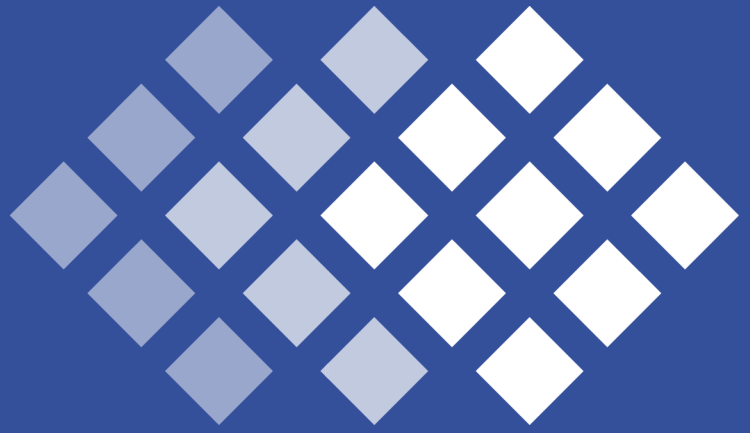
RSC

Group



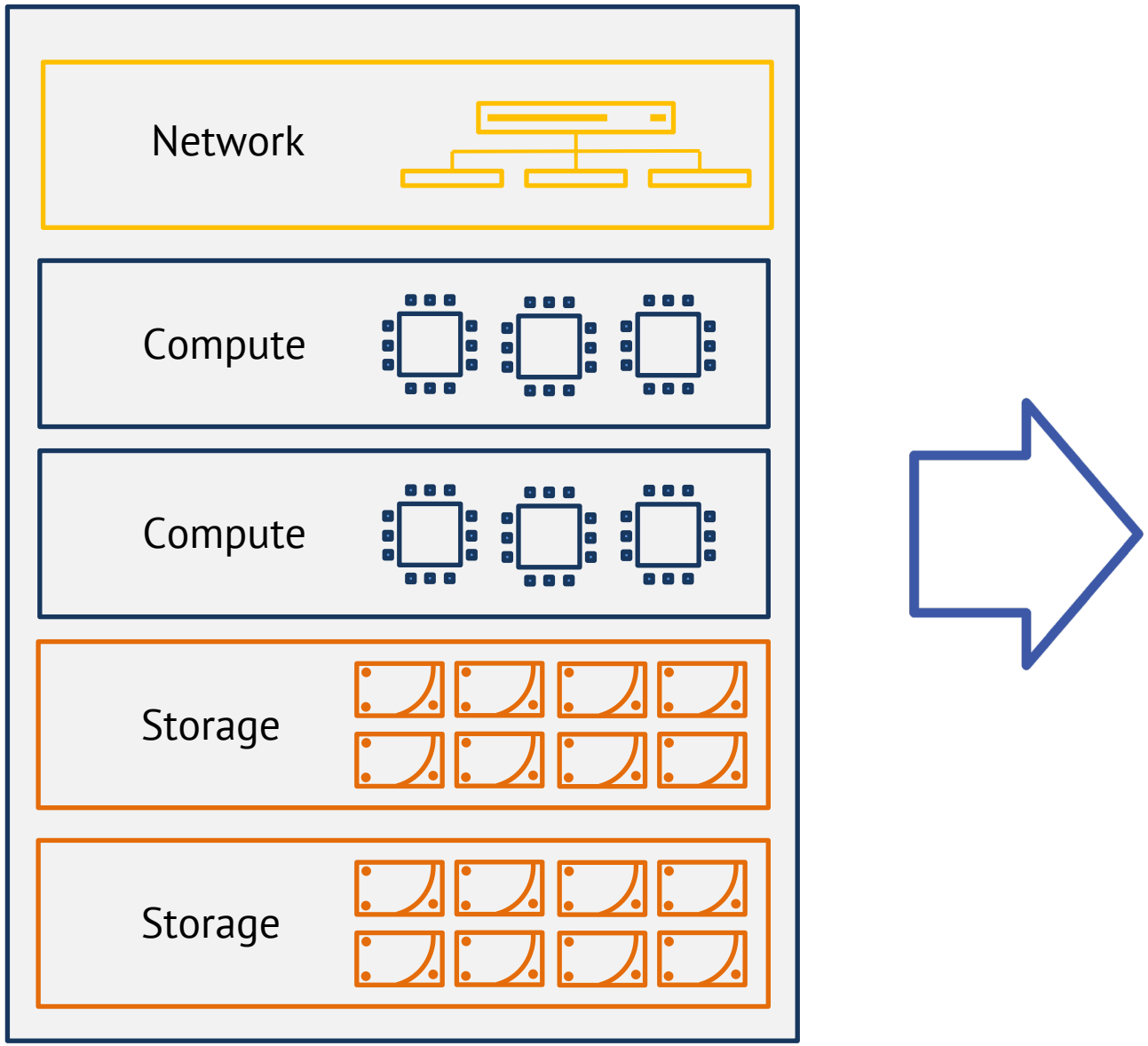
Leading Solutions
for HPC and Data Centers

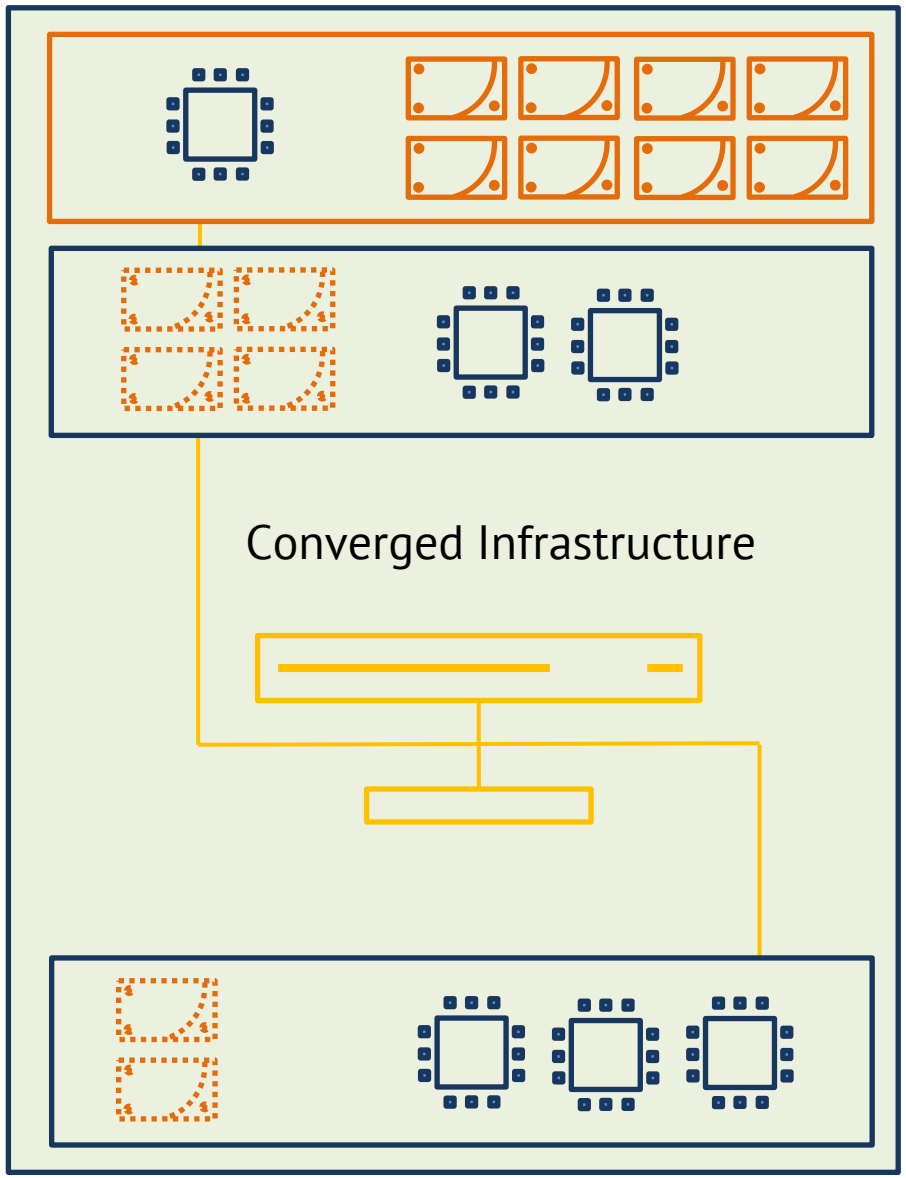
RSC BasIS Platform: Orchestration for High Performance Composable Storage Architectures

Pavel Lavrenko, CBDO, RSC

November, 19 2020, DAOS User Group 2020

From Rackscale to Composable





Current DAOS status

DAOS is great but still has a lot of complications:

- DAOS requires specially-designed hardware platforms to deploy
- DAOS deployment is tricky
- DAOS doesn't fit HPC Cloud

RSC is focused to address these points

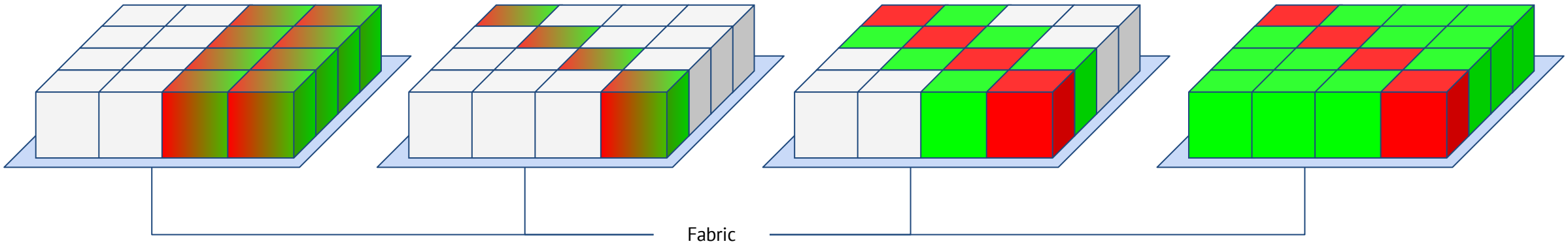
DAOS Storage system models

Pooled Storage Model

Hyper-Converged Storage Model

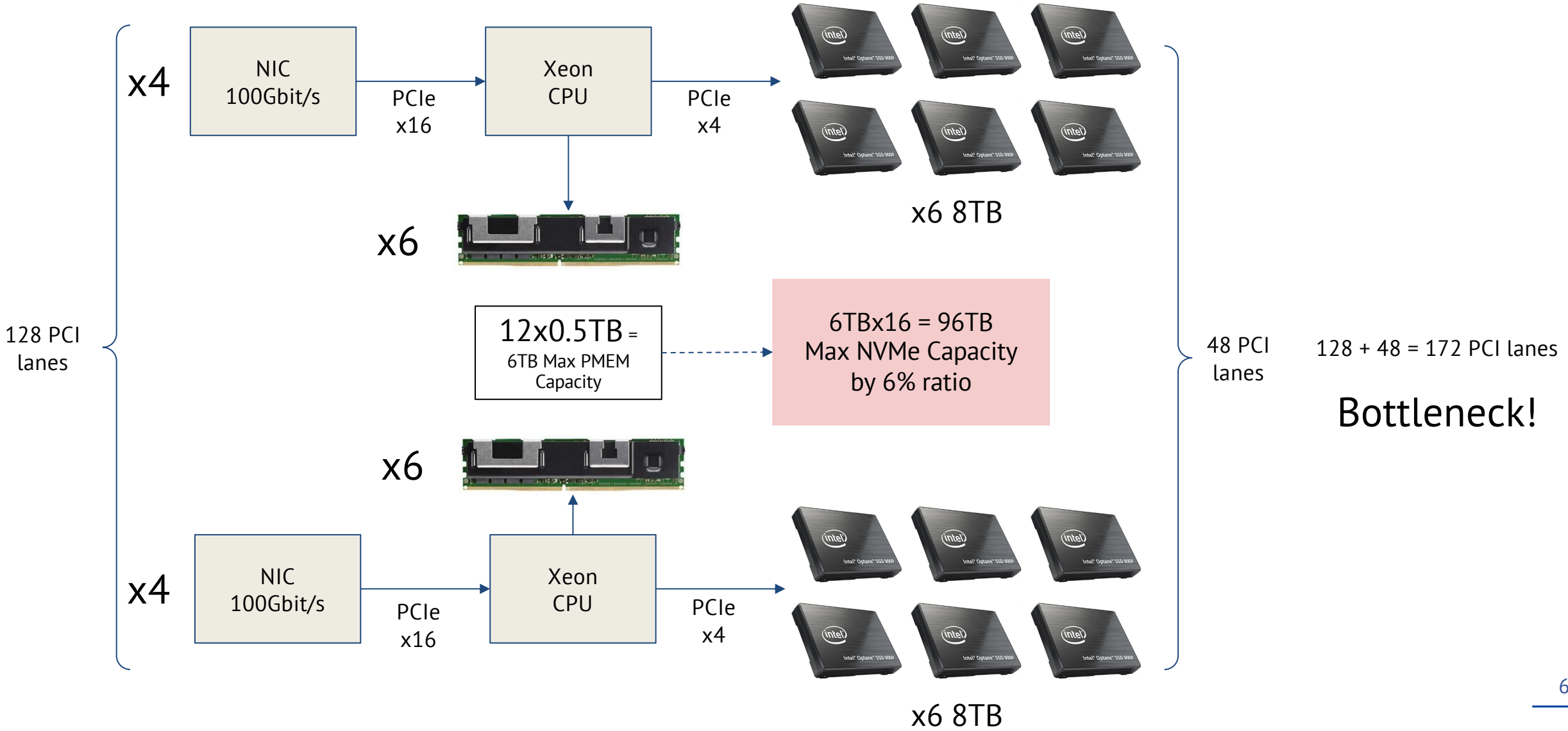
Disaggregated Storage Model (Dedicated servers)

Disaggregated Hyperconverged Model (All servers participate in DAOS)

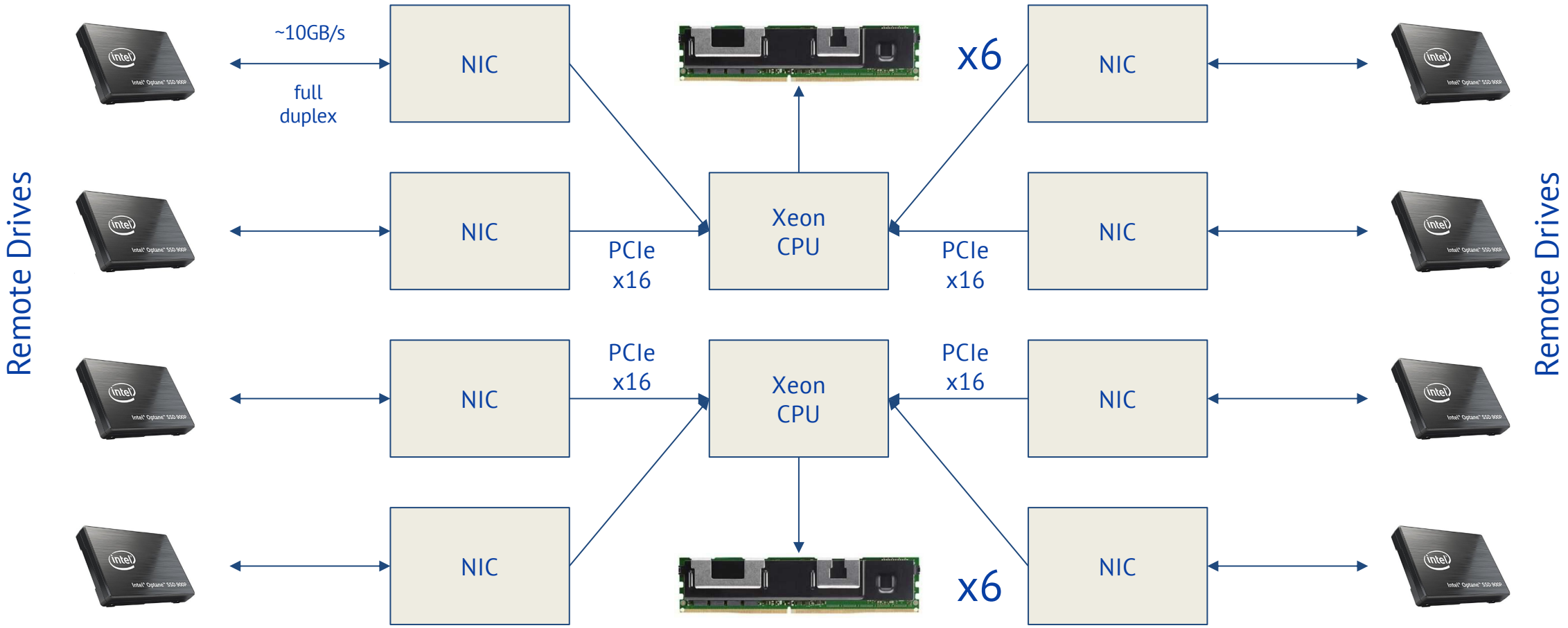


	Nodes With PMEM & NVMe
	Compute Nodes
	Compute with NVMe
	Compute with PMEM

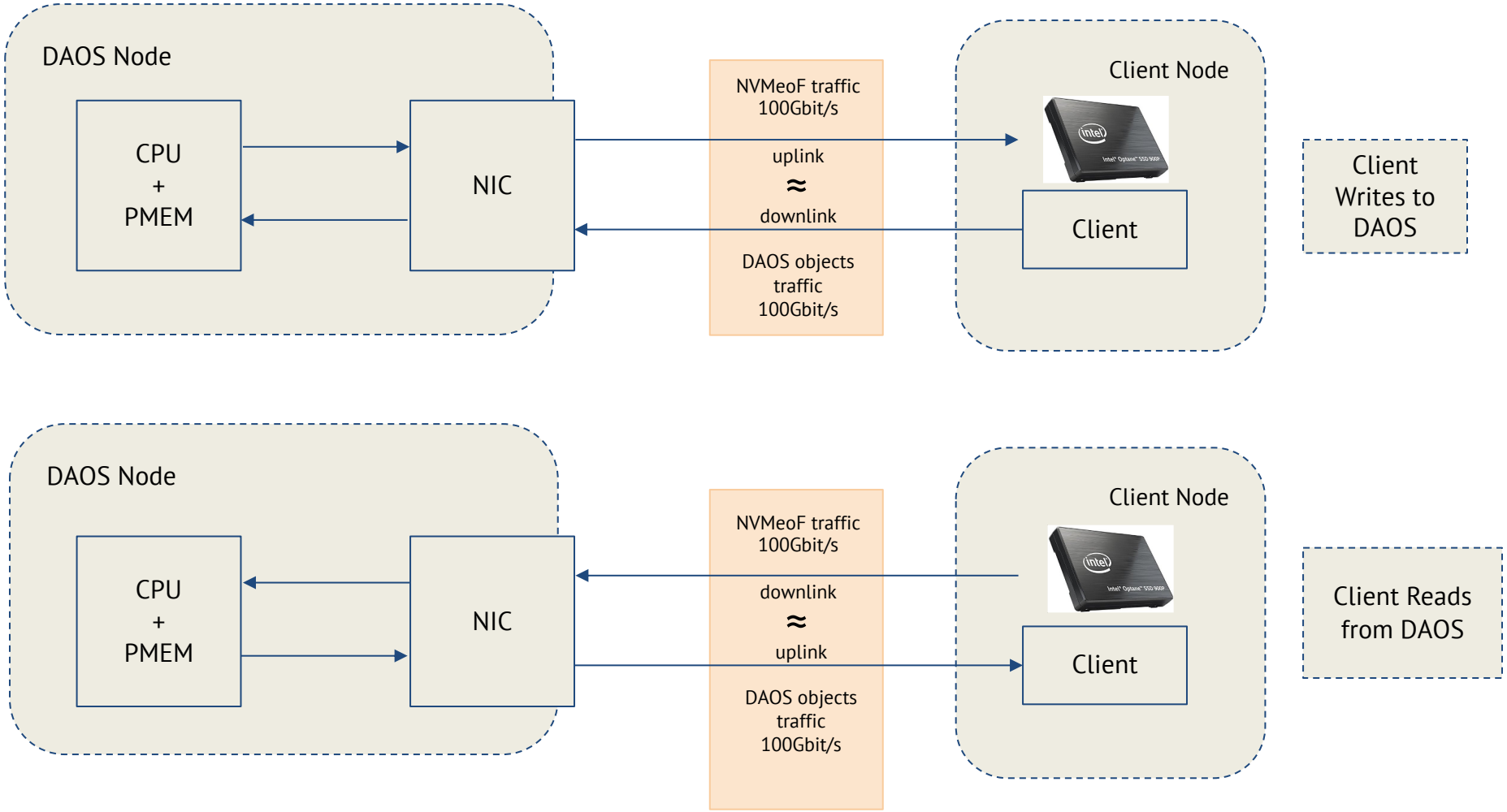
DAOS: Large Capacity Requires Dedicated Storage Servers



Architecture with NVMe-over-Fabrics



Interconnect utilization NVMe-over-Fabrics



Complete utilization of full duplex network:

DAOS data and NVMeOF always move in the opposite directions

Works well when DAOS cluster uses SSD from client nodes

Extra NVMeoF Latency doesn't affect storage performance because of PMEM

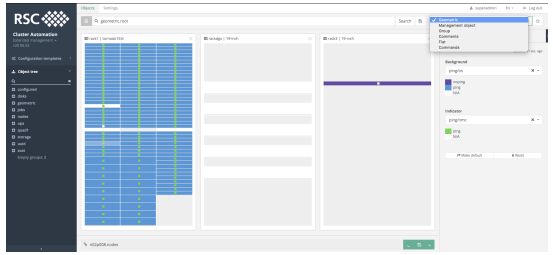
RSC BasIS Orchestration

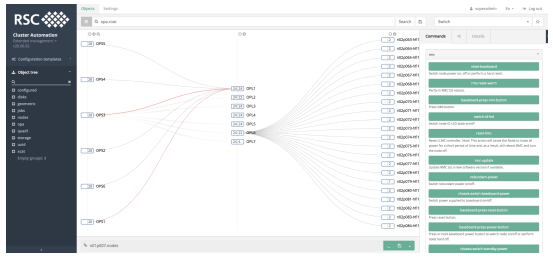
Vertical integration of Hardware, Software and Infrastructure components

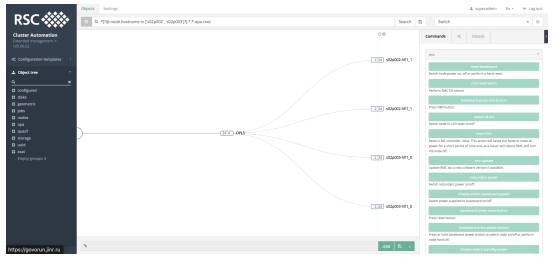
Knowledge about all datacenter objects and their connections

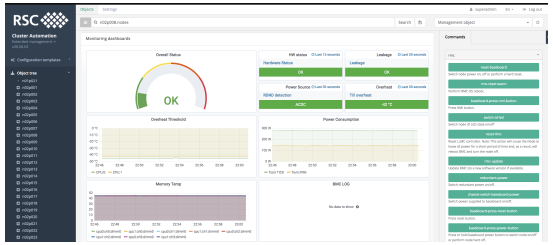
Microagent Mesh for Cluster Automation

- App Repository
- Agents
- Agent Lifecycle
- Messaging system
- SDK









Knowledge of objects

- Auto-discovery
- Inventory and classification
- Knowledge of topologies
- Dynamic selection based on Query language

Continuous configuration

- Repository of configuration
- Maintaining consistency

Group Commands Execution

- Human operator – Platform
- Agent to agent

Monitoring

- Dynamical status representation
- GUI for drill-down analysis
- Problem-oriented dashboards

RSC

Cluster Automation
Extended management
v20.10-rc

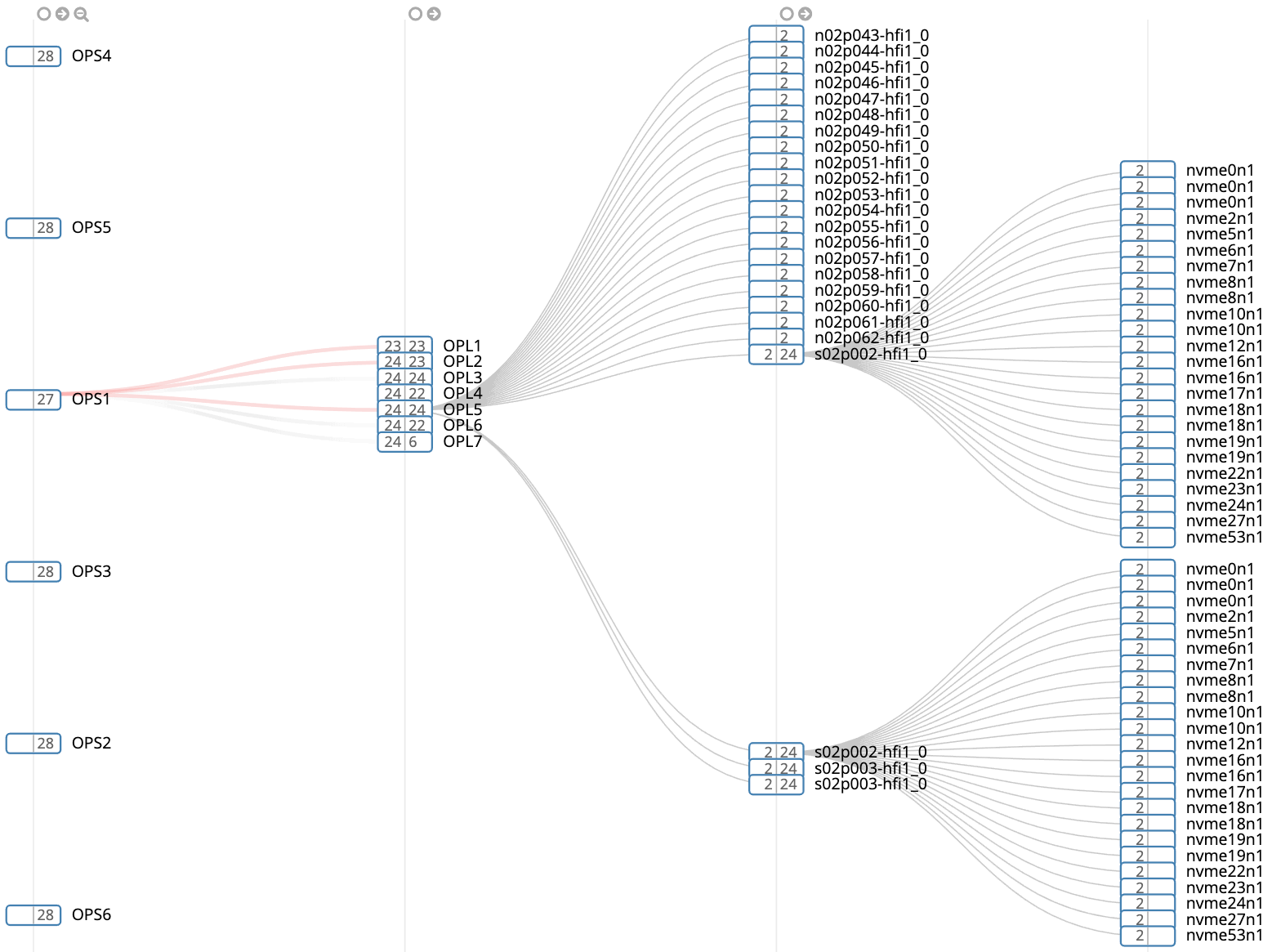
Configuration templates

Object tree

- configured
- disks
- geometric
- jobs
- nodes
- opa
- processes
- quad1
- storage
- uuid
- xcat

Empty groups: 2

opa.root | Search | Switch



Commands | Details

rmc

reset-baseboard
Switch node power on, off or perform a hard reset.

rmc-reset-warm
Perform RMC OS reboot.

baseboard-press-nmi-button
Press NMI button.

switch-id-led
Switch node ID LED state on/off

reset-llmc
Reset LLMC controller. Note: This action will cause the Node to loose all power for a short period of time and, as a result, will reboot BMC and turn the node off.

rmc-update
Update RMC sto a new software version if available.

redundant-power
Switch redundant power on/off.

chassis-switch-baseboard-power
Switch power supplied to baseboard on/off.

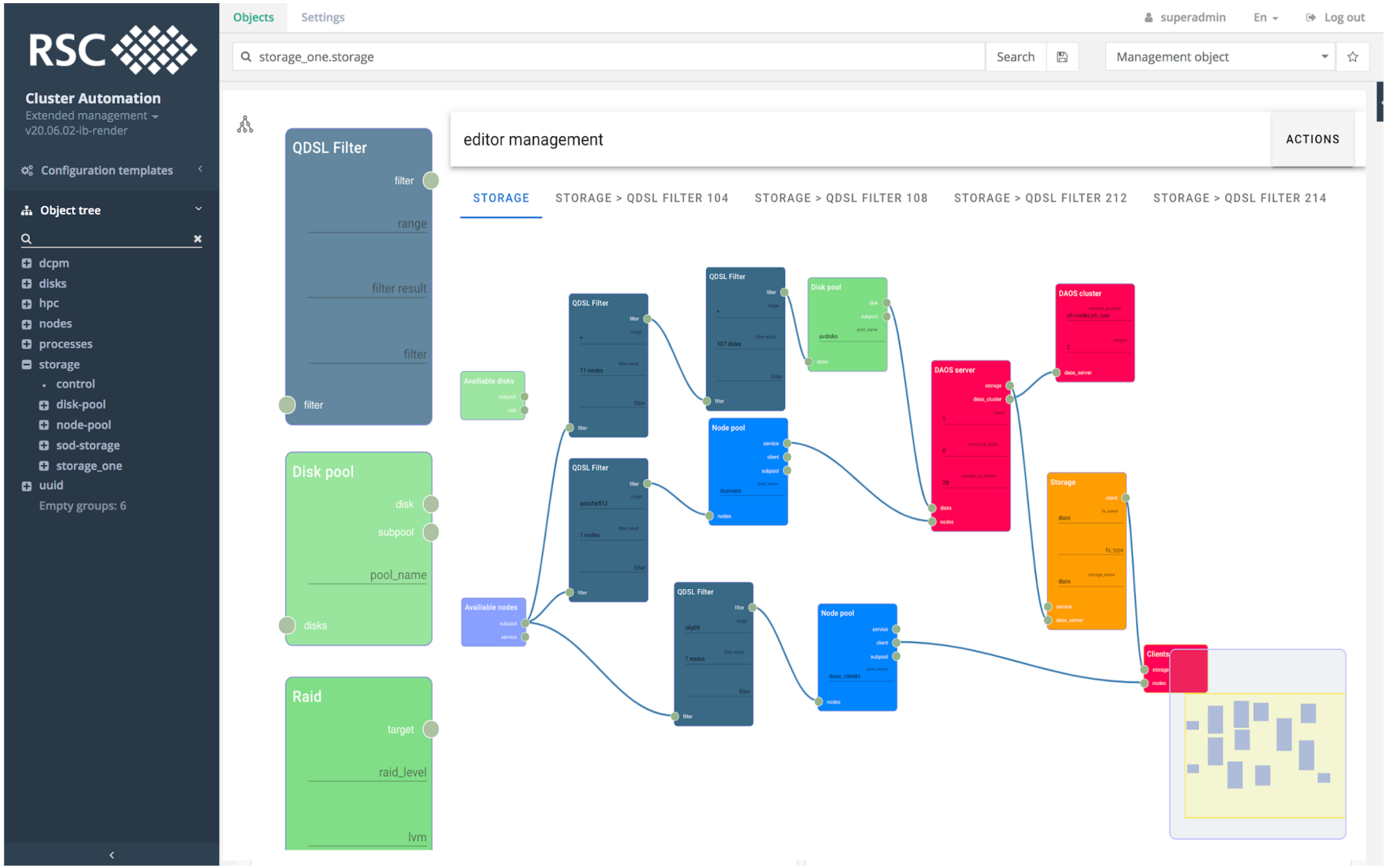
baseboard-press-reset-button
Press reset button.

baseboard-press-power-button
Press or hold baseboard power button to switch node on/off or perform node hard off.

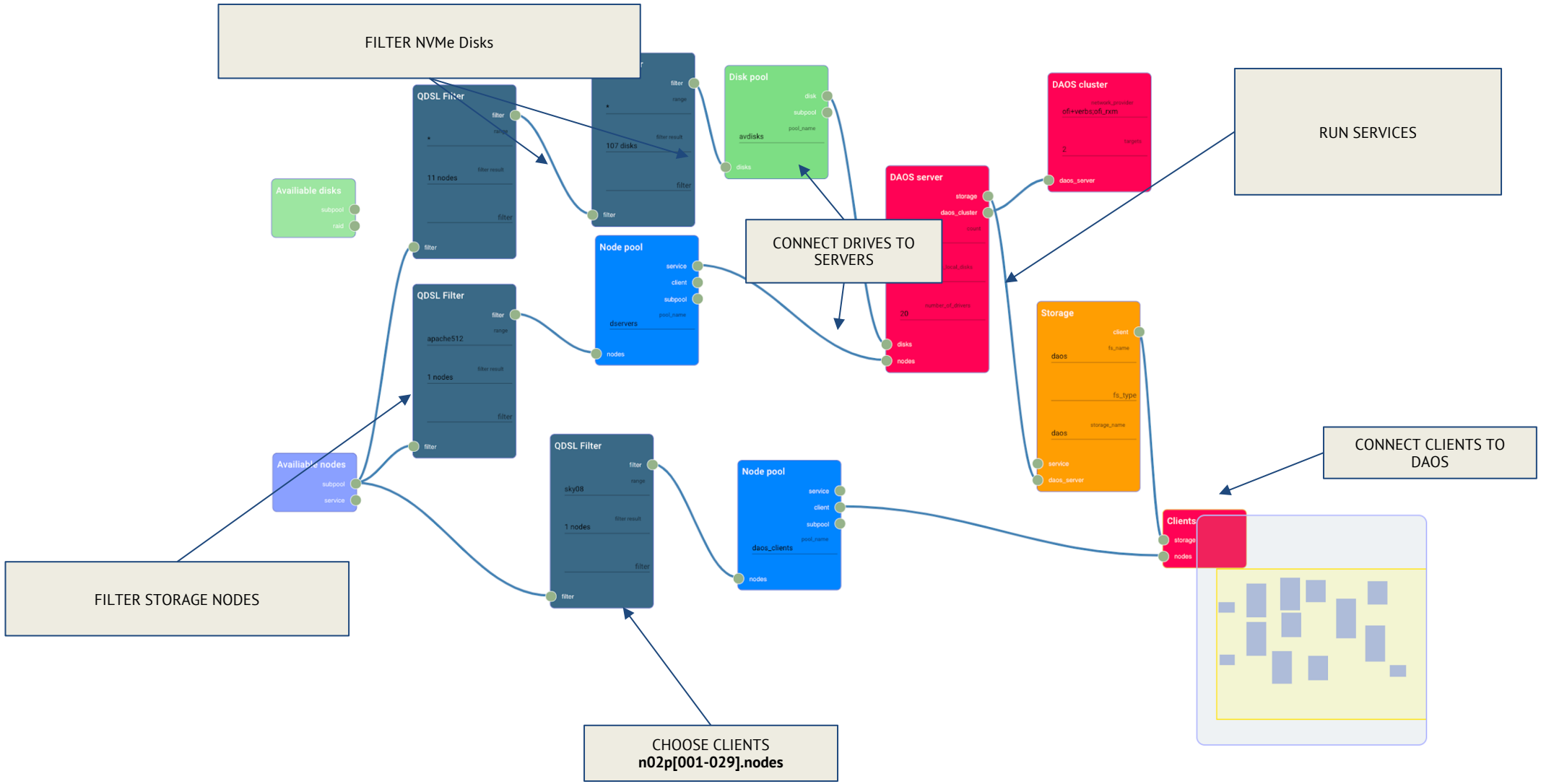
chassis-switch-standby-power
Switch standby power on/off.

rmc-reset-cold
Perform full power cycle on a RMC.

BasIS Storage Orchestrator



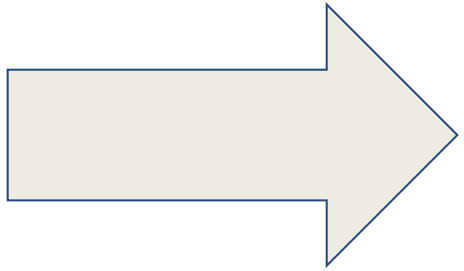
BasIS: DAOS with NVMeOF Pipeline



Flexible PMEM-only server roles

DAOS Server

Storage Server
with PMEM & NVMe



- Less complex
- Cheaper
- More roles

PMEM-only Server

DAOS Server
with NVMeOF Drives

AI

In-Memory DBs

Grid Systems

IOR results with DFS API

Configurations	BW write MB/s	BW read MB/s
2 IO instances and 4 local NVMe drives	2132	2008
2 IO instances and 4 NVMeOF drives	2253	2178
4 IO instances and 8 local NVMe drives	4679	3935
4 IO instances and 8 NVMeOF drives	4248	4268

NVMe drives - Intel P4510 2TB - W: 2 GB/s and R: 3.2 GB/s by specs
DAOS: kdev (AIO Linux driver) was used for NVMeOF drives, 2 targets per disk, max pool size, service replica = 1, ofi+sockets provider through Intel Omni-path
MPI: np=104 from 3 clients

Conclusions

What have we archived with RSC BasIS Storage Orchestration:

- ~~DAOS requires specially-designed hardware platforms to deploy~~
  - You can have DAOS cheap - just buy PMEM and compose DAOS over a fabric
  - Existing servers can share their NVMe drives

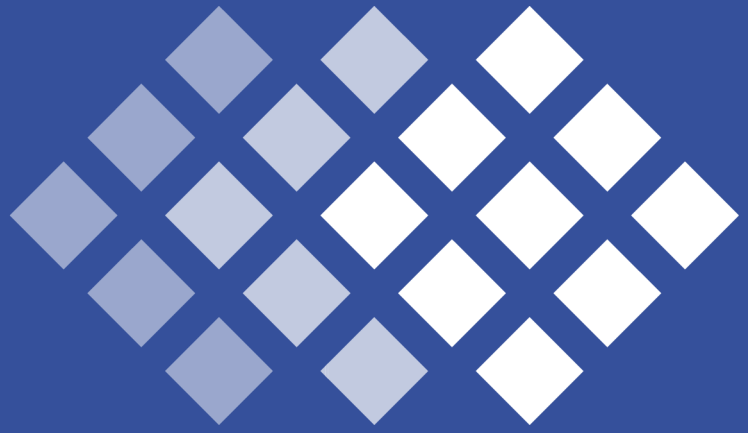
- ~~DAOS deployment is tricky~~
  - Software orchestration significantly simplifies DAOS deployment

- ~~DAOS doesn't fit HPC Cloud~~
  - Composable Disaggregated approach gives flexible ways to use PMEM nodes
  - DAOS can be dynamically assembled when needed

RSC Announces DAOS Support in its storage orchestration platform
https://www.hpcwire.com/off-the-wire/rsc-announces-intel-ice-lake-sp-and-daos-support-introduces-tornado-afs-storage/

RSC

Group



rscgroup.ru
hq@rsc-tech.ru