Storage Strategy for HPC users

Thomas Eifert

IT Center | RWTH AACHEN UNIVERSITY

Storage Parameters

- Size / Capacity
  - Total amount of storage
  - Number of (small) files
- Performance
  - GB/s
  - Op/s
  - Time to data
- Safety
  Protection against
  - accidental deletion
  - System desaster

Satisfying Demands

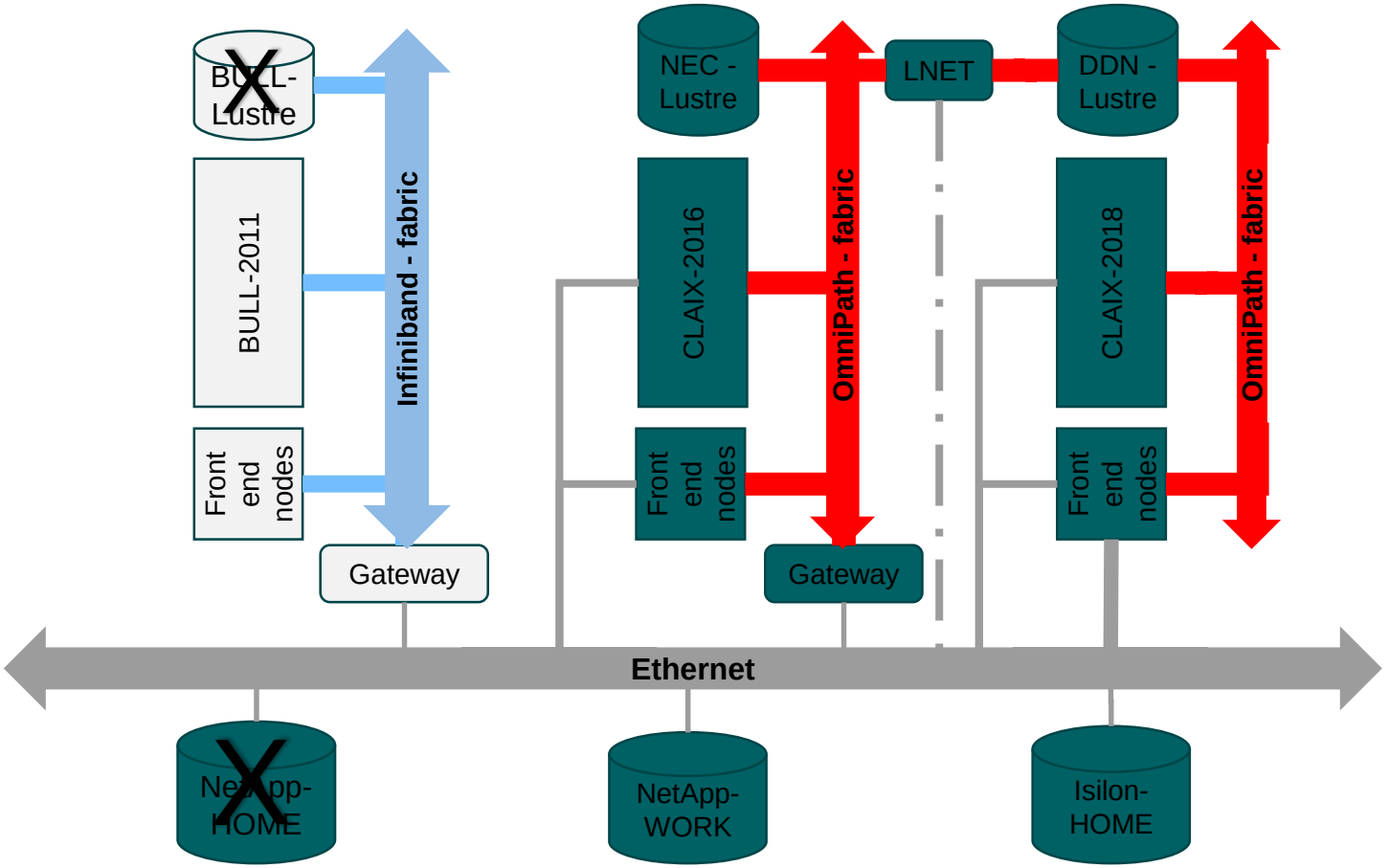
- Common environments: compromise for optimal mix

- HPC: environments for not-so-common demands
  ➔ very large / very fast / safe

- Challenge 1: satisfy the demands ✓
  we offer 10 PB, 150 GB/s, Snapshots and Backup

- Challenge 2: in one system ↯
  ➔ not affordable

RWTH Compute Cluster Configuration Overview



IT Center | RWTH AACHEN UNIVERSITY

What do we offer?

	Type	Total Size	Snapshot / Backup	Def. Quota	
`$HOME`	NFS/CIFS, Cluster-wide	1.5 PB	X / X	150 GB	10 GB/s
`$WORK`	NFS/CIFS, Cluster-wide		X / -	250 GB	3 GB/s
`$HPCWORK`	Lustre Cluster-wide, 2 Instances	C'16: 2.3 PB C'18: 10 PB		1 TB	C'16: ~30 GBs C'18: ~150 GB/s
`$TMP`	Ext4 node-local	Several GB .. 1.5 TB			

https://doc.itc.rwth-aachen.de/display/CC/Available+File+Systems

Storage Strategy

- Your decision:
  - Depending on Your various types of data You have the choice:
    - „hand made“ data + results should reside in `$HOME`
      - Good for small files
      - Protected by snapshot + backup
    - Results can also go to `$WORK` : similar characteristics, but no backup
    - Snippets of intermediate stuff may go to `$TMP` : node-local, job-local
    - Large intermediate data should go to `$HPCWORK`
      large, fast, no loss protection
- Your obligation

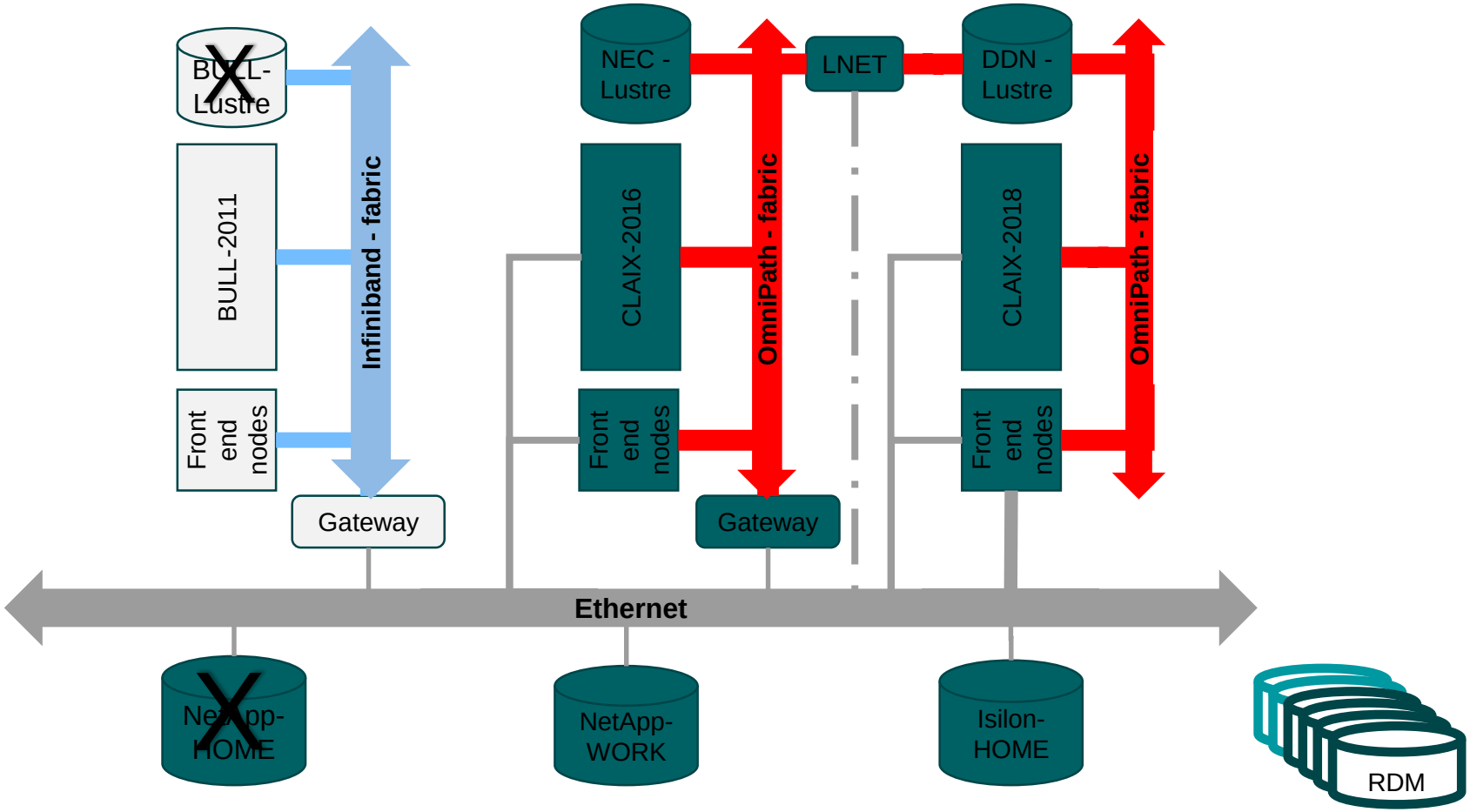
Non-HPC Storage

- „Result“: data leading to knowledge
- Connecting the realm of HPC with ResearchDataManagement (RDM)
- Aims: store data according to FAIR principle
  (findable, accessible, interoperable, reproducable)
- Here: store resulting data in a way that
  - They can be linked together with the publication
  - The parameters are stored: of simulation, of experimental setting, …

Under Construction: Dedicated Storage for RDM

- Currently implemented: testbed
- Funding granted for RD storage
- Storage characteristics:
  - Large
  - ~ slow
  - „persistency layer“ ➔ guaranteed safety
  - Structures and processes for lifetimes far beyond the hardware

RWTH Compute Cluster Configuration Overview



IT Center | RWTH AACHEN UNIVERSITY

Thank You!