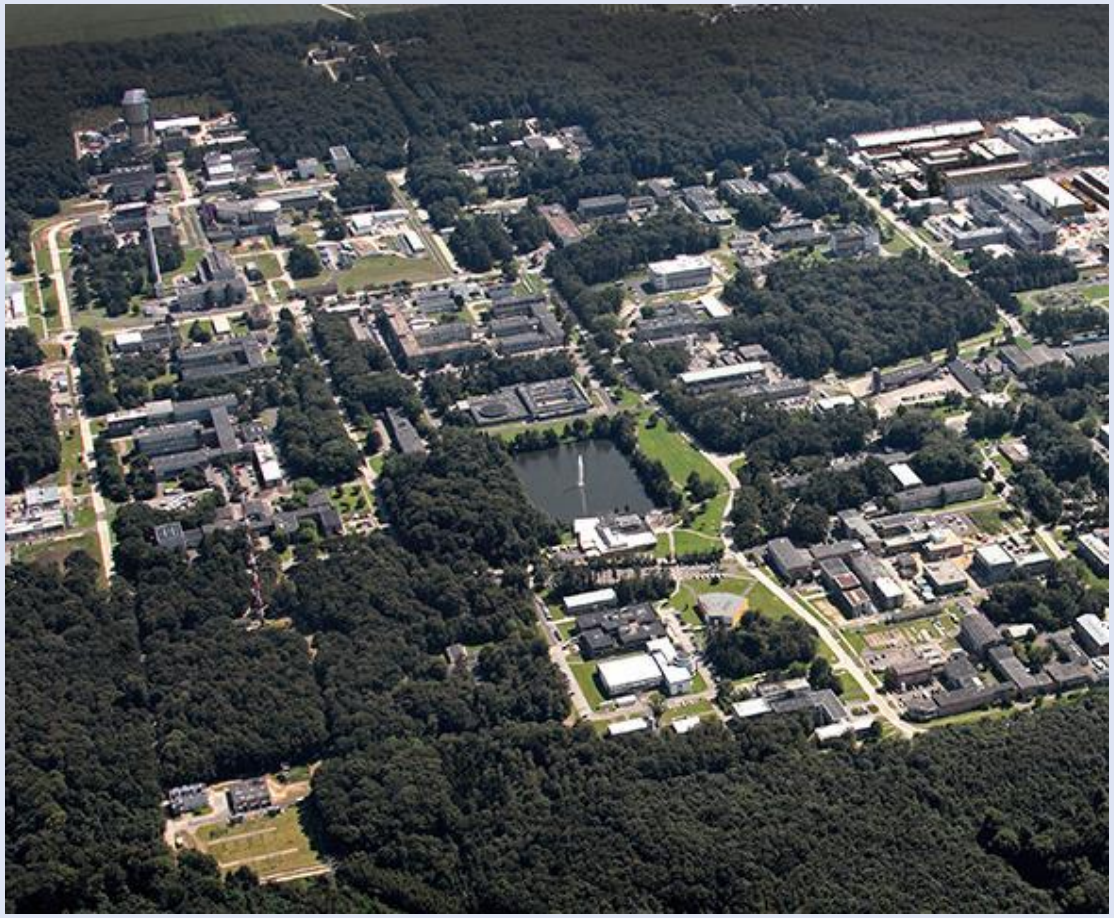
NEW STORAGE FOR JUELICH

EXABYTE FOR EXAFLOP

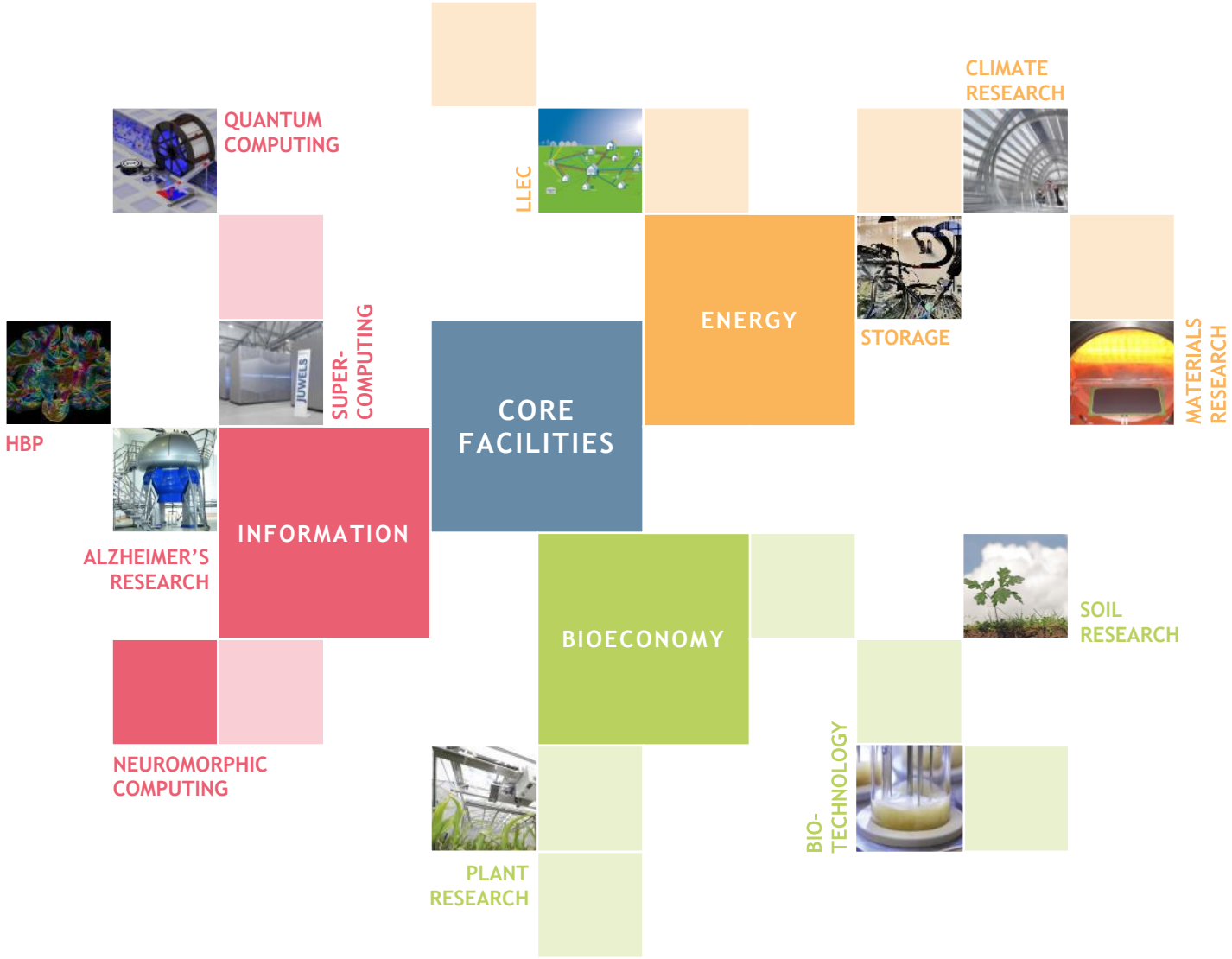
06. MARCH 2024 I SALEM EL SAYED, MARTIN LISCHEWSKI

RESEARCH AND DEVELOPMENT

on 2.2 Square Kilometres



STRATEGIC PRIORITIES



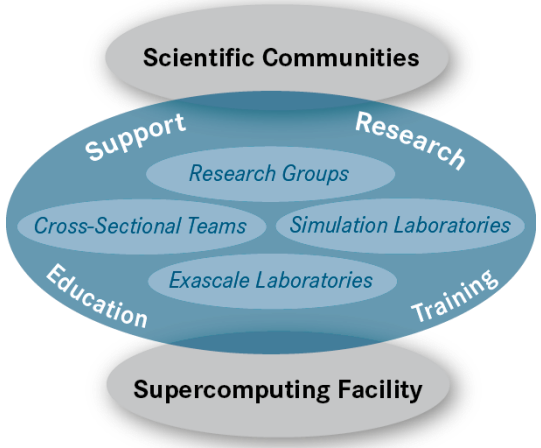
JÜLICH Forschungszentrum

JÜLICH SUPERCOMPUTING CENTRE

JÜLICH Forschungszentrum

JÜLICH SUPERCOMPUTING CENTRE

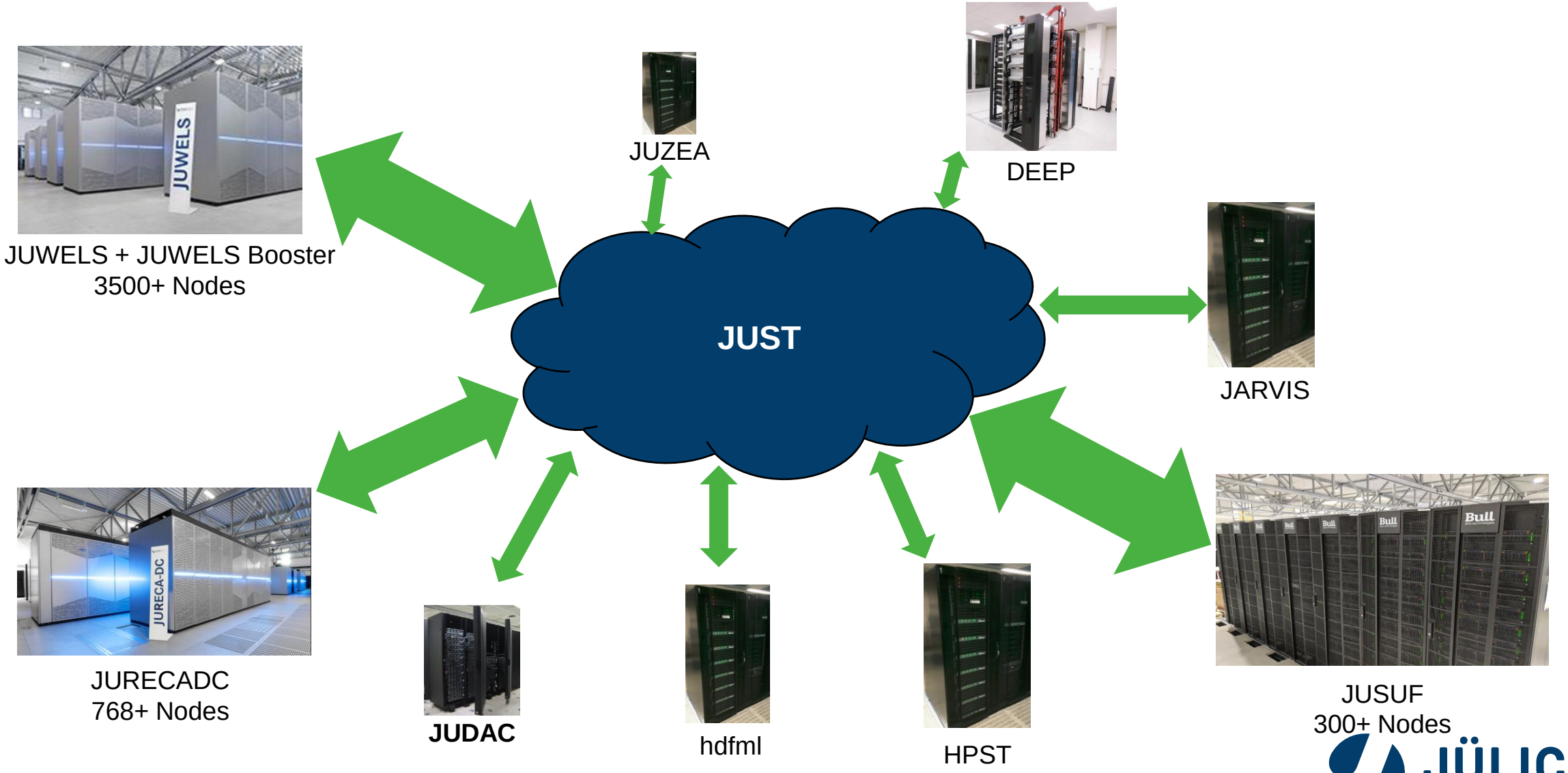
- **Supercomputer operation for:**
  - Center – FZJ
  - Region – RWTH Aachen University
  - Germany – Gauss Centre for Supercomputing / John von Neumann Institute for Computing
  - Europe – PRACE, EU projects
- **Application support**
  - Unique support & research environment at JSC
  - Peer review support and coordination
- **R-&-D work**
  - Methods and algorithms, computational science, performance analysis and tools
  - Scientific Big Data Analytics
  - Computer architectures, Co-Design, Exascale Laboratories: EIC, ECL, NVIDIA
- **Education and Training**



DEEP

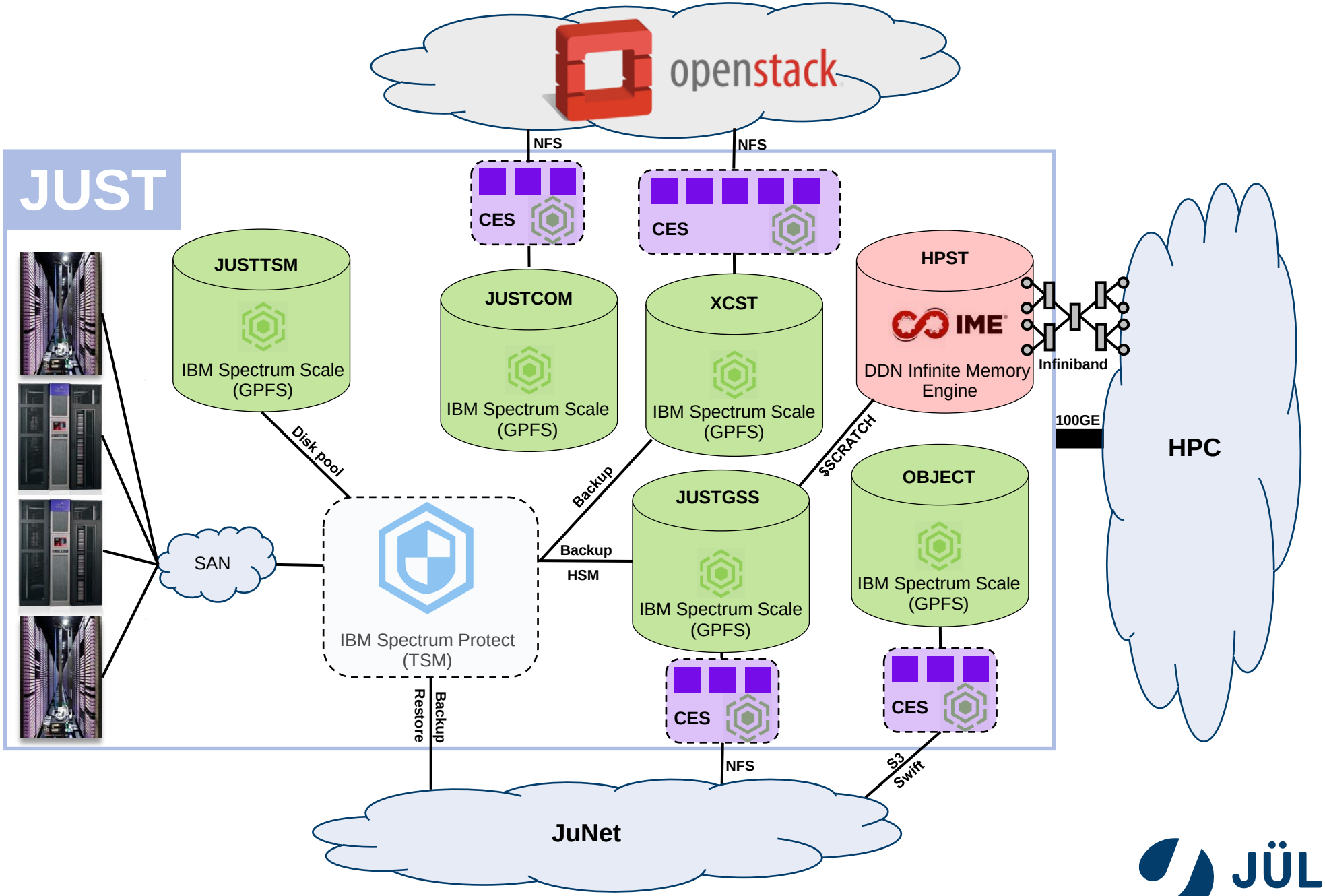
JÜLICH Forschungszentrum

JUELICH STORAGE (JUST)



JUST5

Logical view

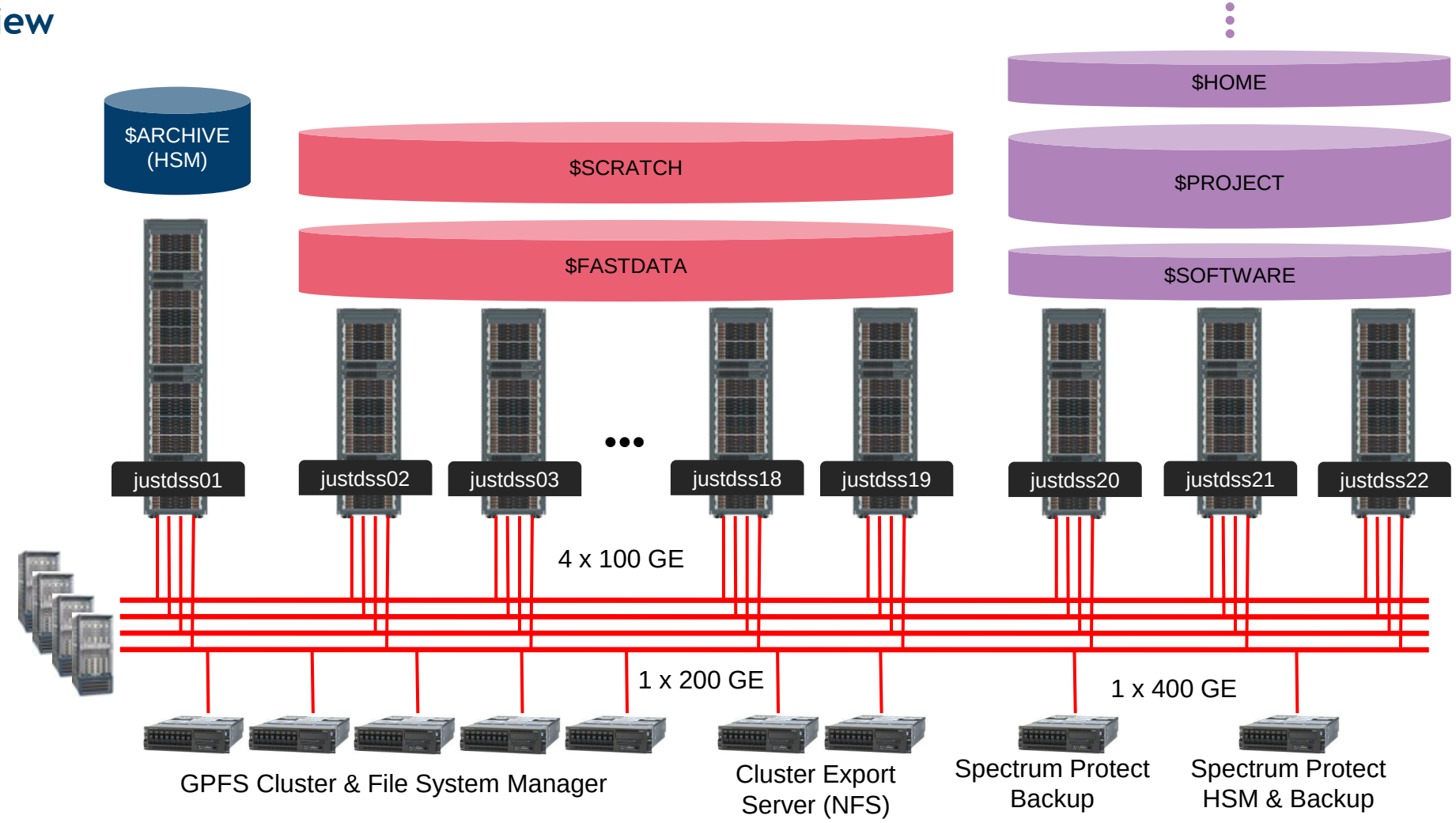


JÜLICH Forschungszentrum

JUST5

Physical view

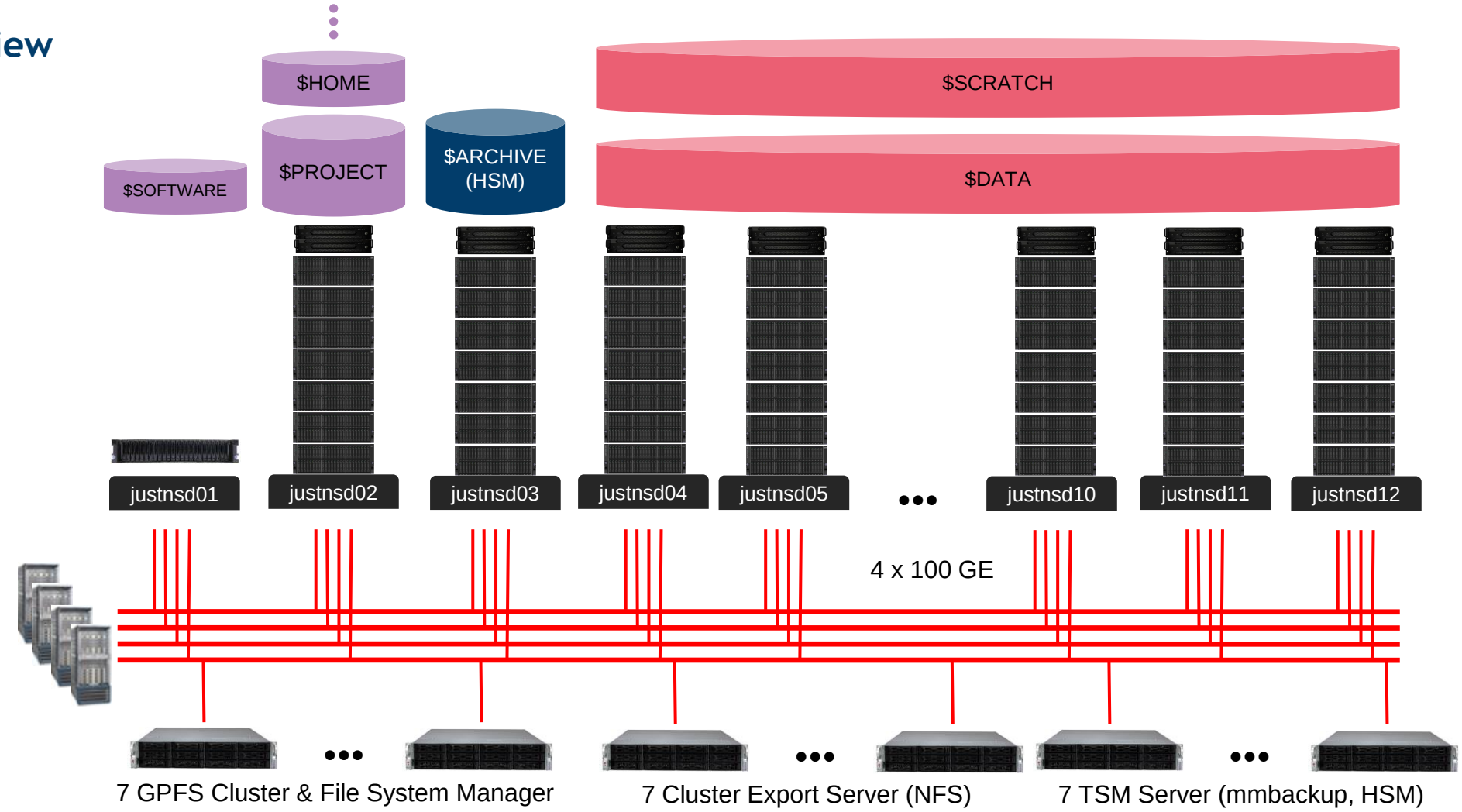
1 x DSS260
21 x DSS240
44 NSD Server
90 Enclosure
7500 10TB Disks
5 GPFS Manager
2 CES Server
2 TSM Server



JUST NEXT ITERATION

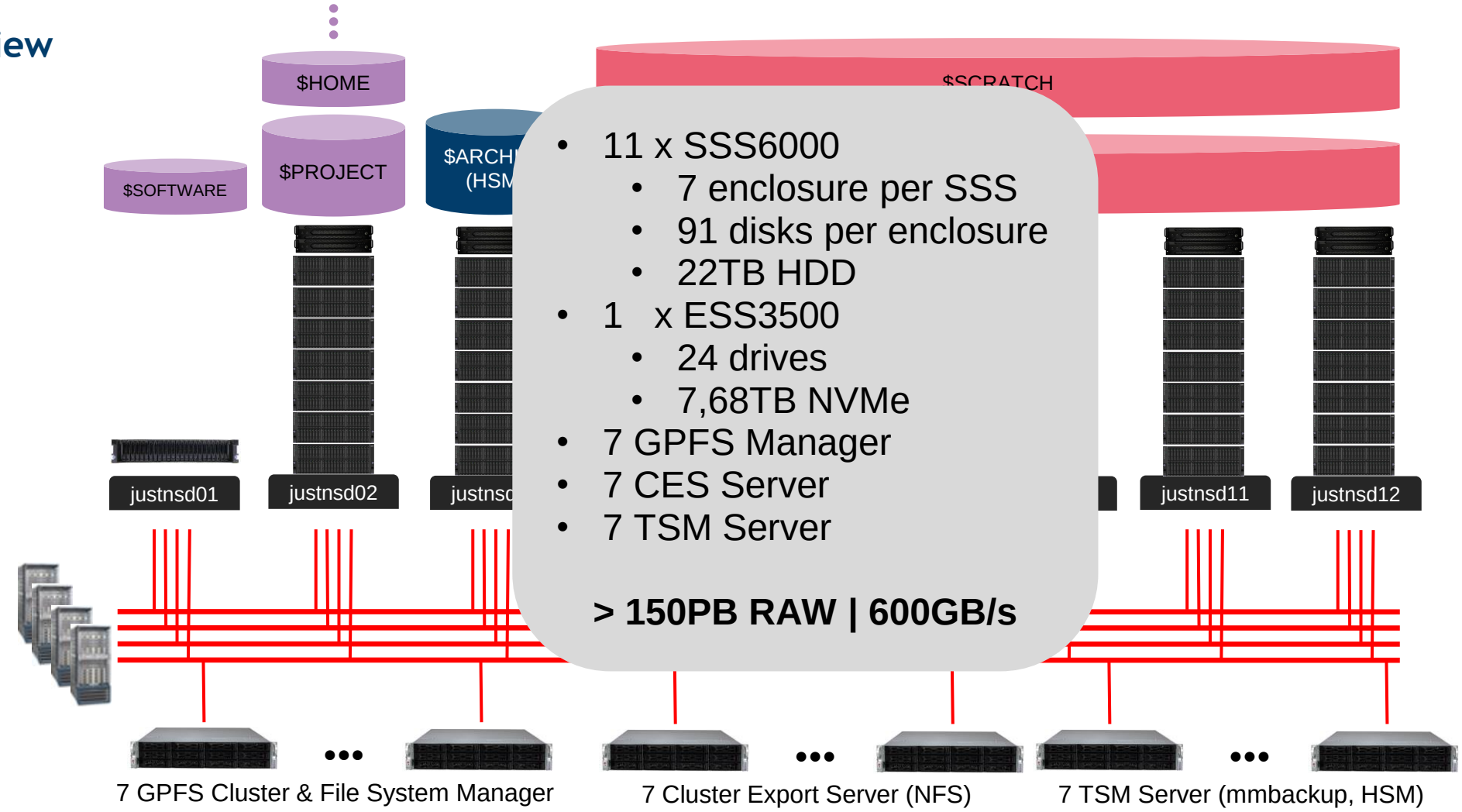
JUST6

Physical view



JUST6

Physical view

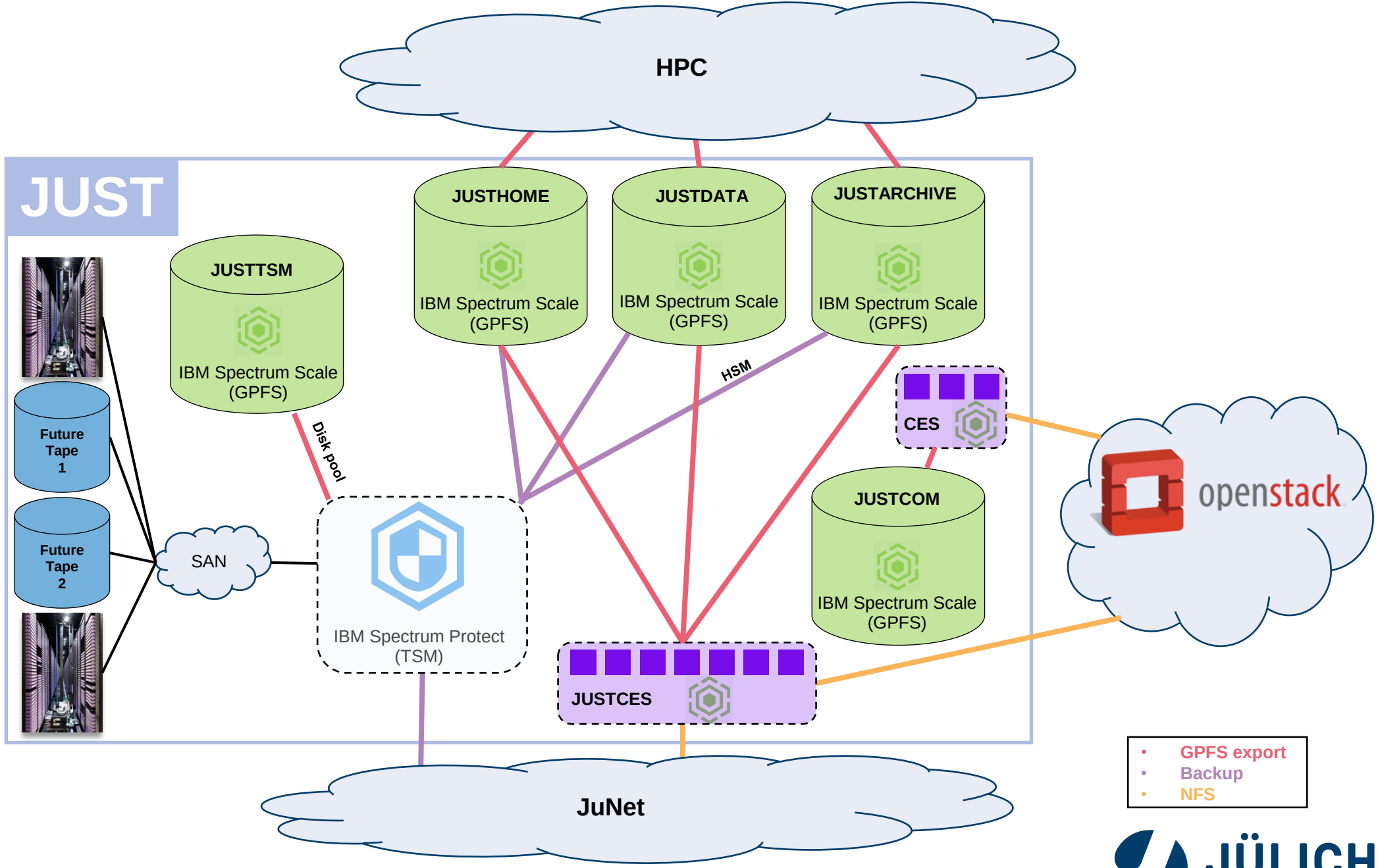


- 11 x SSS6000
  - 7 enclosure per SSS
  - 91 disks per enclosure
  - 22TB HDD
- 1 x ESS3500
  - 24 drives
  - 7,68TB NVMe
- 7 GPFS Manager
- 7 CES Server
- 7 TSM Server

> 150PB RAW | 600GB/s

JUST6

Logical view



JÜLICH Forschungszentrum

Filesystem	Usage	#inodes	Dest. Filesystem	GPFS Cluster	Migration method
arch	500TB (27PB)	22M	arch	justarchive	nsd restripe
arch2	500TB (21PB)	9M	arch2	justarchive	nsd restripe
scratch	9,5PB	420M	scratch	justdata	will be newly created
fastdata	6,6PB	111M	data1	justdata	AFM
largedata	15,6PB	111M	data1	justdata	AFM
largedata2	6,5PB	39M	data1	justdata	AFM
home	28TB	56M	home	justhome	copy
project	3,6PB	702M	project1	justhome	AFM
hpcmon	6TB	3M	hpcmon	justhome	copy
usersoftware	2TB	20M	usersoftware	justhome	copy
*_software	16TB	105M	software	justhome	copy
canary	<1TB	<1M	canary_<cluster>	one for each	will be newly created
cesSharedRoot	<1TB	<1M	?	?	

ACTIVE FILE MANAGEMENT (AFM)

Enables sharing of data across clusters

- Create association between two GPFS-Clusters or to a NFS data source
- Use cases:
  - One common namespace between two storages
  - Automate flow of data
  - Duplicate data
  - Caching data
- Only possible using independent filesets
- Backend protocol is NSD or NFS

https://www.ibm.com/docs/en/storage-scale/5.1.9?topic=reference-active-file-management-afm

JUPITER - THE ARRIVAL OF EXASCALE IN EUROPE

Member of the Helmholtz Association

EuroHPC Joint Undertaking

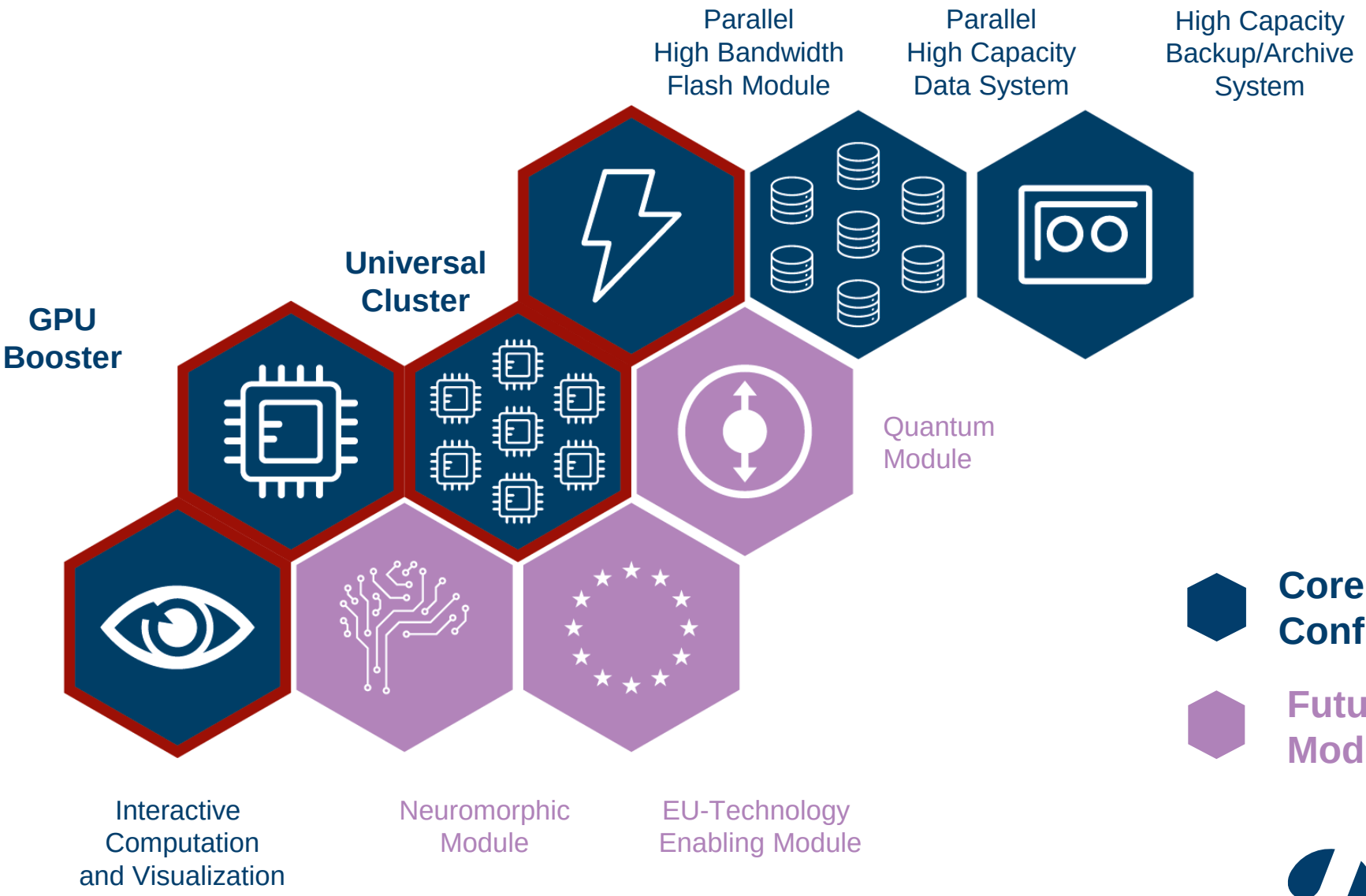
Bundesministerium für Bildung und Forschung

Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen

JÜLICH Forschungszentrum

Shaping Change

JUPITER – HIGH-LEVEL ARCHITECTURE



Core Configuration

Future Call Modules

JÜLICH Forschungszentrum

EXASTORE

- Same HW as JUST6 for flexibility to move HW
- 22 x IBM Storage Scale System 6000
  - 154 enclosures
  - 14014 x 22TB HDD
- 1 x IBM Storage Scale System 3500
  - 24 x 7.68 TB NVMe
- Data Movers to move data between ExaStore and JUST6

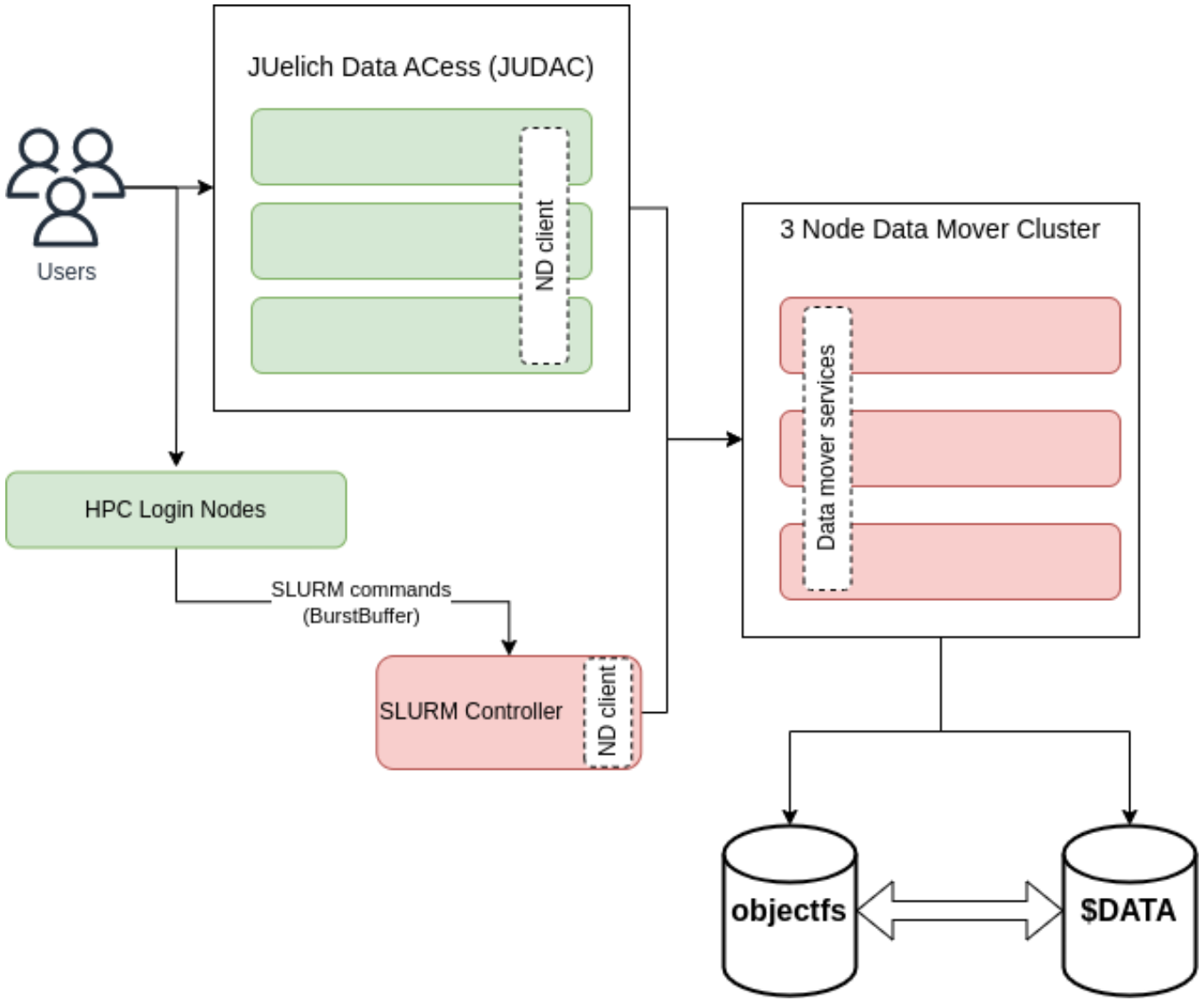
308 PB RAW | 1200 GB/s

DATA MOVER

Nodeum

A dedicated data mover with its own nodes triggered using various interfaces

- Deployed using ANSIBLE
- Allows for scaling by expanding data mover cluster
- Login nodes are not loaded with copy operations
- Various degrees of resiliency levels built into the Data Mover components



EXAFLASH

NVMe based storage integrated in the JUPITER fabric

- 20x IBM SSS6000
- **29 PB raw**
- **2,1 TB/s read, 3,1 TB/s write**
- 20 x 48 x 30 TB NVME
- IBM Storage Scale
- InfiniBand
- Data Movers to migrate data between ExaFlash and ExaStore (evaluating: AFM)

DESIRED FEATURES

- Stable Dynamic Page Pool → Better control of default size
- REST API → no root ssh needed
- Tape optimized backup of HSM migrated data **SPF-I-3049**
- mmbackup support for ACLs **SPF-I-2996**
- C-Group version 2 support for QoS **GPFS-I-903**
- Reliable Quota **GPFS-I-807**
- Capacity quota support for HSM managed file systems **GPFS-I-1001**
- Prometheus Exporter for zimon metrics **GPFS-I-889**
- Data integrity checksum on stored data **GPFS-I-851**

GREETINGS FROM THE STORAGE TEAM