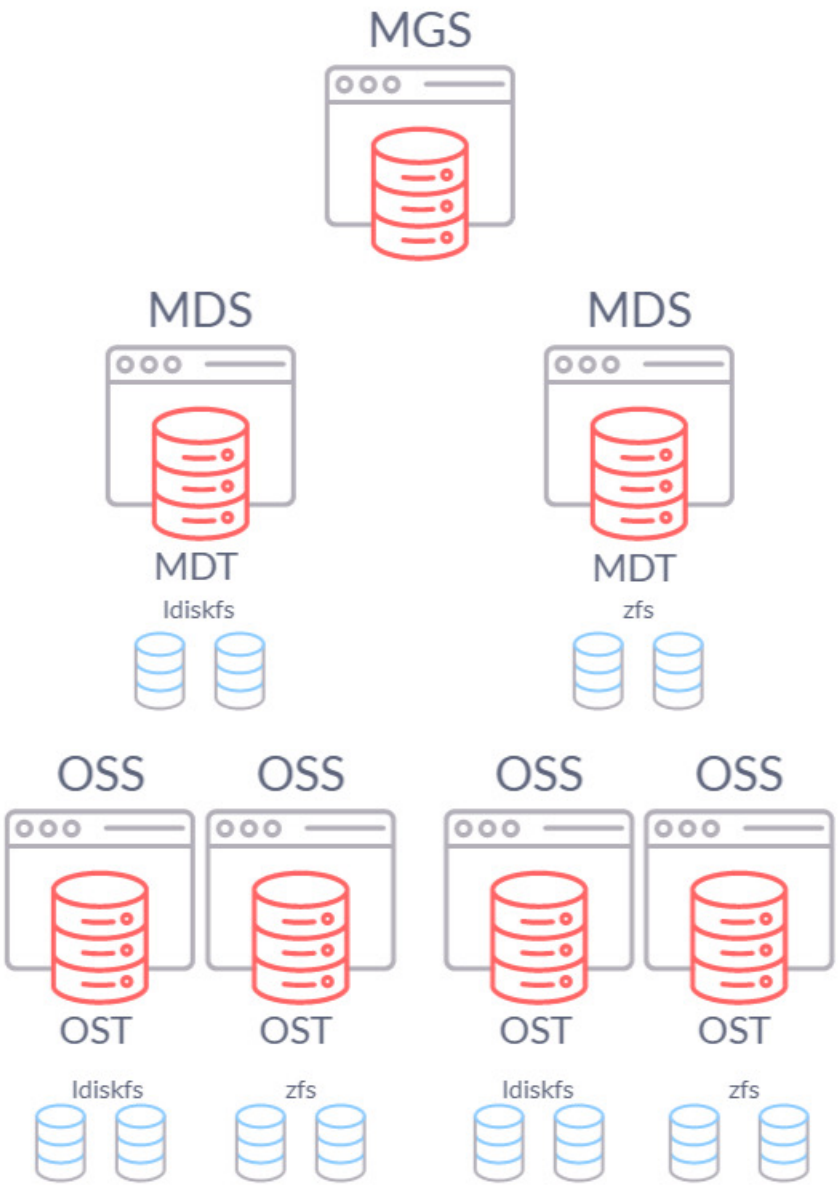
Clusters / HPC and Data storages systems



Data migration problem for BM@N, MPD and SPD experiments



Lustre filesystem components



Lustre Servers:

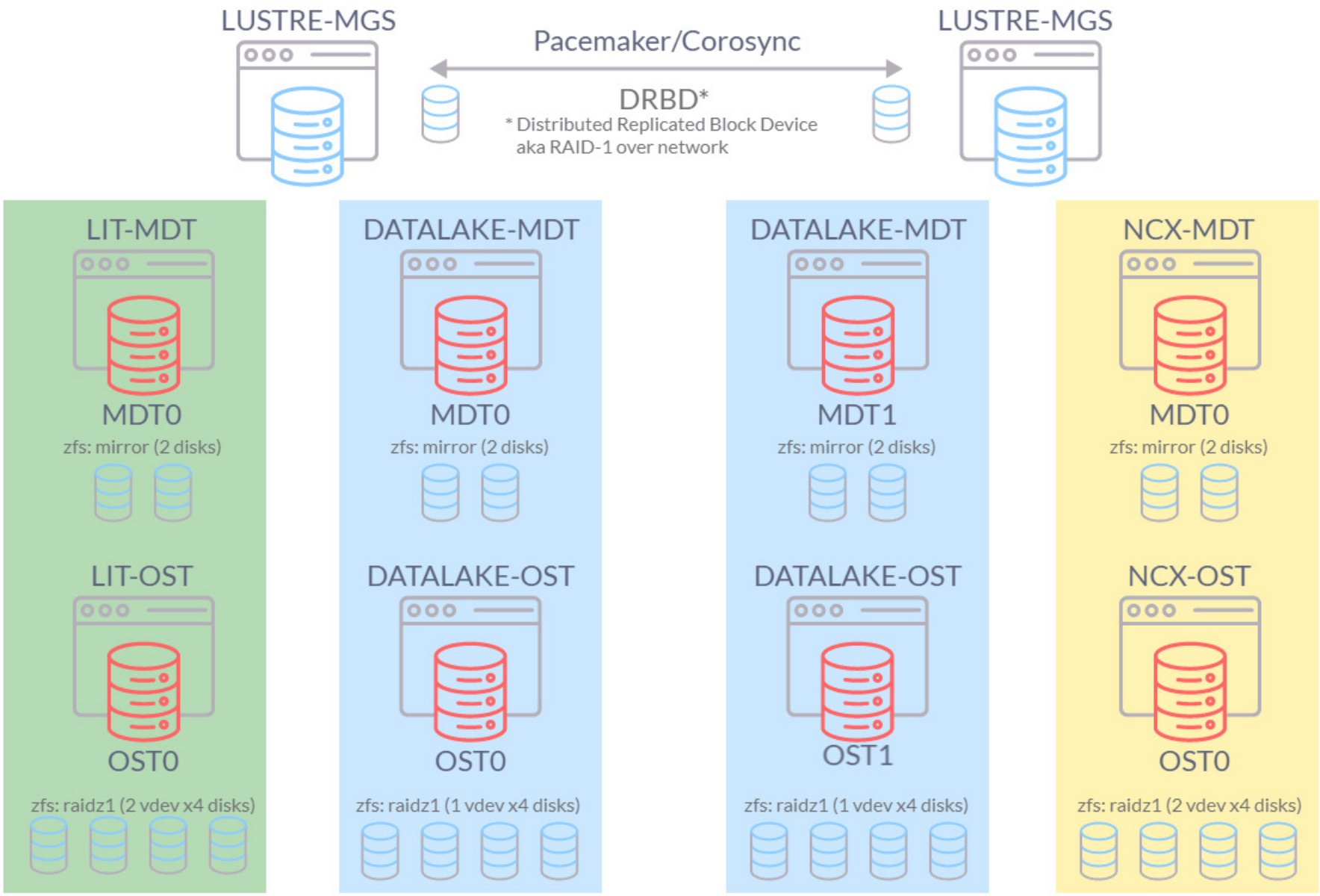
MGS Management Server
MDS Metadata Server
OSS Object Storage Server

MGT Management Target
MDT Metadata Target
OST Object Storage Target

Lustre Clients:

MGC Management Client
MDC Metadata Client
OSC Object Storage Client

Architecture for Data-Lake based on Lustre for mega-science project NICA



Stages

Team LHEP

Practice Deploy new filesystem from prepared RPMs

Team MLIT

Proto type

- Build Lustre Server and Client RPMs from sources
- Deploy Lustre «LIT», «NCX», «Data-Lake» at MLIT servers
- Setup failover for Lustre Services

Work

- Deploy Lustre «LIT» at MLIT servers and «NCX» at LHEP servers
- Deploy Lustre «Data-Lake» at MLIT and LHEP servers
- Setup a mirror feature on Lustre «Data-Lake»

Tests Run user's jobs for testing Lustre «LIT», «NCX» and «Data-Lake»

Hardware

MLIT servers

2x
Dell PowerEdge R730xd

2x 160 TB, SAS

Motherboard	PowerEdge R730/R730xd System Board
Processor	2x Intel Xeon E5-2660 v4 @ 2.00 GHz
Memory	8x Micron DDR4 2400 MHz, 16 GB (128 GB)
RAID	Dell PERC H730P
Disk	2x Dell MFC6G (Samsung) SSD SAS, 400 GB (2x 400 GB) 16x HGST UltraStar HE10 SAS, 10TB (160 TB)
Network	Dell 99GTM (Intel X540-T2 2x 10 Gb/s + Intel I350 Dual Port 2x 1 Gb/s)
Power	2x 750W Redundant Power Supply

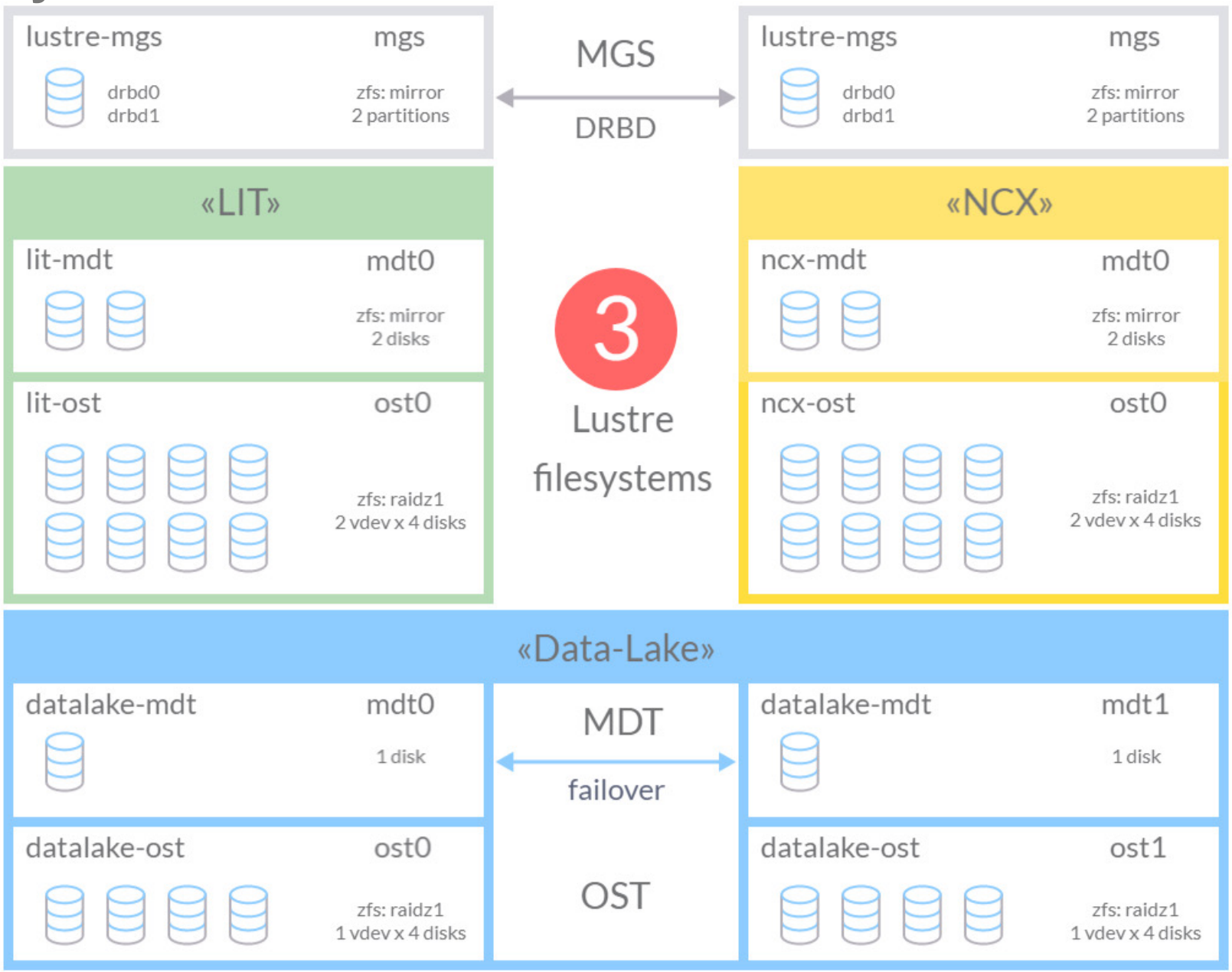
LHEP servers

2x
Supermicro SSG 1029P-NEL32R

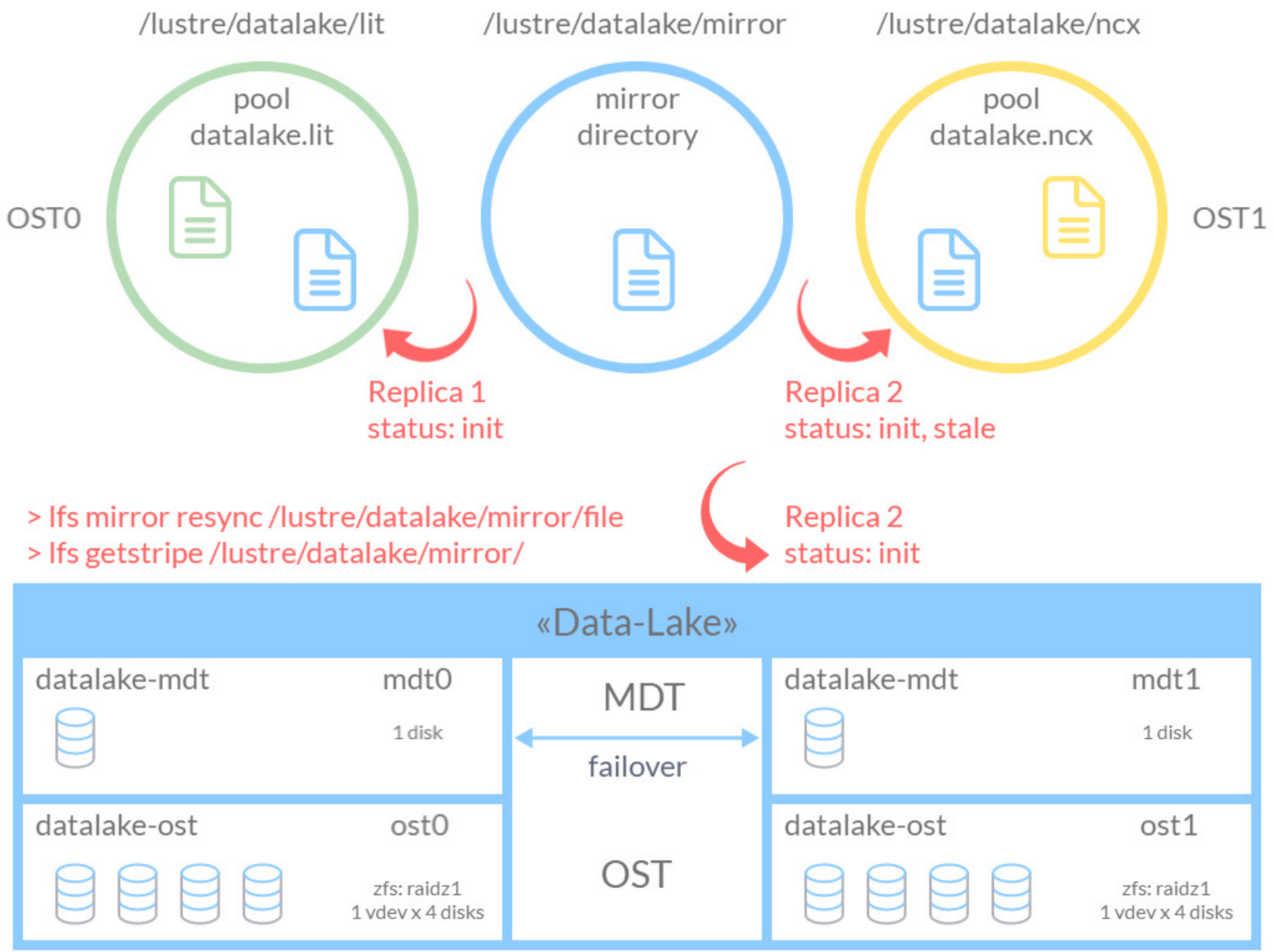
2x 244.8 TB, NVMe (Rulers)

Motherboard	Supermicro X11DPS-RE
Processor	2x Intel Xeon Gold 6230R @ 2.10 GHz
Memory	12x Samsung DDR4 2993 MHz, 64 GB (768 GB)
Disk	2x Apacer SSD NVME m.2 , 512 GB (2x 512 GB) 16x Intel DC P4510 SSD NVME (Ruler), 15.3TB (244.8 TB)
Network	Intel X550-T Dual Port 2x NVidia (Mellanox MT27800) ConnectX-5 Dual Port 2x 100 Gb/s Ethernet
Power	2x 1600W Redundant Power Supply

Data-Lake based on Lustre
Deploy at MLIT and LHEP servers



Data-Lake based on Lustre Pools and Mirror directory



Results

Data-Lake for mega-science project NICA

Theory — Create Architecture for Data-Lake based on Lustre filesystem

Practice — Setup Lustre filesystem at different servers from prepared RPMs by MLIT and LHEP teams

Work

- Build Lustre Server and Client RPMs from sources
- Deploy 3 Lustre filesystems at MLIT and LHEP servers
- Setup failover for Lustre Services (MGT and MDT)
- Setup a mirror feature (Lustre pools and mirror directory)

Future tasks

?

- Setup Pacemaker/Corosync for failover MGS/MGT
- Run user's jobs for testing Lustre «Data-Lake»

Thank You
for attention!