SURF: site update

Data Processing services,

Role in the LTA,

And future plans

Lodewijk Nauta

23 September, 2024

General update

Snellius (HPC: https://visualization.surf.nl/snellius-virtual-tour/#/)

Simulations and modelling that not only demand a lot of computing power and memory but also a lot in terms of communication between the various processors

Supercomputer:
The most powerful HPC system for research in NL
(peak performance 38 Pflop/s – phase 3)

Machine Learning:
AI workloads with fast processors (higher RAM) and GPUs (higher training throughput)

Energy efficiency and sustainability are key drivers

Broad software stacks built with EasyBuild

Many cores (predominantly AMD)
Large symmetric multi-processing nodes
High memory nodes (4 TB and 8 TB)
A fast interconnect
A lot of work space on disk
or a fast I/O subsystem

Phase – 3 (H100 GPUs) has just arrived !

Snellius (phase 3 adds another 88 nodes with 352 H100 GPUs)

System at a glance

1.396	214K	408	>24PF
Compute nodes	CPU Cores	GPUs	Rpeak

580 + 21
AMD ROME CPU compute nodes

36 + 30
Nvidia A100 GPU compute nodes

76,8K + 2.7K Cores

144 + 120 A100 GPUs

3PF +0.15PF CPU Rpeak

2,8PF + 2,3PF GPU Rpeak

Phase 1

NL-SAS, NVMe and SSD storage

>13.3PiB Usable storage capacity

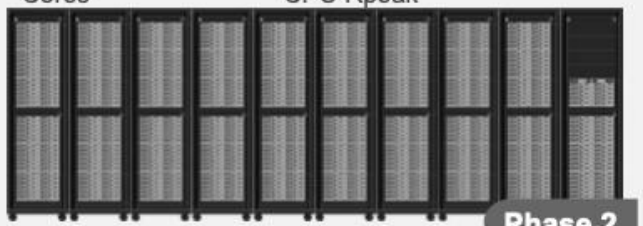
>300GiB/s Sequential perfromance

Storage

714 + 72
AMD NG CPU compute nodes

137K + 14K Cores

5,63PF + 0,5PF CPU Rpeak



Phase 2

66 + ?
Nvidia Hopper GPU compute nodes

264 + ? GPUs

13PF + ? GPU Rpeak

Phase 3

HDR and NDR
Fat-Tree Mellanox infiniband fabric

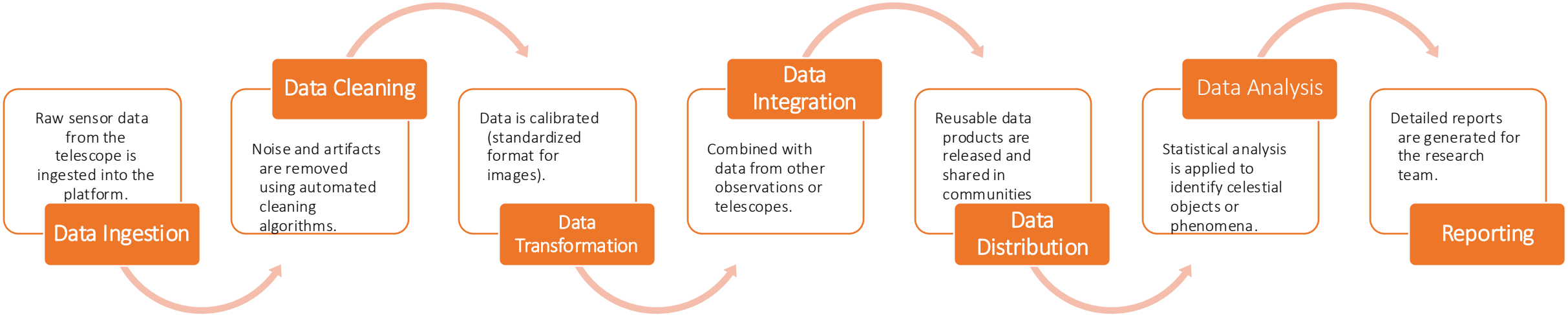
High speed network

SURF

Data processing services

Intro: Data processing life-cycle

From project initiation to completion,
we offer comprehensive support and consulting.



Data Processing proposition

Data Processing provides **comprehensive platforms** for researchers.

Capabilities:

Processing, analyzing, and managing large-scale data

Scientific instruments:

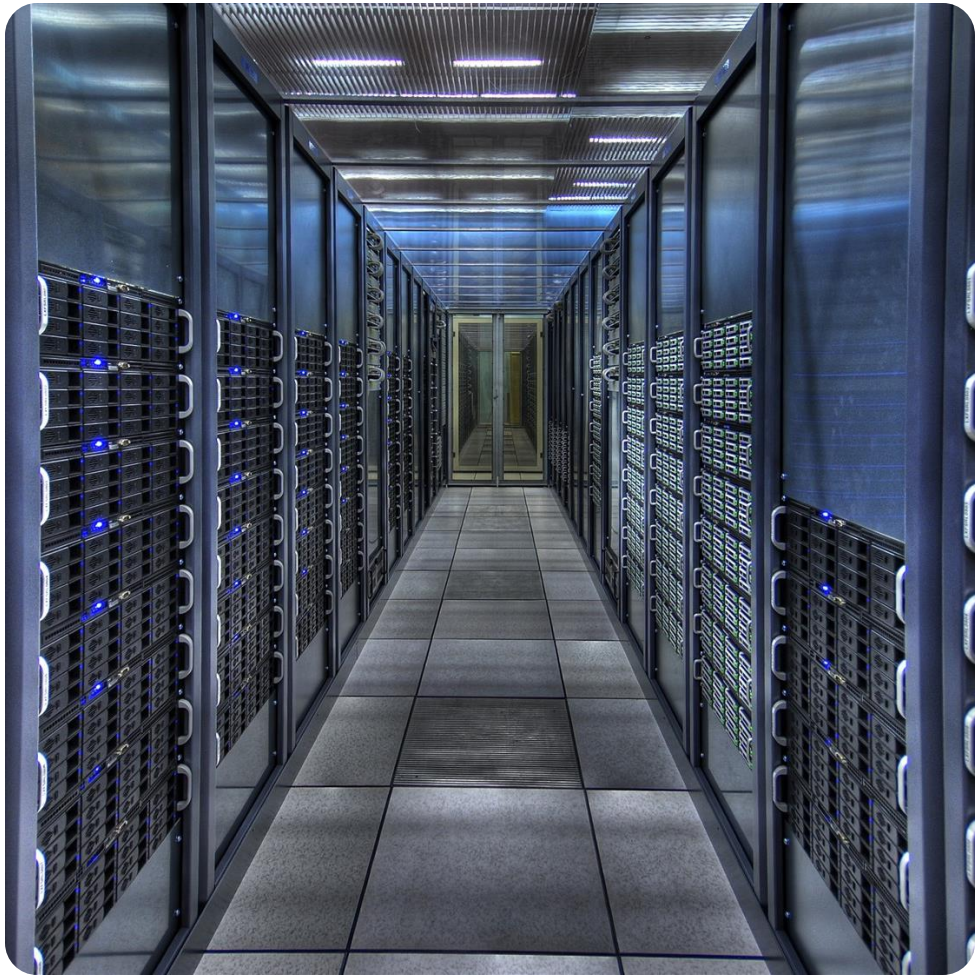
Telescopes, genetic sequencers, particle detectors, satellites

Data Processing services:

Spider: Cloneable, cloud based local compute cluster at SURF

Grid: Federated international collaboration between compute centers
(Implemented as a “clone” of Spider at SURF)

dCache: storing and retrieving huge amounts of data



Goal of Data Processing

The Data Processing team **facilitates** scientists in building and running **data-intensive** workflows for **multi-year** projects

“Empowering scientists with long term Data Processing solutions”

Cloud Research Consultancy (CRC)

- Tailor-made solutions to solve specific research problems
- Co-development with researchers and research supporters
- Private, public or hybrid cloud infrastructure
- Long-term, sustainable production support
- Used to be MS4 and SDA

Cloud Research Consultancy (CRC)

Examples:

- Big Data analytics
- Streaming data
- Internet of things
- Machine Learning

Spider & Grid (HTC)

Data intensive projects processing instrument data from sensors, sequencers, telescopes, and satellites during the entire mission lifetime

Data volumes:
parallel processing of large amounts of data, from many Terabytes to Petabytes

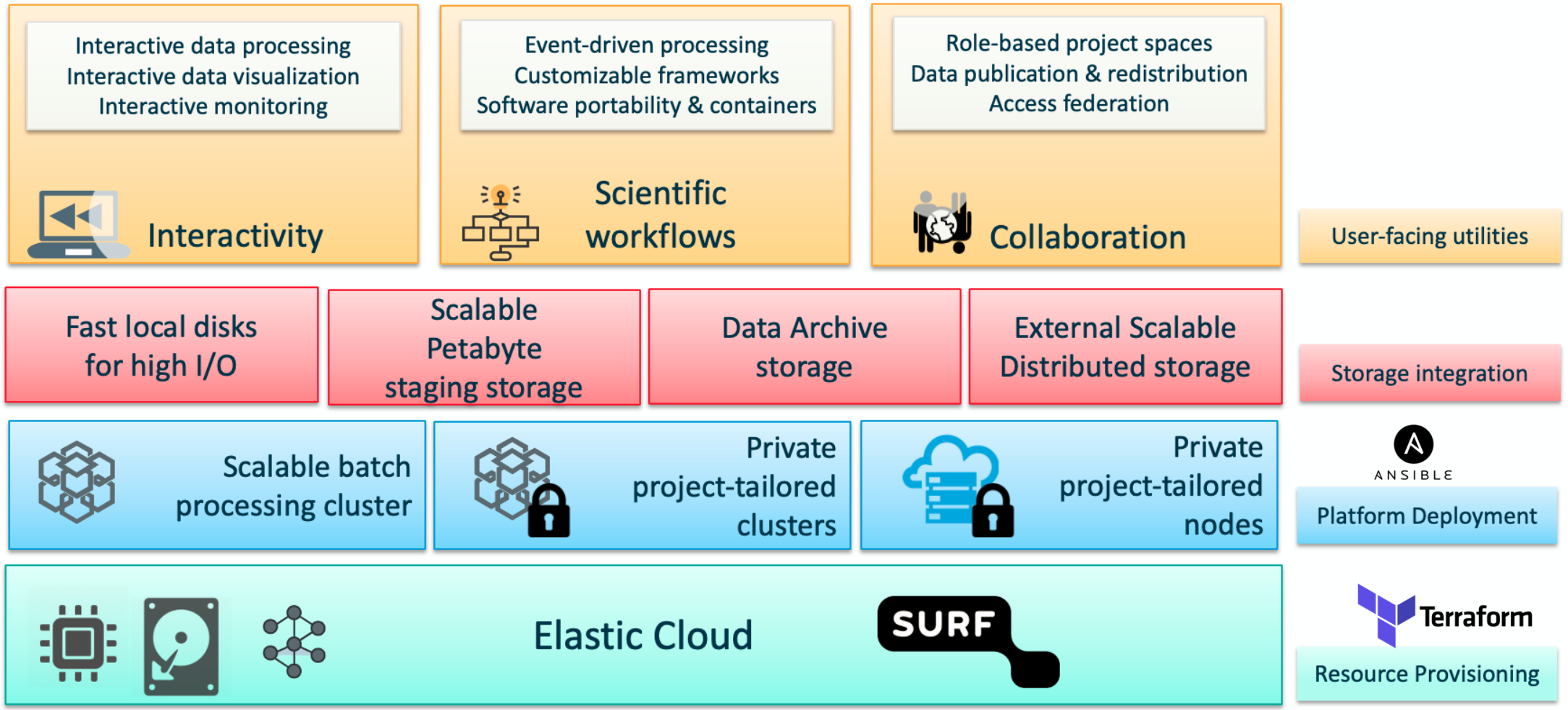
Processing pipelines:
steady production workflows with semi-continuous data flows

Project organisation:
international collaborations working on a shared set of data and software

Ecosystem:
modern cloud-based solutions with automated deployment, optimized for data processing

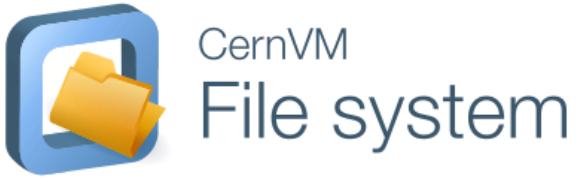
Built on the internal, elastic Cloud (OpenStack)
Fast network to external sources/dCache (1200 Gbit/s EVPN)
External connectivity for each compute node (2x25Gbit)
Fast local disks (up to 12 TB NVME SSD)
Powerful compute nodes (64 core, up to 1TB RAM)
GPUs: 16 A100 and 16 A10 GPU's

Spider architecture



Grid Services

- File Transfer Service (FTS)
  The service moves data from location A to
  location B using a third-party copy method
- perfSONAR
  performance monitoring of network and computer systems
- DIRAC
  Interware solution for distributed heterogeneous resources
  Job submission, storage management, catalogues
- CVMFS / softdrive.nl
  Software distribution platform
- ARGO
  Infrastructure monitoring
  (used for monitoring of Spider clone)



Data Processing projects examples and scale

Over 89PB* of storage and over 79** million core hours are consumed together by various scientific domains in April 2024

	WLCG Worldwide LHC Computing Grid	Amsterdam UMC Alzheimercentrum	LOFAR
Core hours / year allocated	38 mil	3 mil	6.6 mil
Domain	high energy physics	genomics	radio astronomy
Disk used	13 PB	550 TB	1.4 PB
Tape used	48 PB	1.8 PB	23 PB

Part of the LOFAR radio telescope Credit: RUG

DC24 (WLCG data challenge 2024)
Connection between CERN and Amsterdam (NIKHEF and SURF)
Atlas sending data with FTS
661 Gbit/s reached (target was 400 Gbit/s)

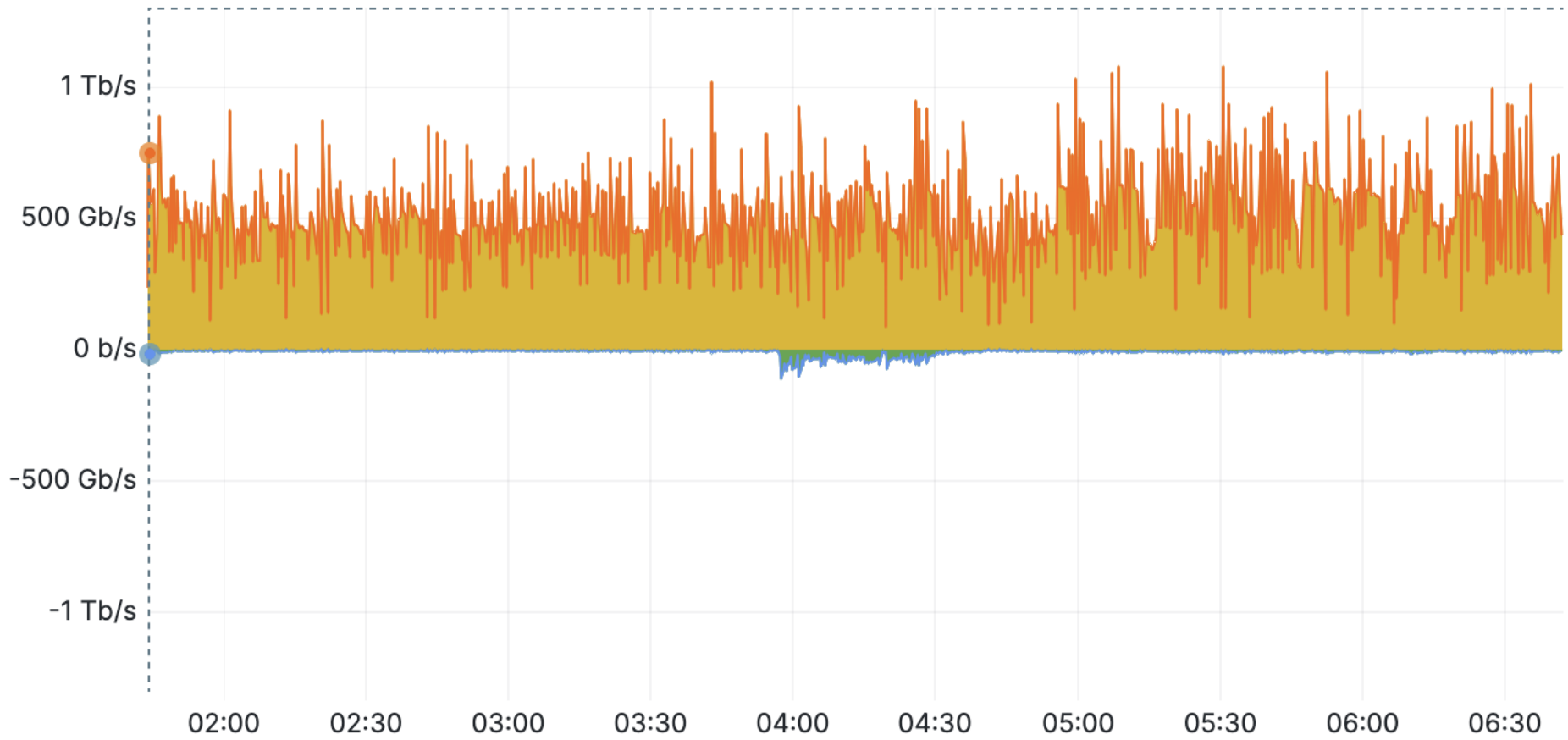
*(3.1+18+68) PB between CephFS storage, dCache disk, and tape

**(26 + 53) M c-h between Spider and Grid

DC24 (WLCG data challenge 2024)

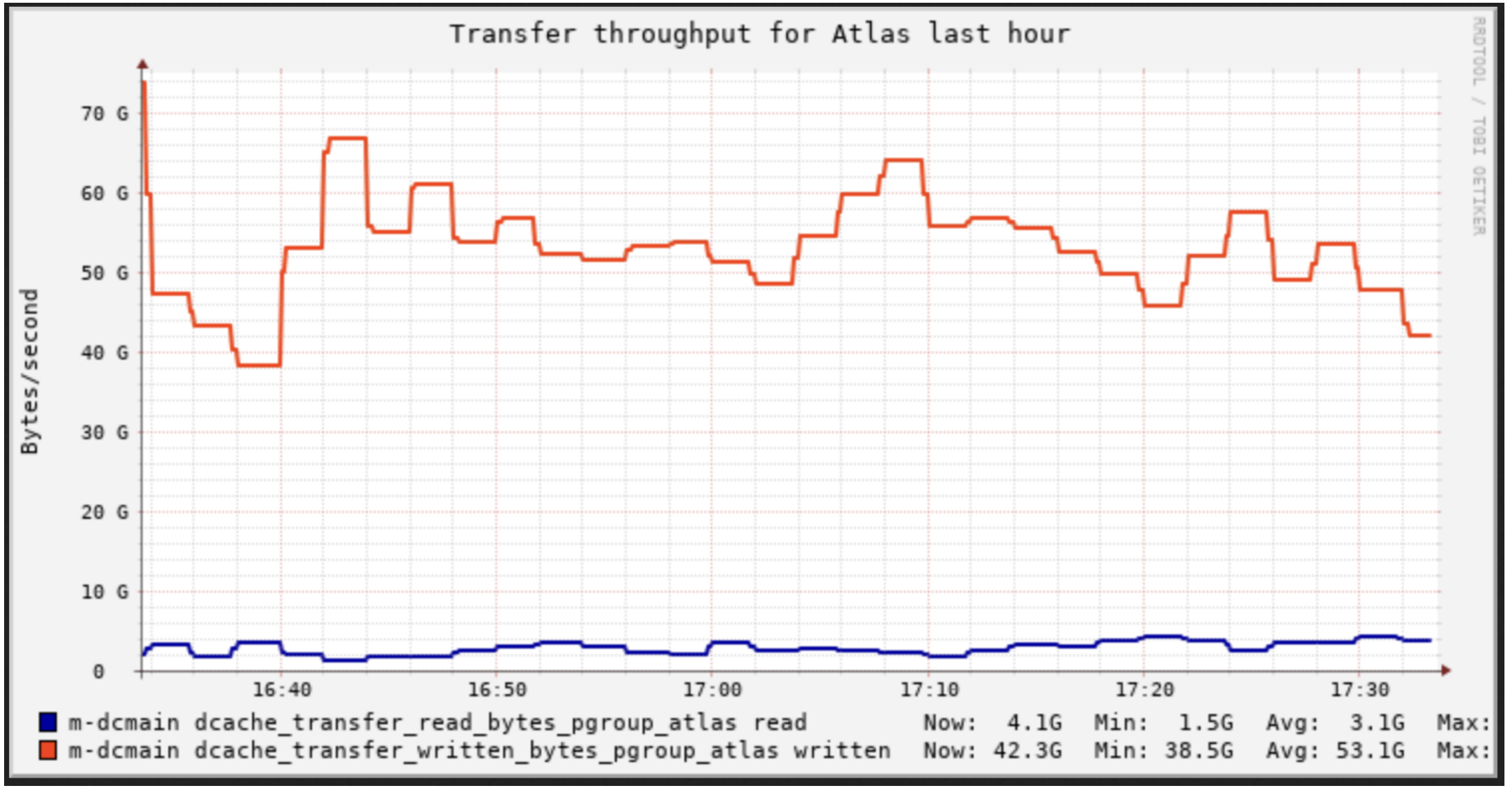
- Test: 800 Gbit/s connection between CERN and Amsterdam (NIKHEF and SURF)
- Nokia network equipment
- 1648 km fiber
- Atlas sending data with FTS from EOS to NIKHEF and SURF
- Using 101 pools at SURF
- 661 Gbit/s reached (target was 400 Gbit/s)

Total throughput ASD-GEN (stacked)



Name	Mean	Last *	Max
lag 1 - RX	12.0 Gb/s	7.07 Gb/s	115 Gb/s
lag 1 - TX	504 Gb/s	434 Gb/s	1.08 Tb/s
lag 2 - RX	0 b/s	0 b/s	0 b/s
lag 2 - TX	0 b/s	0 b/s	0 b/s

DC24 Nokia networking test: 53 GB/s average for an hour



Success!

dCache could handle the load with ease.

But we can make it even better...

International collaborations

Data Processing users work on **multi-year** projects and can be **international** scientific groups.

“Fostering global cooperation and enabling groundbreaking research”

Cloning Spider

Spider runs on SURFs internal cloud DAPHNE.

This setup enables:

- Clonable building blocks
- Rapid, customizable, scalable deployment

Allowing the creation of ‘clone instances’ for specific use-cases:

- Large projects: GRID, SKA
- Require strict security measures:
  Alzheimer’s Genomics Hub (AGH)

Spider clones

The platforms' (internal) **cloud based deployment** allows for easy **cloning and customization** of the data processing **platform**

"Ensuring scalability and improving resources utilization"

Interoperability

To serve many different communities Data Processing services are set up to be **interoperable**.

“Through interopability, national and international communites are connected”

Data Processing Services in a nutshell

Data Processing

- Provides consulting from **project start to finish**
- **facilitates** scientists in executing **data intensive multi-year** projects
- where services are **interoperable** to serve many communities
- used by **international** scientific groups

Cloud based deployment allows for

- easy **cloning** of the platform for specific use-cases

Many projects!

Project Mine is granted with new Computing Time

12-02-2022

Via the call Computing Time on N Facilities by NWO, Project MinE is computing time and [...]

Read more ›

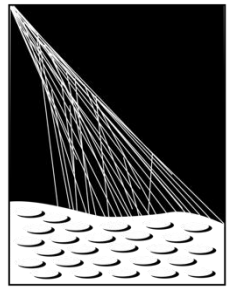
project MinE
Make it yours

PACE
BLOG

Plankton, Aerosol, Cloud, ocean Ecosystem

PACE Mission Launch: A Milestone in Earth Observation

SPEX one



PIERRE AUGER OBSERVATORY

DUNE
DEEP UNDERGROUND NEUTRINO EXPERIMENT

LHCb

NOS Nieuws • Dinsdag 19 februari 2019, 21:58

Onderzoekers brengen 300.000 sterrenstelsels in kaart

LOFAR

KM3NeT

Alzheimercentrum Amsterdam
Amsterdam UMC

nieuwsuur

Woensdag 10 augustus 2022, 20:00

Methaansatelliet vindt vuilnisbelt met klimaatimpact van 1,5 miljoen auto's

TROPOMI

XENON

ALICE

VIRGO

ATLAS EXPERIMENT

SURF

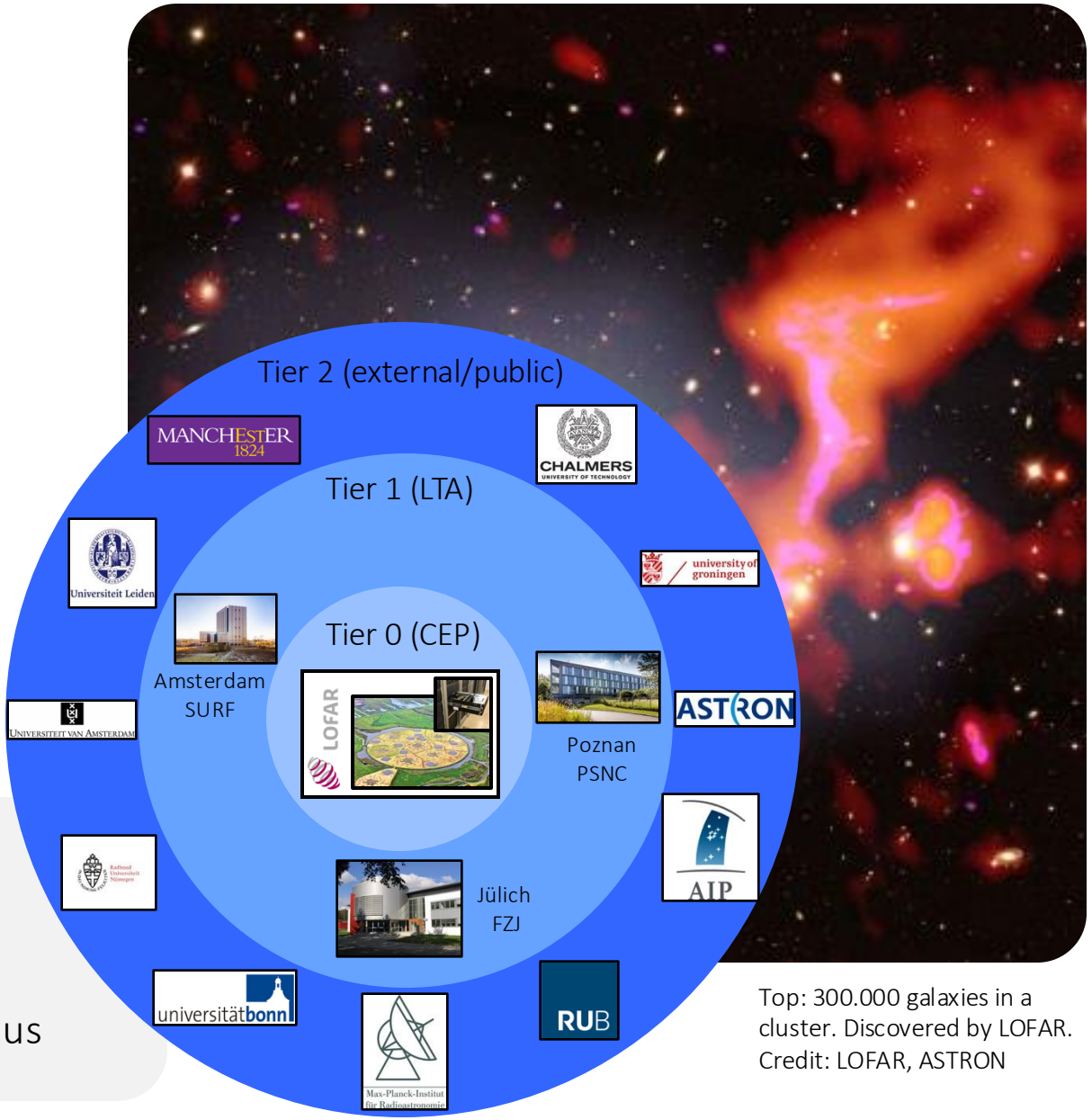
Role in the LTA

Infrastructure provider, service co-development, community support

SURF supports LOFAR projects – Example SKSP

LOFAR Surveys Key Science (LoTSS) Project:
Transform massive amounts of data into high-quality images

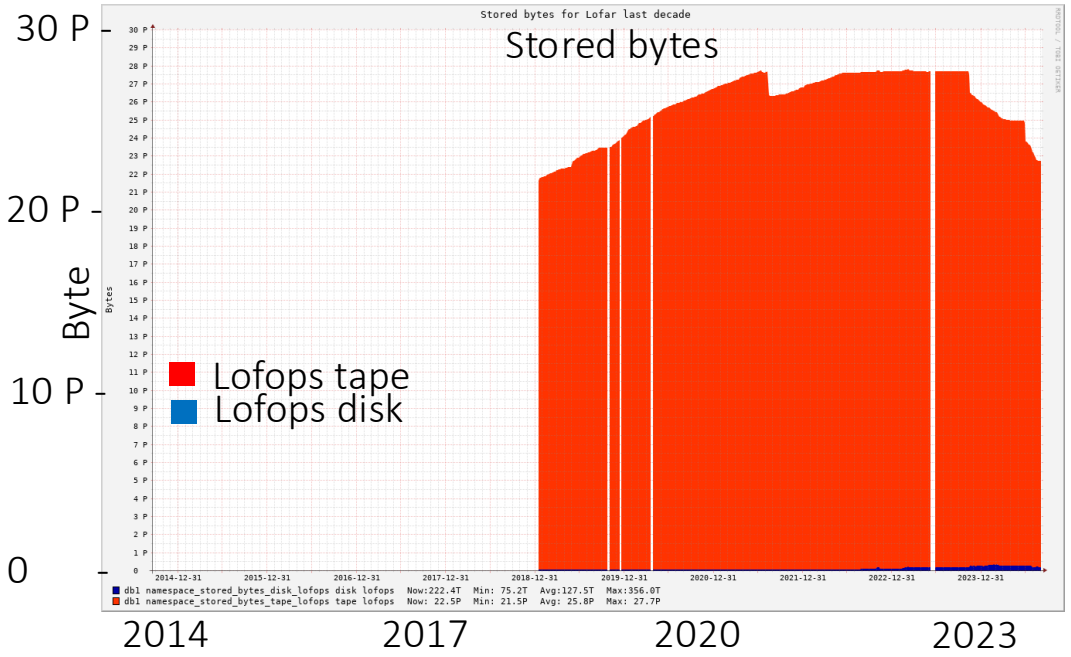
- 300 members
- 60 institutes
- 18 countries
- 175 papers
- Distribution of data products
- Processed over 1000 LOFAR observations (4M core hours, 300TB disk and 700TB tape)
- Science ready data products: publicly available through SURF data repository

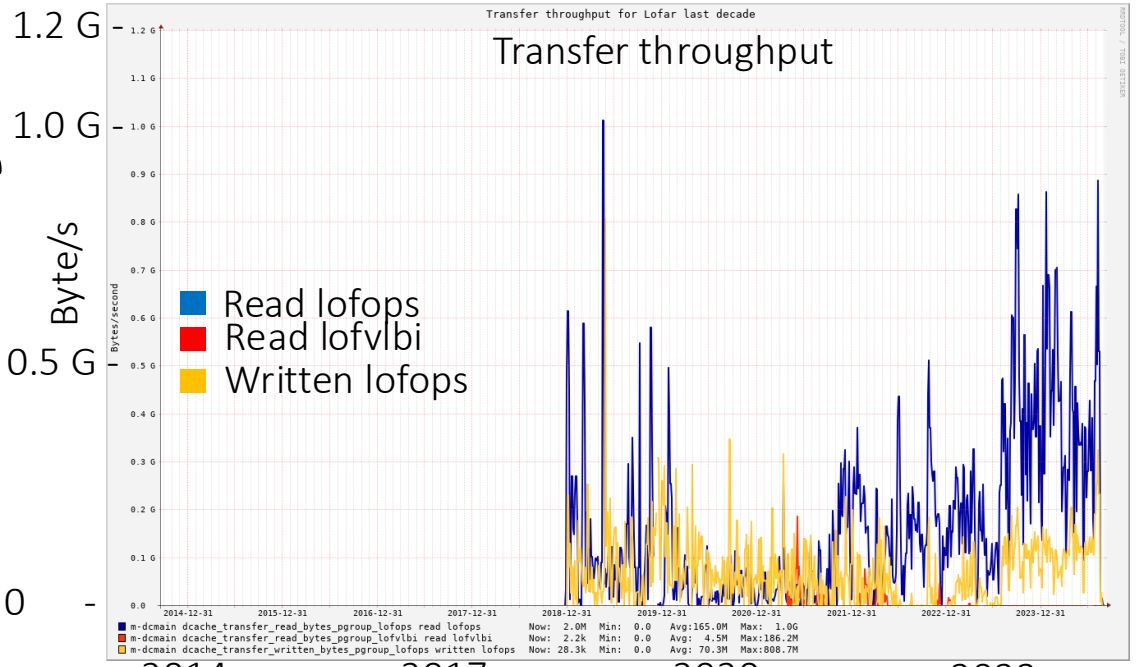


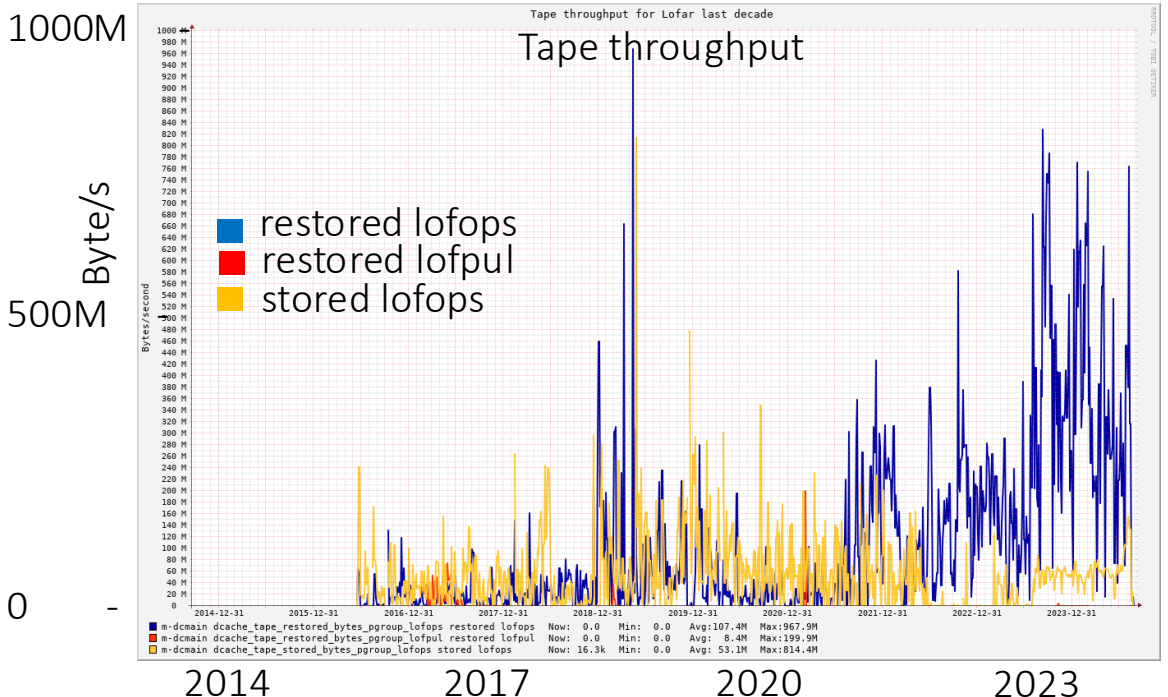
Top: 300.000 galaxies in a cluster. Discovered by LOFAR. Credit: LOFAR, ASTRON

Compute, storage, services
23 PB stored at SURF
High-speed connection to the LTA
Processing power: Grid, Spider, Snellius

LTA numbers on SURF dCache







Examples of different Lofar workflows

SURF contribution

Type	Data acquisition workflows	LTA operations workflows	1st level End user workflows	2nd level End user workflows
Description	Correlator processing	LTA data management	LTA data (pre)processing	Science ready products analysis
Examples	Data classification Bad data deletion …	Data Ingestion Quality checks Searching data Data Staging Data Retrieval Data Reduction (LDV, Rapthor) Data Deletion …	LoTSS/LoTSS deep/I-LoTSS Lofarvwf Lofarvlbi Lspc NLsrc …	Products discovery Combining images SKSP WebDAV …

SKA (Square Kilometre Array)

SKA clone on Spider

Joined SKA challenge ‘25

Joined test datalake

First dCache site to join SKA

OIDC token authentication

⇒ Raymond talk today:
The SKA Regional Centre Network

⇒ Onno talk today:
Token authentication in dCache

Application Roadmap for Lofar 2.0

How to cope with a 4x data increase?
How to deliver science ready products?

Resources allocation & usage

Goals and Objectives: Identify the specific problems or to address, such as improving computational efficiency, enabling large-scale simulations, or accelerating data analysis.

Current State Assessment: Evaluate the organisation's or research field's existing compute and data infrastructure, applications, and workflows. Assess the strengths, weaknesses, and gaps in the current system.

Collaboration with Scientific Community

Requirements Gathering: Engage with stakeholders, including researchers, domain experts, and IT professionals, to understand their requirements for HPC/HTC applications. Identify the computational needs, data sizes, software dependencies, and performance expectations for different use cases.

Technology assessment: Evaluate and select the appropriate technologies, architectures, and tools based on the identified requirements.

Build a Roadmap for Applications

Infrastructure Planning: Define the infrastructure needed to support application usage and development. Determine the required hardware resources, such as high-performance computing clusters or cloud-based instances and design the network and storage systems accordingly.

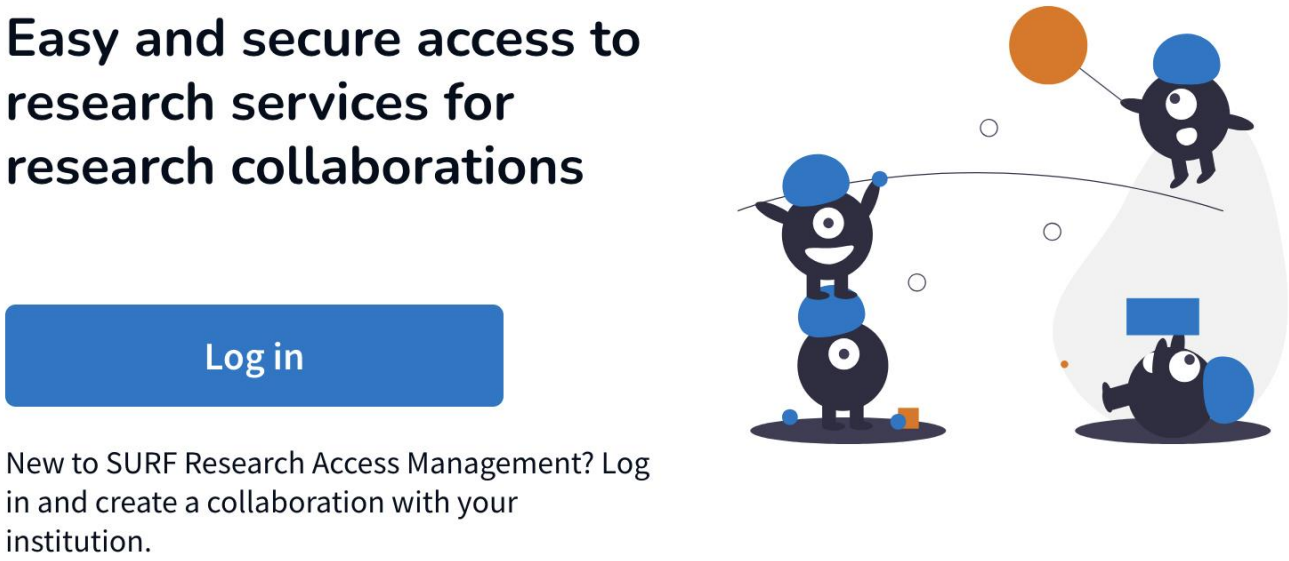
Collaboration and Knowledge Sharing among researchers, developers, and users of the applications. Encourage the reuse and dissemination of successful implementations.

SURF

Future plans

Spider Roadmap

- SRAM authentication
- Token based authentication
- GinA and Spider merge
- Slurm Upgrade
- CephFS improvements
- Workflow management tools
- SKA clone (NLSRC v0.1)
- Automated testing and monitoring
- Exploring kubernetes for distributed high-throughput workflows



dCache Roadmap

- Tape system replacement: DMF → Versity
- Ada (API client): add bulk-requests, labels
- Reorganize WebDAV doors
- Split dCache in two instances: evaluate benefits over horizontal scaling
- OIDC token support
- FTS & tokens support
- Metadata support (labels)
- FSSPEC support (dcachefs)
- STAC support (stac2dcache)

SURF

Innovation outcomes

- RUCIO server testing (done, but needs to community driven and deployment is K8S focused)
- Software (EESSI)
- non-posix filesystems
- energy aware computing (e.g. EAS/EAR)
- token support for storage and compute
- accelerators and non-conventional hardware
- AI & HPML
- Quantum computer
- FAIR data and implementation Open Science policies (e.g., PID, RAID, metadata updates)

Wrapping up

Data Processing Services:

- Spider: local compute cluster for High Throughput Computing
- Grid: federated compute infrastructure where SURF is tier-1 location for LTA / CERN
- MS4 (part of CRC): custom cloud setup for specific machines in High Throughput setups
- dCache: massive disk and tape storage for scientific data

Services in the Basic Service Package for members

Storage & Data Management
SURFfilesender

Identity & Access Management
SURFconext
eduID

Security
SURFaudit
SURFcert
Cyber crisis exercise OZON
Cybersave Yourself

Network Connectivity
Internet
EVPN - P2P
EVPN multipoint and L3VPN multipoint
eduroam
eduroam Visitor Access
iotroam

Procurement & Contracting
SURF IT Procurement
SURF Vendor Compliance
Distribution of hardware and software

Publishing
Auteursrechten.nl

Additional Services

SURF Consultancy

Procurement & Contracting
SURF Content Procurement
Credit via SURFspot

Compute
National Supercomputer Snellius
Data Processing
dCache
SURF Research Cloud
SURFcumulus and SURFcumulus Professional Services
Cloud Research Consultancy

Storage & Data Management
Research Drive
SURFdrive
Object Store
Data Archive
Yoda Hosting
RDM Storage Scale-out

Publishing
CopyrightCheck
SURFsharekit
edusources
Publinova
HBO Kennisbank
SURF Data Repository
Persistent Identifiers

Educational Logistics
Kies Op Maat
edubadges
SURFeduhub

Identity & Access Management
SURF Research Access Management
SURFsecureID

Security
SURFmail filter
eduVPN
SURFcertificates
SURFsoc

Network Connectivity
Secondary Site Connection
SURFwireless
SURFfirewall
NetherLight
SURFdomeinen