CLAIX: the HPC environment at RWTH Aachen University for National HPC and Data Services

Matthias Müller, Thomas Eifert

IT Center | RWTH AACHEN UNIVERSITY

RWTH Aachen University

A leading university with strong research

- One of the leading Technical Universities in Germany (TU9)
- One of eleven German Universities of Excellence
- University ranked #5 in Germany in THE 2024
- Computer Science ranked #2 in Germany in THE 2024
- University is #1 in Germany creating start-up companies
- Central node in the German Initiative for Research Data Management (NFDI)
- Host of National High Performance Computing Center for Engineering Sciences (NHR4CES)

Studies and Teaching

Excellent Teaching, Learning and Assessment

- 47,078 Students
- 14,150 International Students
- 173 courses of study

Employees and Finances

- 10,272 Employees
- ~1.2 Billon Euro annual budget

IT Center @ RWTH Aachen University

Mission

IT-Service Provider for RWTH Aachen University

- From network infrastructure to HPC systems
- E-Learning and SLCM
- Responsible to support Research Data Management at RWTH

National Mission

- **HPC for Computational Engineering Sciences (NHR4CES)**
- **Important node of the NFDI network**
- **Infrastructure provide for AI Service Center WestAI**
- **Service provider for 42 universities in the state of North Rhine-Westphalia**

Staff and finances

- 360 employees
  (111 scientists, 130 staff, 46 apprentices, 74 students)
- 50+ M€ annual budget

Long history of HPC at RWTH Aachen University

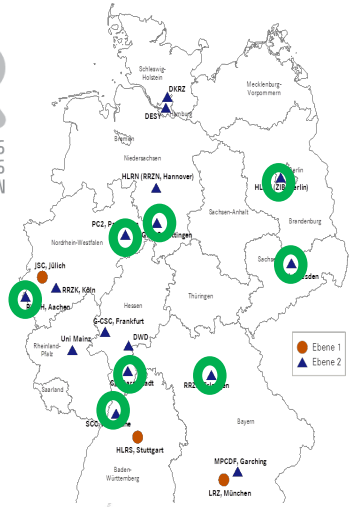
- 1958: Zuse Z22
- 1959: Siemens 2002
- 1966: CDC 6600 (1st installation in Germany, 2nd in Europe)
- 1976: CDC Cyber 175
- 1989: IBM 3090/600S/VF
- 1993: Fujitsu/SNI: S600/20 (VP2600/10) – fastest in Germany
  55th fastest supercomputer in the world (TOP500 06/1993)
- 2001: SunFire 6800
- 2002: SunFire 15k
- 2008: Fujitsu Siemens HPC Cluster
- 2011: Bull Cluster (#32 in TOP500 06/2011)
- 2016: NEC Cluster „CLAIX 2016“
- 2018: NEC Cluster „CLAIX 2018“
- 2023: NEC Cluster „CLAIX 2023“
- 2025: NEC Cluster „CLAIX 2025“

Abb.: Der neue Höchstleistungs-Universalrechner IBM 3090/600S/VF. Foto: Wolfgang W. Nitsche

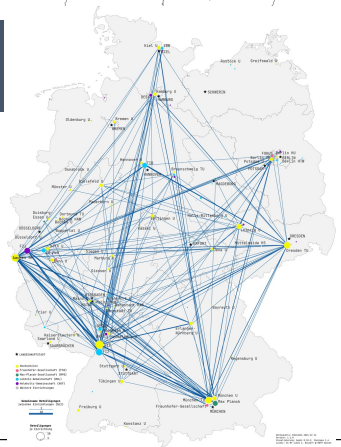
VP2100 series, Source: https://museum.ipsj.or.jp

CLAIX 2023 from NEC at RWTH Aachen University

Two national initiatives: NHR (National HPC) and NFDI (National Research Data Management)

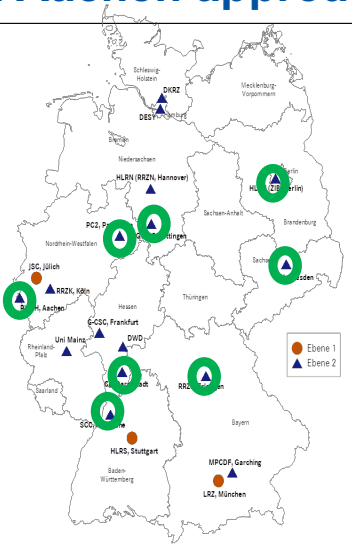


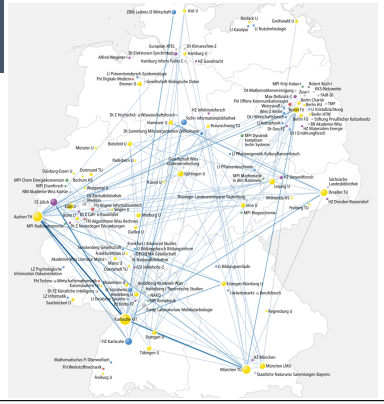
- **National High Performance Computing (NHR)**
  - Tier-2 HPC systems
  - Nine academic centers for scientists from universities
  - Funding pays for HPC systems, operation costs and staff
  - 9 selected centers:
    - **Aachen**, Berlin, Erlangen, Darmstadt, Dresden, Goettingen, Southwest, Karlsruhe, Paderborn



- **National Research Data Infrastructure (NFDI)**
  - 26 consortia for different scientific fields and one consortium for basic services
  - NFDI pays for staff, but not for IT infrastrcucture
  - The NFDI projects build a strong network of 190 partners in Germany
  - RWTH is one of the central nodes of the network

RWTH Aachen approach: two major platforms to integrate HPC and RDM

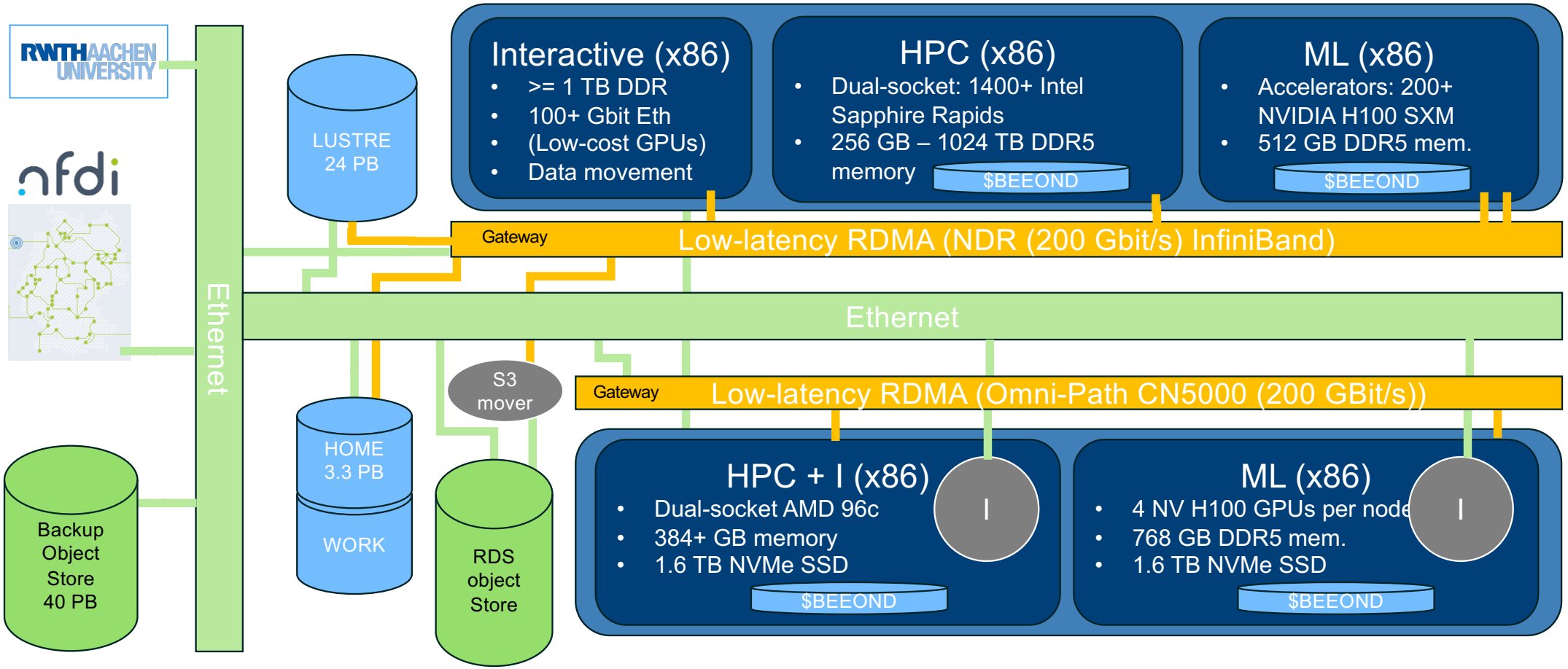




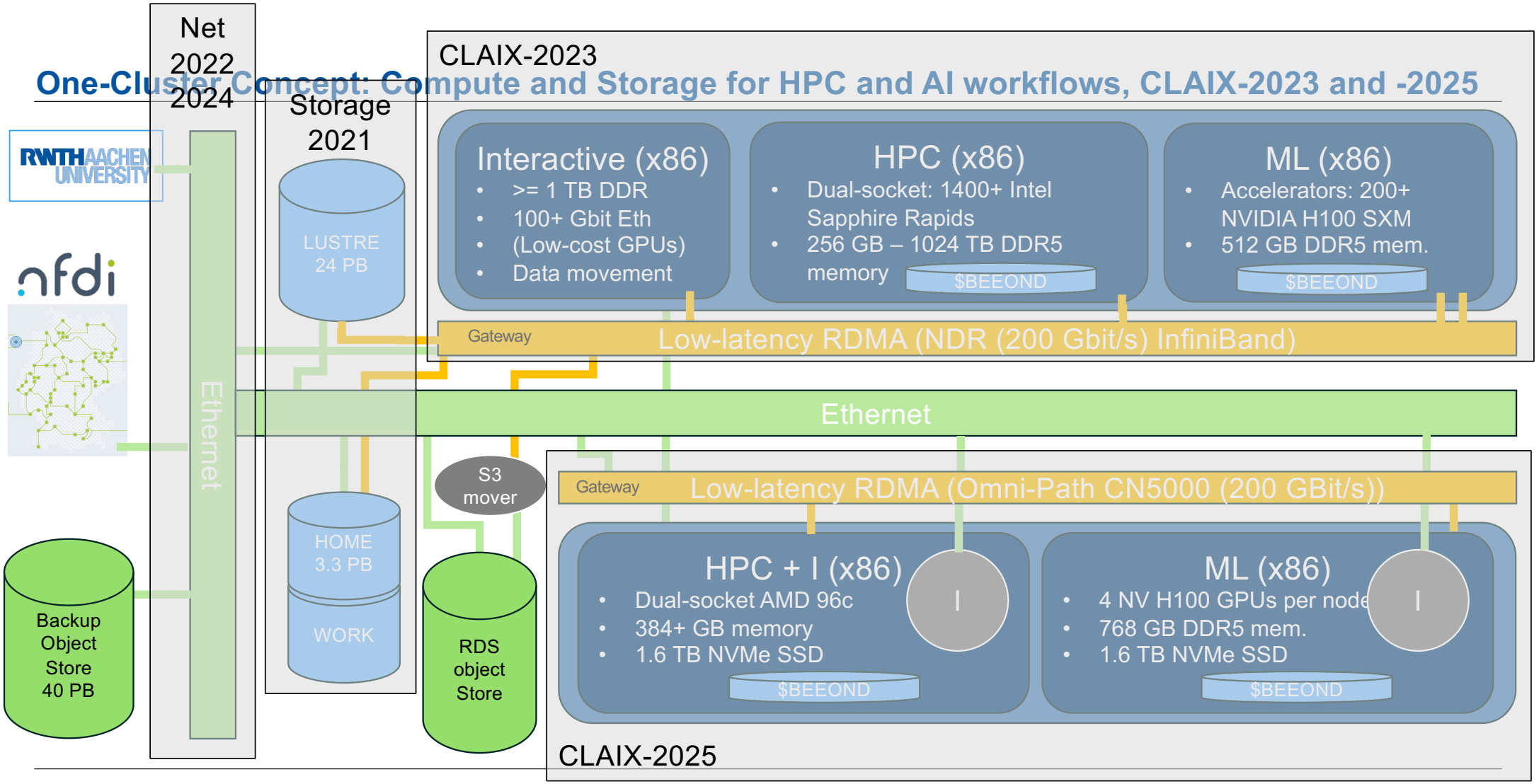
- **An integrated One-Cluster Concept for HPC**
  - Different Segmens for
    - Classical HPC
    - Maschine Learning
    - Interactive usage, Jupyter Hub
  - Different storage classes for specific needs
    - Home file system for secure, permanent storage
    - Project file system for high bandwidth demand
    - On demand file system for high metadata performance

- **Coscine – the data management platform for FAIR data**
  - a data storage and linking environment for arbitrary data sources
  - Datastorage.nrw as a federated storage backend
  - implements FAIR principles based on the FAIR DO concept

One-Cluster Concept: Compute and Storage for HPC and AI workflows, CLAIX-2023 and -2025



One-Cluster Concept: Compute and Storage for HPC and AI workflows, CLAIX-2023 and -2025

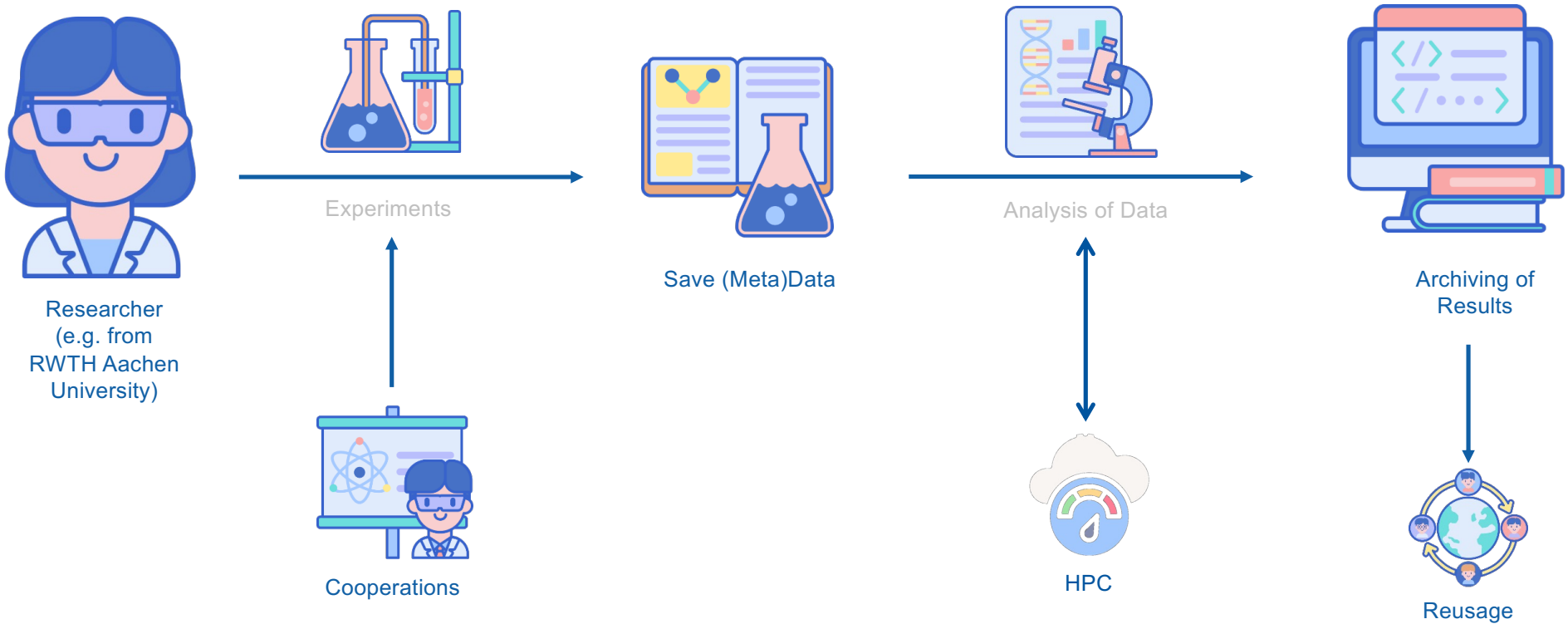


Evolution of the One-Cluster Concept with Open Source

Feature	Bull Cluster	CLAIX-2016	CLAIX-2018	CLAIX-2023	CLAIX-2025
Integrator	Bull	NEC	NEC	NEC	NEC
SW stack by	IT Center	IT Center	IT Center	IT Center	IT Center
OS	Scientific Linux	CentOS	CentOS	Rocky Linux 8.x	Rocky Linux 9.x
Batch System	LSF	Slurm	Slurm	Slurm	Slurm
Deployment via	Ethernet	Ethernet	Ethernet	IB	Omni Path CN5000
RDMA network	InfiniBand-40	Omni-Path-100	Omni-Path-100	InfiniBand-400	Omni Path CN5000
Home FS	Netapp (NFS)	Isilon (NFS)	Isilon (NFS)	DDN (GPFS)	DDN (GPFS)
Parallel FS	DDN (Lustre)	NEC (Lustre)	DDN (Lustre)	DDN (Lustre)	DDN (Lustre)
OnDemand FS	-	-	BeeGFS on local SSD	BeeGFS on local SSD	BeeGFS on local SSD
Host processor	Intel	Intel	Intel	Intel	AMD
Accelerators	-	NVIDIA	NVIDIA V100	NVIDIA H100	NVIDIA H100
Container support	-	-	NVIDIA + any singularity	NVIDIA + any apptainer	NVIDIA + any apptainer + secure container

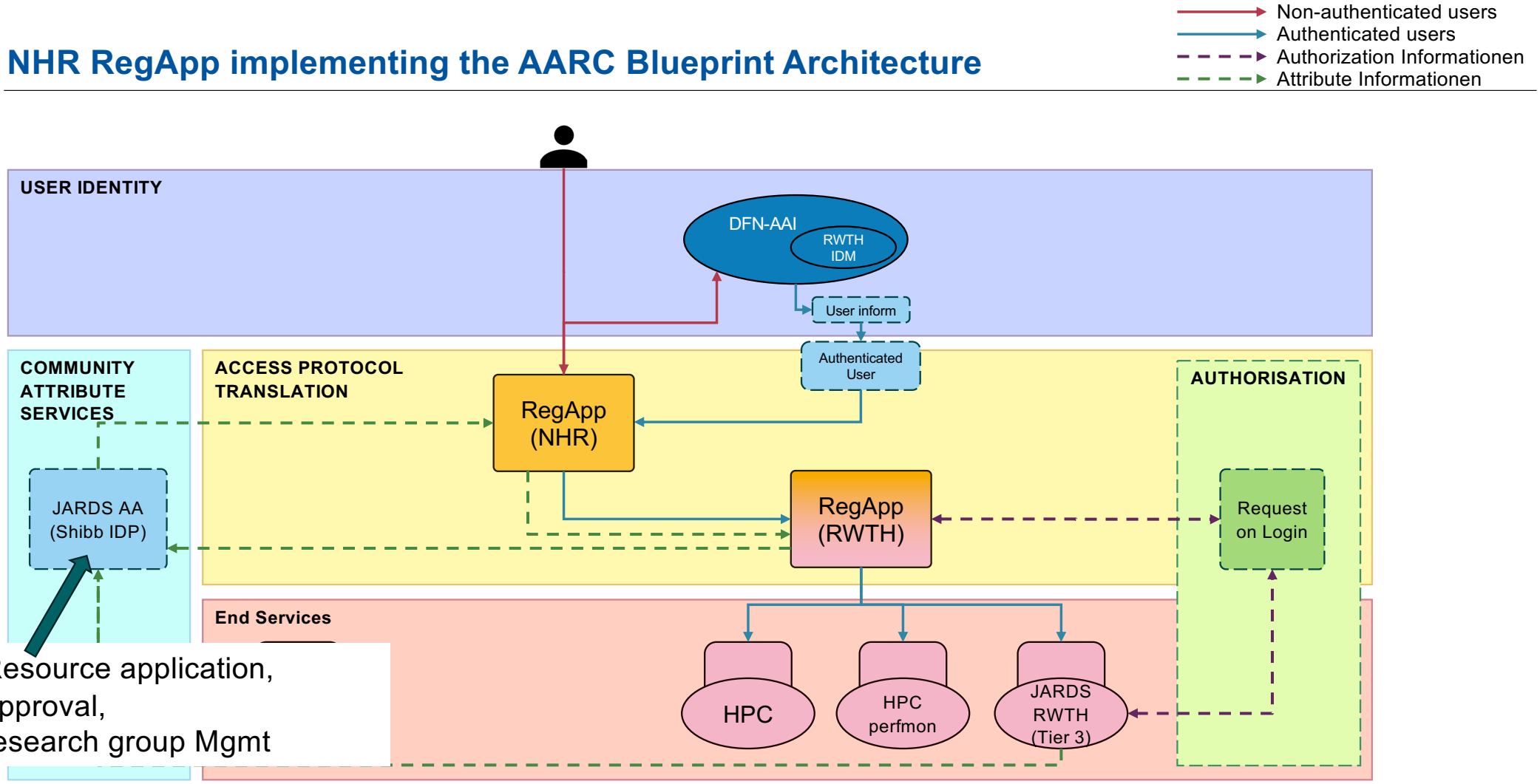
Coscine – Collaborative Scientific Integration Environment

Example for Coscine Usage



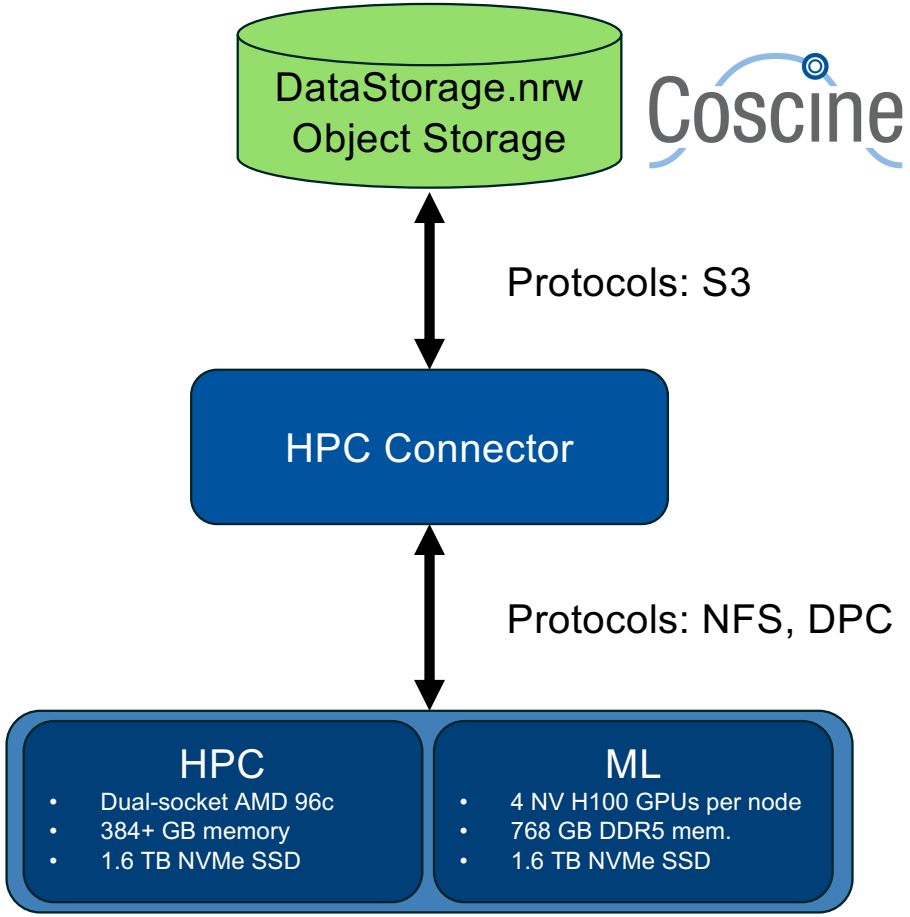
Icons by Flaticon (Flat Icons)

NHR RegApp implementing the AARC Blueprint Architecture



What is the HPC Connector?

- The HPC-Connector is a small storage placed between our Object Storage and the HPC Cluster
  - ~ 300 TB Storage on the HPC Connector (all Flash)
  - ~ 25 PB Storage on the Object Storage (HDD)
- It has an excellent connection to the cluster
  - 26 * 200 Gb/s InfiniBand (13 nodes)
- In the transfer test between the HPC-Connector and the HPC Cluster we archived
  - NFS ~ 60 GB/s, DPC ~ 260 GB/s
- The connection between the Object Storage and the HPC-Connector allows automatic synchronization of changes



Research Data Management Platform Coscine – Core Features

Science led allocation and provisioning of storage

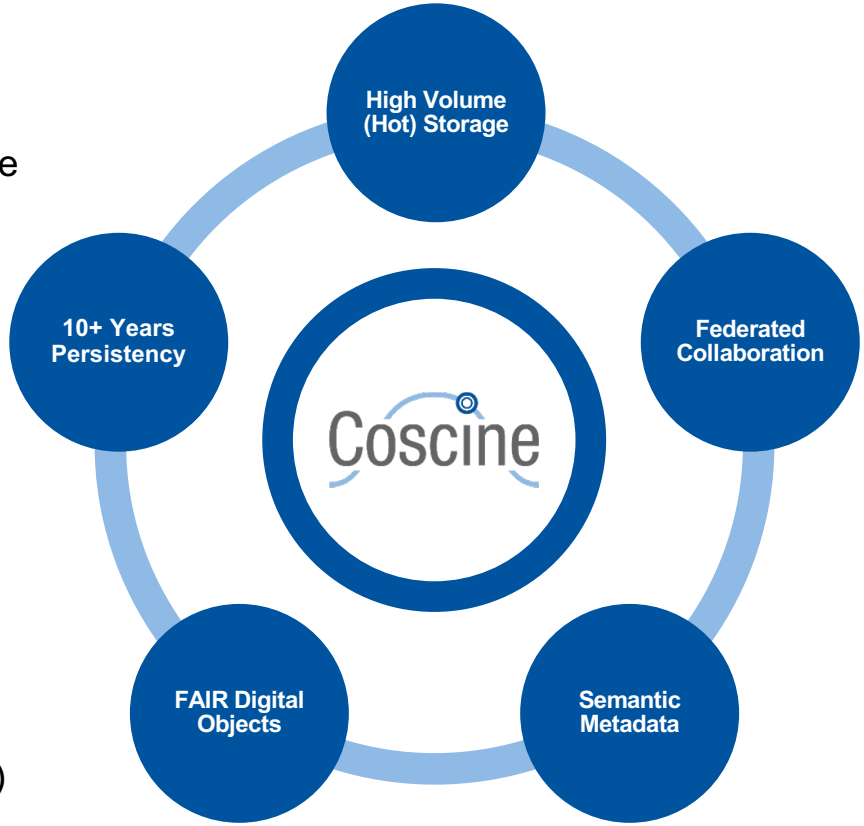
- Storage applications via peer review process
- Persistent ID & FAIR Digital Object layer on top of industry grade storage

Federated Access management

- Low threshold institutional login via DFN-AAI/eduGAIN (SAML)
- External (industry) collaborators via ORCID (OIDC)

Data annotation according to semantic standards

- Validation using discipline specifc metadata profiles (SHACL)
- Access via standard protocols: Linked Data Platform & SPARQL
- Connection to International Data Spaces / EOSC / Gaia-x (DCAT / EDC)



Structured Data Management with Coscine - Reporting

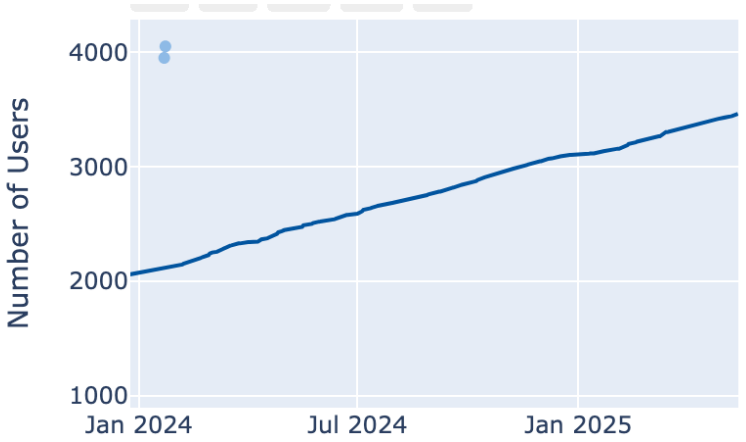
NUMBER OF ORGANISATIONS
204
(+1 in the last week)

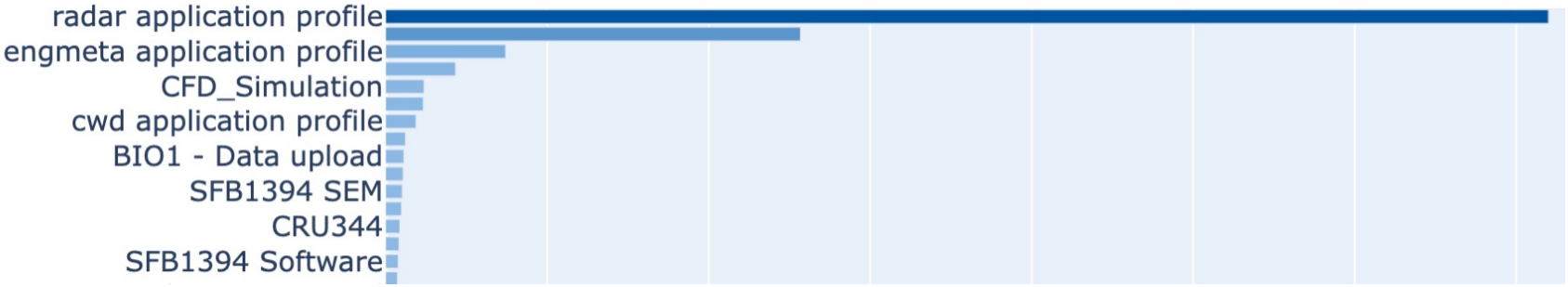
NUMBER OF PROJECTS
4,880
(+3 in the last week)

NUMBER OF USERS
3,447
(+13 in the last week)

Increasing use of Coscine

- *3500 Users (+20/week)*
- *200+ Organisations*
- *5000 Projects (+25/week)*
- *6000+ Datasets (S3 Buckets) (+26/week)*
- *5,7PB (+50TB/week)*
- *Large number of metadata profiles:*





Summary

- RWTH Aachen is a good place to combine research data management and HPC
  - National Center for HPC with a long tradition
  - Strong node in the National Research Data Infrastructure (NFD) participating in many projects
- Integration of reseach data management and HPC is a focus area
  - User management from application process to identity and access management
  - Fast migration of data between external storage/RDM storage to HPC system
- Coscine is used as research data management platform
  - Developed at RWTH
  - Rapid increase of usage
- The combination of open source software and suitable infrastructure was able to combine RDM and HPC
  - HPC cluster operated with open source software
  - Coscine as open source software

Vielen Dank
für Ihre Aufmerksamkeit