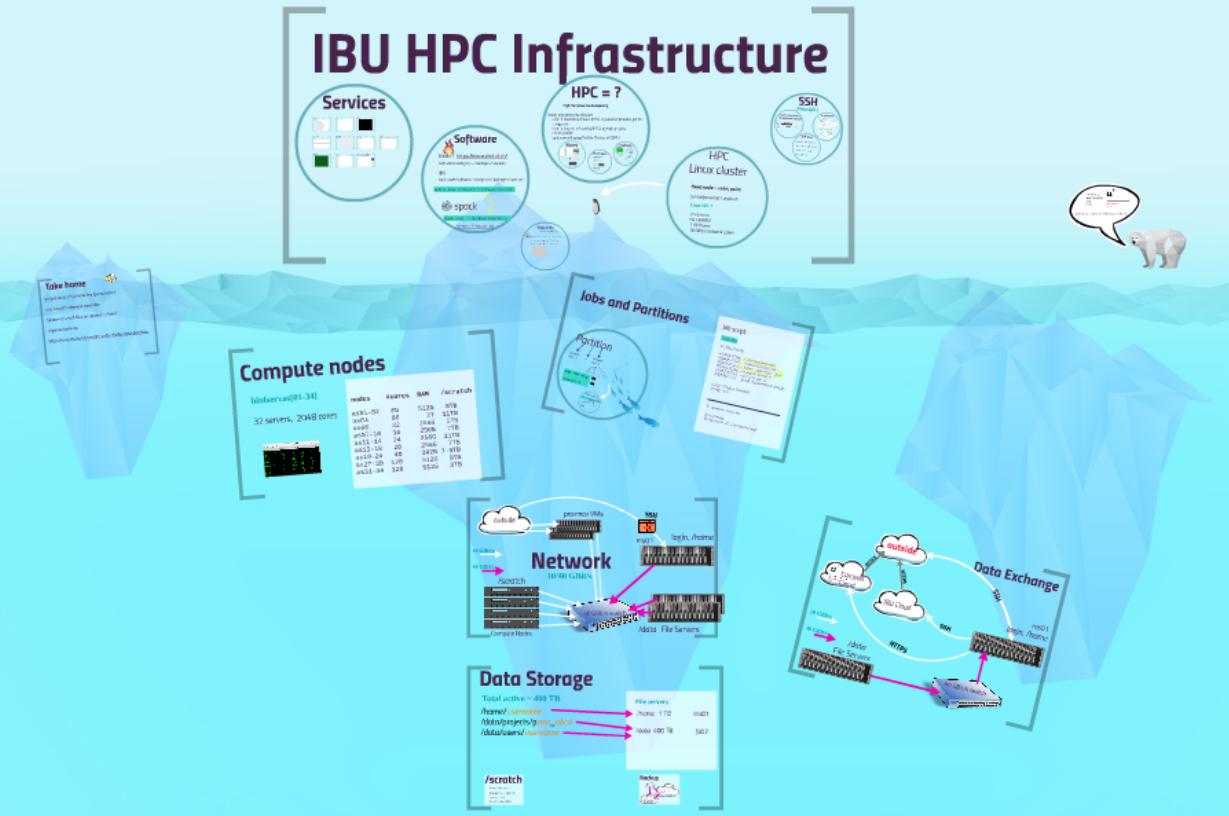


Interfaculty Bioinformatics Unit

Pierre Berthier, University of Bern, 26.10.2020



Interfaculty Bioinformatics Unit

Pierre Berthier, University of Bern, 26.10.2020

Interfaculty
Bioinformatics
Unit
(IBU)

u^b

b

**UNIVERSITÄT
BERN**

pierre.berthier@bioinformatics.unibe.ch



IBU HPC Infrastructure

Services



Software

Vital-IT <https://www.vital-it.ch/>
/software/<category>/<package>/<version>

IBU
`/opt/cluster/software/<category>/<package>/<version>`

```
module load <category>/<package>/<version>
spack
spack load -r <package>@<version>
https://spack.io
```

HPC = ?

High Performance Computing

- Using large amounts of power
- over a short time (hours) (HPC), eg weather forecast, genetic diagnostic
 - over a long time (months) (HTC), eg Astrophysics
 - "many tasks"
 - grid computing (eg Particle Physics at CERN)



HPC Linux cluster

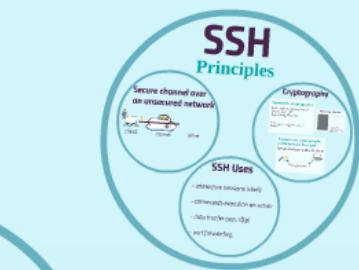
Head node = entry point

`ssh binfservms01.unibe.ch`

Cent OS 7

2 x 6 cores
64 GB RAM
1 TB /home
10 GBit/s Network uplink

SSH Principles



Jobs and Partitions

Partition

Job script

```
test.sh
#!/bin/bash
#SBATCH
```

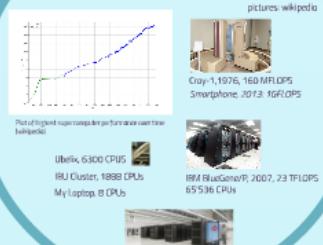
HPC = ?

High Performance Computing

Using large amounts of power

- over a short time (hours) (HPC), eg weather forecast, genetic diagnostic
- over a long time (months) (HTC), eg Astrophysics
- "many tasks"
- grid computing (eg Particle Physics at CERN)

History



Challenges

Electrical Power

Piz Daint: 3 MW
IBU: 15 kW
My Laptop: 50 W
City of Bern: 114 MW



Data flow

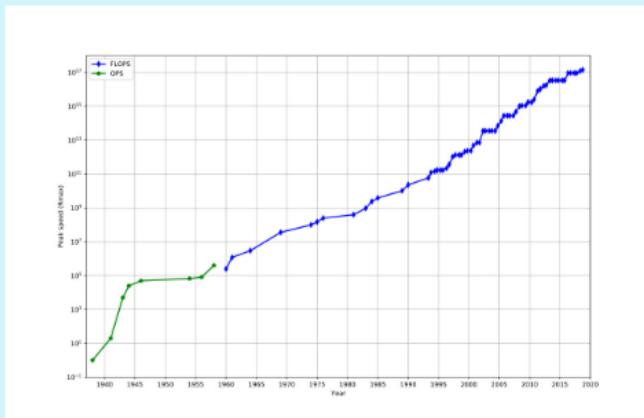
IBU Cluster: 1 PB Data
Network: 10 GBit/s

Features



History

pictures: wikipedia



Plot of highest supercomputer performance over time
(wikipedia)



Cray-1, 1976, 160 MFLOPS
Smartphone, 2013: 1GFLOPS



IBM BlueGene/P, 2007, 23 TFLOPS
65'536 CPUs

Ubelix, 6300 CPUS

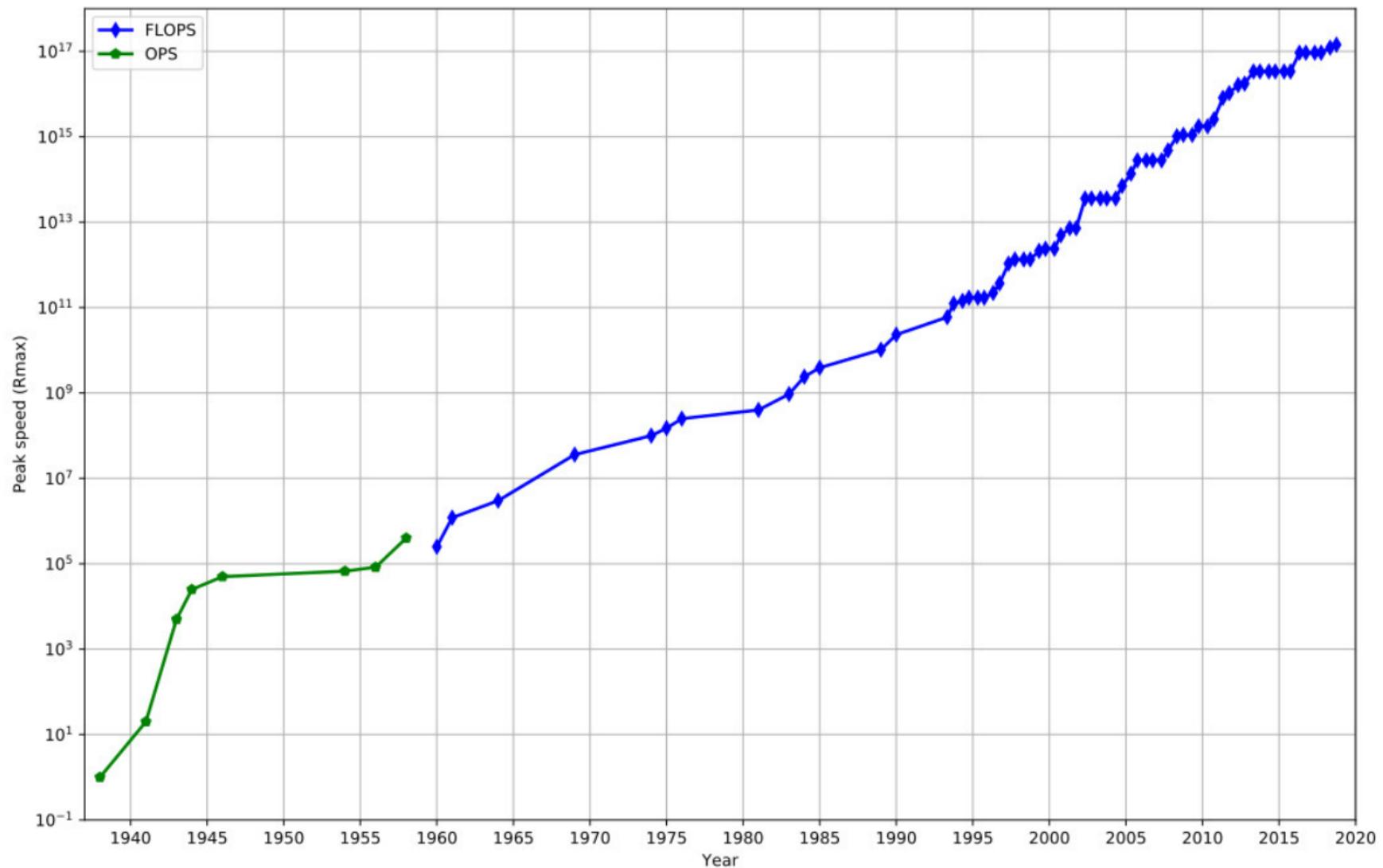


IBU Cluster, 1888 CPUs

My Laptop, 8 CPUs



Cray XC50, 2017, 27 PFLOPS
133'716 CPUs (Piz Daint, CSCS)



Plot of highest supercomputer performance over time
(wikipedia)



Cray-1, 1976, 160 MFLOPS
Smartphone, 2013: 1GFLOPS



IBM BlueGene/P, 2007, 23 TFLOPS
65'536 CPUs



Cray XC50, 2017, 27 PFLOPS
133'716 CPUs (Piz Daint, CSCS)

Ubelix, 6300 CPUS

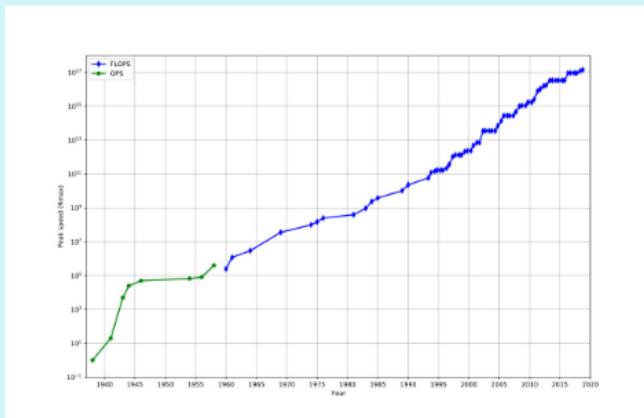


IBU Cluster, 1888 CPUs

My Laptop, 8 CPUs

History

pictures: wikipedia



Plot of highest supercomputer performance over time
(wikipedia)



Cray-1, 1976, 160 MFLOPS
Smartphone, 2013: 1GFLOPS



IBM BlueGene/P, 2007, 23 TFLOPS
65'536 CPUs

Ubelix, 6300 CPUS



IBU Cluster, 1888 CPUs

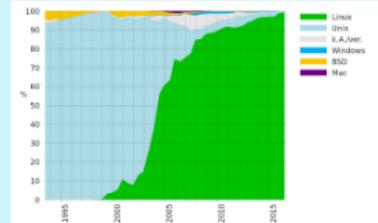
My Laptop, 8 CPUs



Cray XC50, 2017, 27 PFLOPS
133'716 CPUs (Piz Daint, CSCS)

Features

Operating System

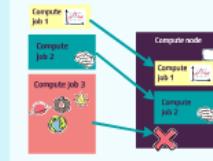


Operating systems used on top 500 supercomputers (wikipedia)

Queuing System

Concurrency on resources (CPUs, RAM) for users and jobs

Optimal usage of resources



Storage

Large capacities

1 Hard Disk: 16 TB

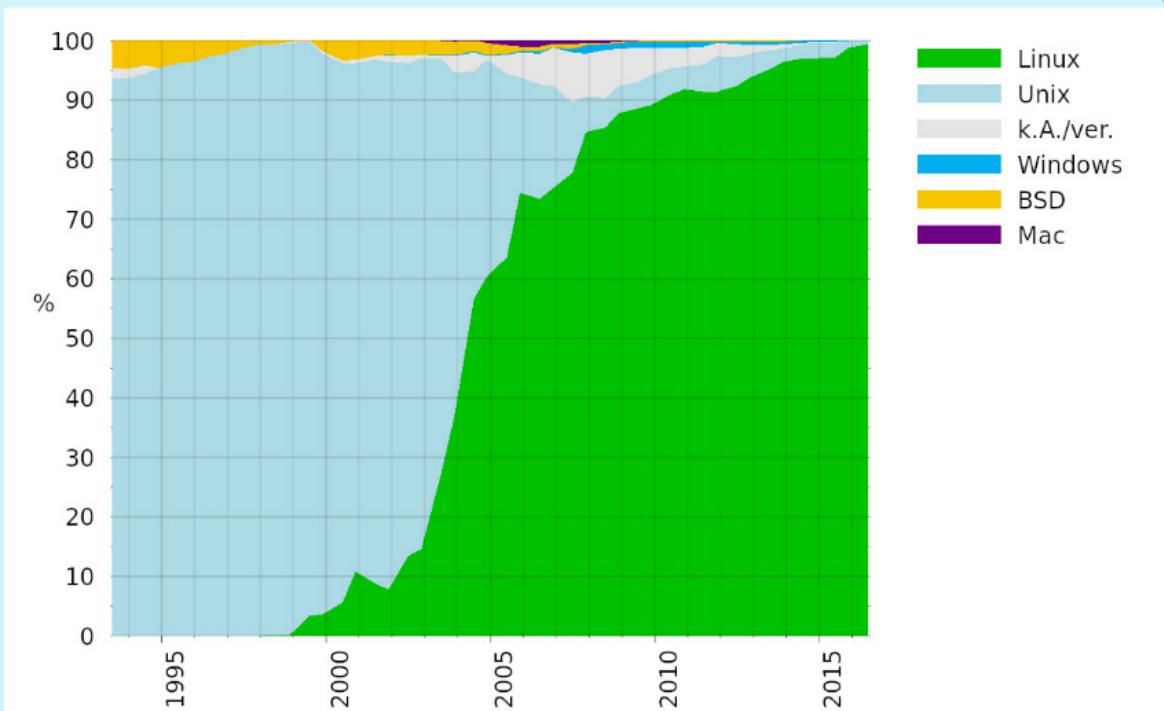
Piz Daint: 8'000 TB
Ubelix: 3'000 TB
IBU: 1'000 PB

Network

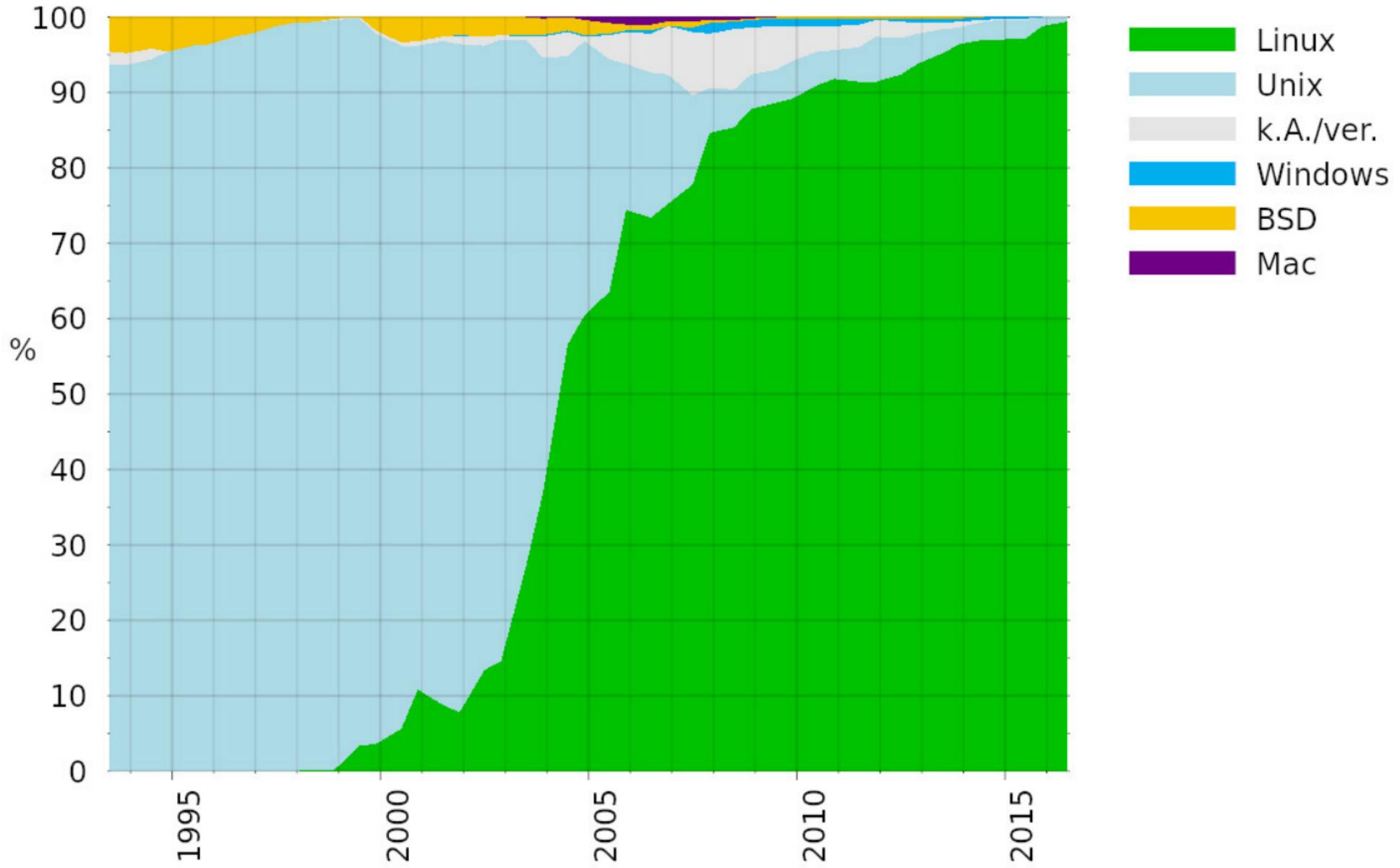
Nodes Interconnect

Outbound connection

Operating System



Operating systems used on top 500 supercomputers (wikipedia)



Operating systems used on top 500
supercomputers (wikipedia)

Storage

Large capacities

1 Hard Disk: 16 TB

Piz Daint: 8'000 TB

Ubelix: 3'000 TB

IBU: 1'000 PB

Network

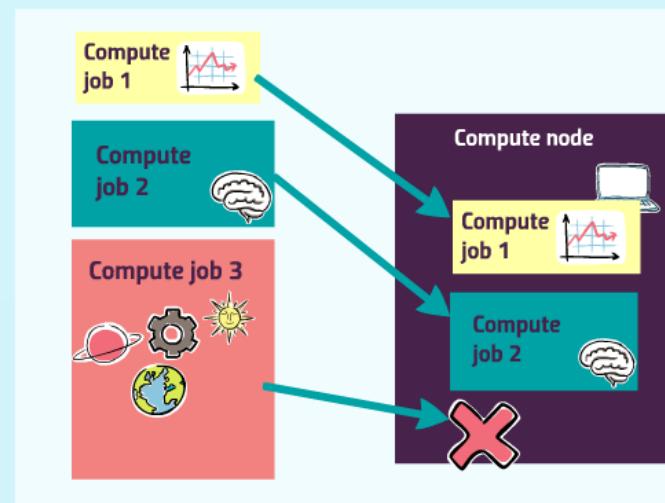
Nodes Interconnect

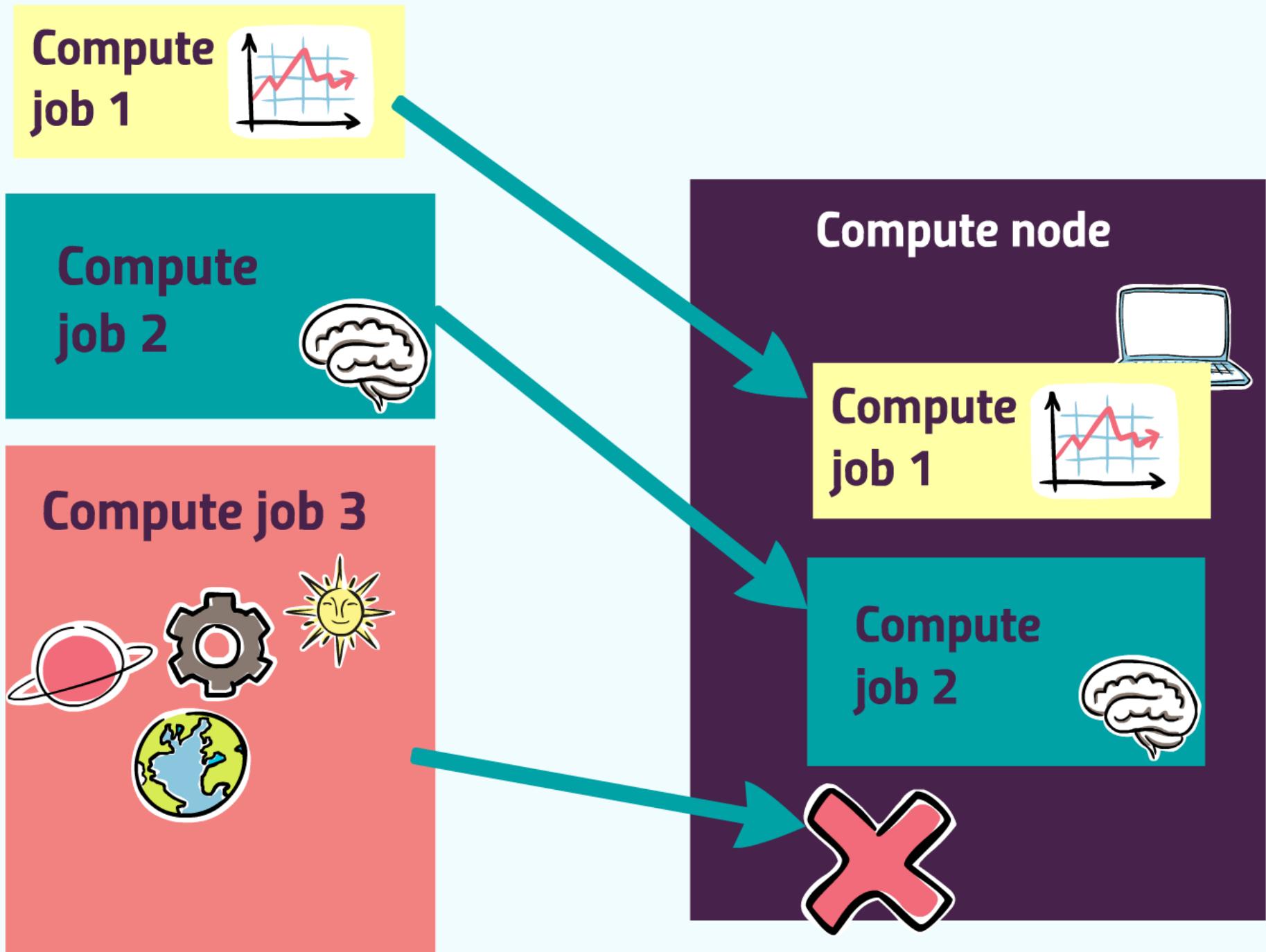
Outbound connection

Queuing System

Concurrency on resources (CPUs, RAM) for users and jobs

Optimal usage of resources





Challenges

Electrical Power

Piz Daint: 3 MW

IBU: 15 kW

My Laptop: 60 W

City of Bern: 114 MW

Cooling

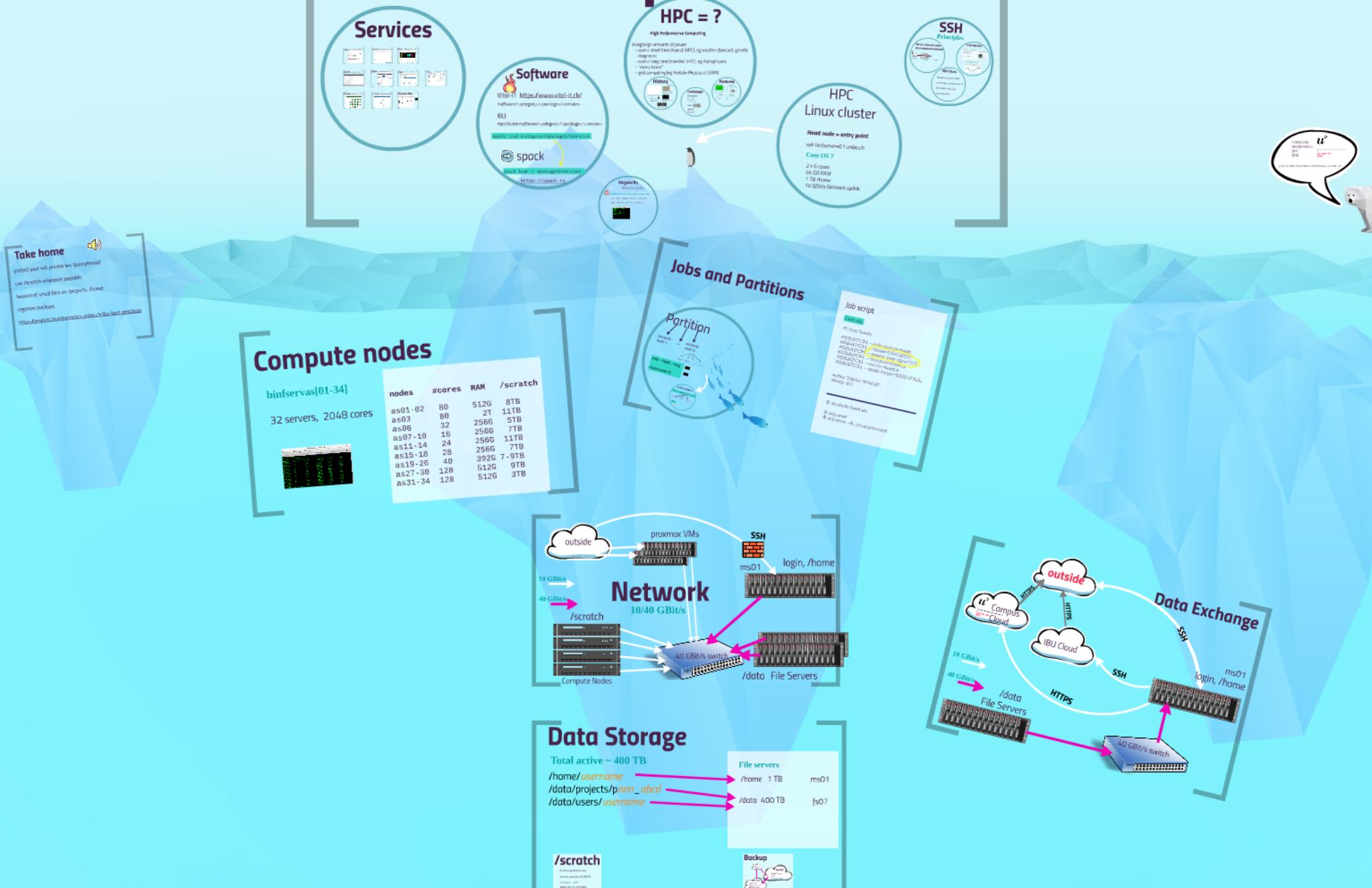


Data flow

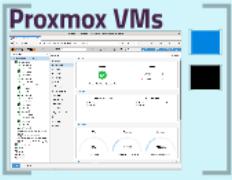
IBU Cluster: 1 PB Data

Network: 10 GBit/s

IBU HPC Infrastructure

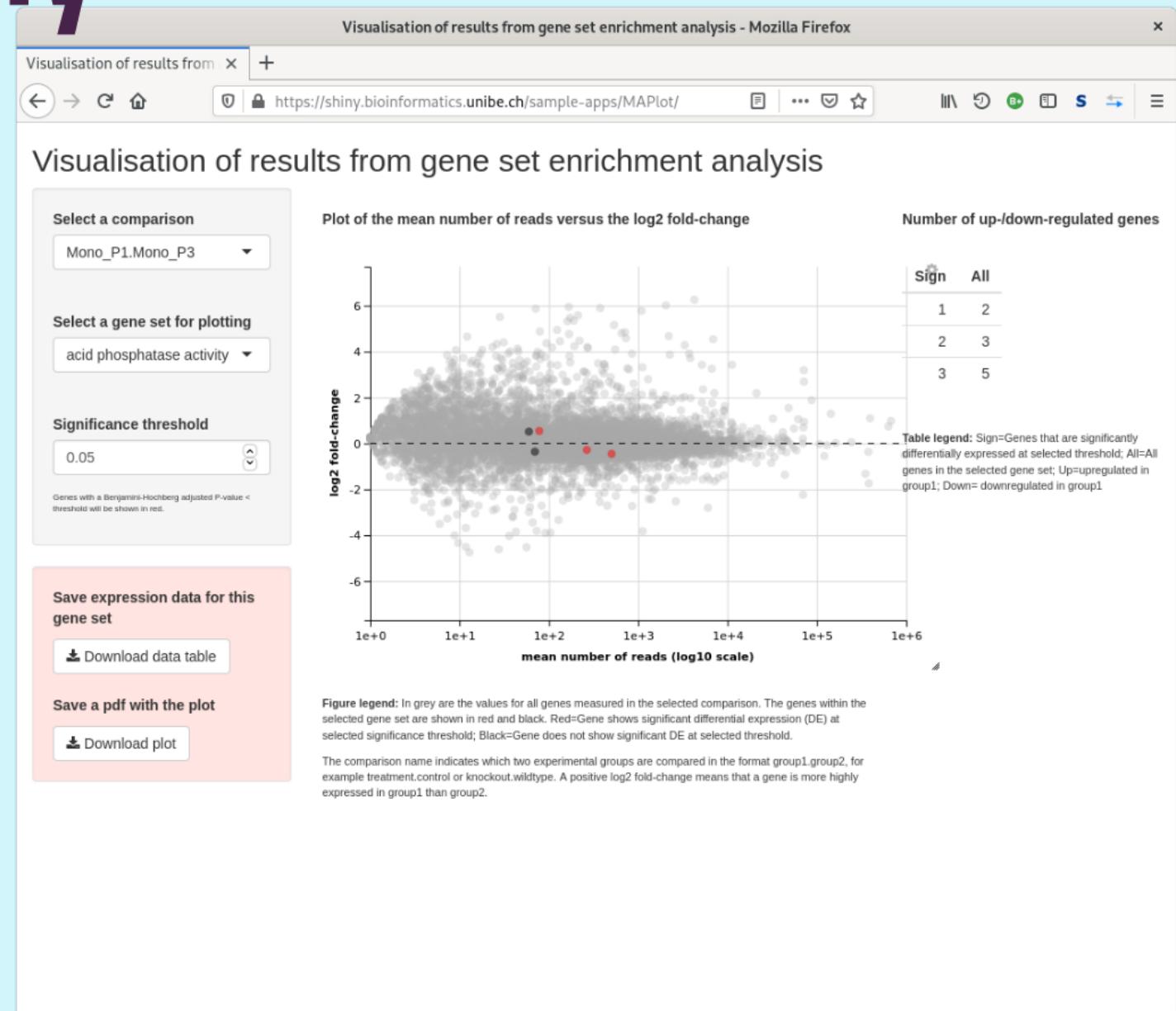


Services



Rshiny

<https://shiny.bioinformatics.unibe.ch/sample-apps/MAPlot/>



Sequenceserver

<https://xentobox.bioinformatics.unibe.ch/sequenceserver/>

The screenshot shows a Mozilla Firefox browser window with the title "SequenceServer: Custom BLAST Server - Mozilla Firefox". The address bar displays the URL <https://xentobox.bioinformatics.unibe.ch/sequenceserver/>. The main content area is titled "SequenceServer 1.0.11". A large text input field is present with the placeholder text "Paste query sequence(s) or drag file containing query sequence(s) in FASTA format here ...". Below this, a section titled "Nucleotide databases" lists several database options, each preceded by a checkbox:

- idm_dna_assembly.v1
- idm_rna_assembly_BE.v1
- idm_rna_assembly_Cannon.v1
- ssf_rna_assembly_BE.v1
- xnt_dna_assembly.v1
- xnt_dna_assembly.v2
- xnt_rna_assembly_all.v1

At the bottom, there is an "Advanced Parameters:" input field containing the text "eg: -eval 1.0e-5 -num_alignments 100", a help icon (?), and a blue "BLAST" button.

PACMAN: PACbio Methylation ANalyzer - Mozilla Firefox

PACMAN: PACbio Methylation ANalyzer https://bugfri.unibe.ch/

UNI FR UNIVERSITÉ DE FRIBOURG UNIVERSITÄT FREIBURG PACMAN

HOME MSC BIOINFORMATICS SEMINAR RESEARCH ABOUT US

openBIS

<https://bbbhub.unibe.ch/>

The screenshot shows a web browser window for the openBIS platform at <https://bbbhub.unibe.ch/openbis/webapp/eln-lims/?menuUniqueld=20181211162059617-92&v...>. The main content area is titled "Bbb: TEST1".

Left sidebar:

- Datasets
 - Others
 - Bruggmann
 - David Test
 - Default
 - Public
 - Yvesn
 - Project Test1
 - 20181211162058314-30
 - 20181211162058314-31

Top right actions:

- + New
- Edit
- Upload
- More ...

Section: Process all comparisons

Section: Pick groups to compare

Group by	-- select an option --
Group 1	-- select an option --
Group 2	-- select an option --

Section: Start processing

Section: Metadata

Group:
IBU_University of Bern

Analysis:
RNA-Seq

Species:
Homo Sapiens

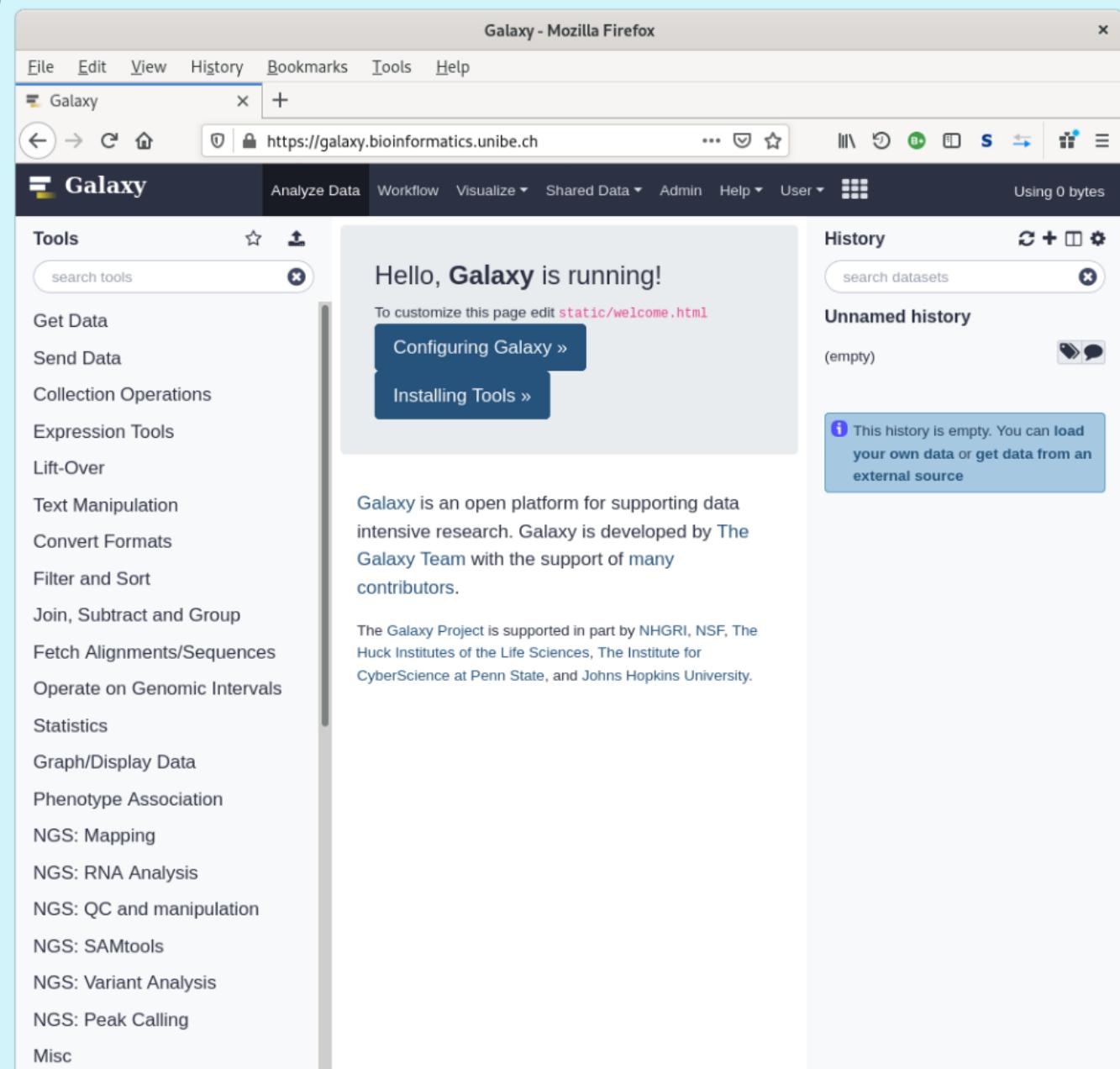
Platform:
Illumina HiSeq 2000

Bottom status bar:

96% 15:59 ENG

Galaxy

<https://galaxy.bioinformatics.unibe.ch/>



The screenshot shows the Galaxy web interface running on <https://galaxy.bioinformatics.unibe.ch/>. The browser window title is "Galaxy - Mozilla Firefox". The main content area displays a "Hello, Galaxy is running!" message with links to "Configuring Galaxy" and "Installing Tools". Below this, a paragraph describes Galaxy as an open platform for supporting data intensive research, developed by The Galaxy Team with contributions from many contributors. It also mentions the project's supporters: NHGRI, NSF, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State, and Johns Hopkins University. On the left, a sidebar titled "Tools" lists various tool categories: Get Data, Send Data, Collection Operations, Expression Tools, Lift-Over, Text Manipulation, Convert Formats, Filter and Sort, Join, Subtract and Group, Fetch Alignments/Sequences, Operate on Genomic Intervals, Statistics, Graph/Display Data, Phenotype Association, NGS: Mapping, NGS: RNA Analysis, NGS: QC and manipulation, NGS: SAMtools, NGS: Variant Analysis, NGS: Peak Calling, and Misc. On the right, there is a "History" panel showing an "Unnamed history" which is currently empty.

Galaxy - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Galaxy https://galaxy.bioinformatics.unibe.ch/

Galaxy Analyze Data Workflow Visualize Shared Data Admin Help User Using 0 bytes

Tools search tools

- Get Data
- Send Data
- Collection Operations
- Expression Tools
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- Fetch Alignments/Sequences
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- NGS: Mapping
- NGS: RNA Analysis
- NGS: QC and manipulation
- NGS: SAMtools
- NGS: Variant Analysis
- NGS: Peak Calling
- Misc

Hello, Galaxy is running!

To customize this page edit [static/welcome.html](#)

[Configuring Galaxy »](#)

[Installing Tools »](#)

Galaxy is an open platform for supporting data intensive research. Galaxy is developed by The Galaxy Team with the support of many contributors.

The Galaxy Project is supported in part by NHGRI, NSF, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State, and Johns Hopkins University.

History search datasets

Unnamed history (empty)

This history is empty. You can load your own data or get data from an external source

Gitlab

<https://gitlab.bioinformatics.unibe.ch/>

The screenshot shows a Mozilla Firefox browser window displaying a GitLab project page. The URL in the address bar is <https://gitlab.bioinformatics.unibe.ch/berthier/spack-ibu-repos>. The page title is "Pierre Berthier / spack IBU repos · GitLab - Mozilla Firefox".

The left sidebar contains a navigation menu with the following items:

- S spack IBU repos
- Project overview
- Details
- Activity
- Releases
- Repository
- Issues 0
- Merge Requests 0
- CI / CD
- Operations
- Analytics
- Wiki
- Snippets
- Settings

The main content area shows the project details for "spack IBU repos" (Project ID: 11). It displays statistics: 12 Commits, 1 Branch, 0 Tags, and 625 KB Files. A description states "Spack Repository for IBU specific packages".

Below the stats, there is a commit history section with a single entry:

new package py-mg-toolkit and dependencies
Pierre Berthier authored 2 months ago

Commit details: 5fe3f9dc

Below the commit history, there are several buttons for repository management:

- README
- Add LICENSE
- Add CHANGELOG
- Add CONTRIBUTING
- Enable Auto DevOps
- Add Kubernetes cluster
- Set up CI/CD

A table lists the repositories in the project:

Name	Last commit	Last update
amos	corrections after flake8 tests	1 year ago
fastp	clean up comments and dependencies	2 months ago
genomethreader	new package genomethreader	2 months ago
libdatrie	clean up comments and dependencies	2 months ago
libdrmaa	clean up comments and dependencies	2 months ago
py-aiodns	new package py-mg-toolkit and dependencies	2 months ago
	new package py-mg-toolkit and dependencies	2 months ago

At the bottom of the page, the URL <https://gitlab.bioinformatics.unibe.ch/berthier/spack-ibu-repos> is repeated.

Rstudio

<https://rstudio.bioinformatics.unibe.ch/users/berthier/>

The screenshot shows the RStudio interface running in a Mozilla Firefox browser window. The title bar reads "RStudio - Mozilla Firefox".

Console pane:

```
R version 3.6.2 (2019-12-12) -- "Dark and Stormy Night"  
Copyright (C) 2019 The R Foundation for Statistical Computing  
Platform: x86_64-pc-linux-gnu (64-bit)  
  
R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.  
  
Natural language support but running in an English locale  
  
R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.  
  
Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.
```

Environment pane:

Global Environment

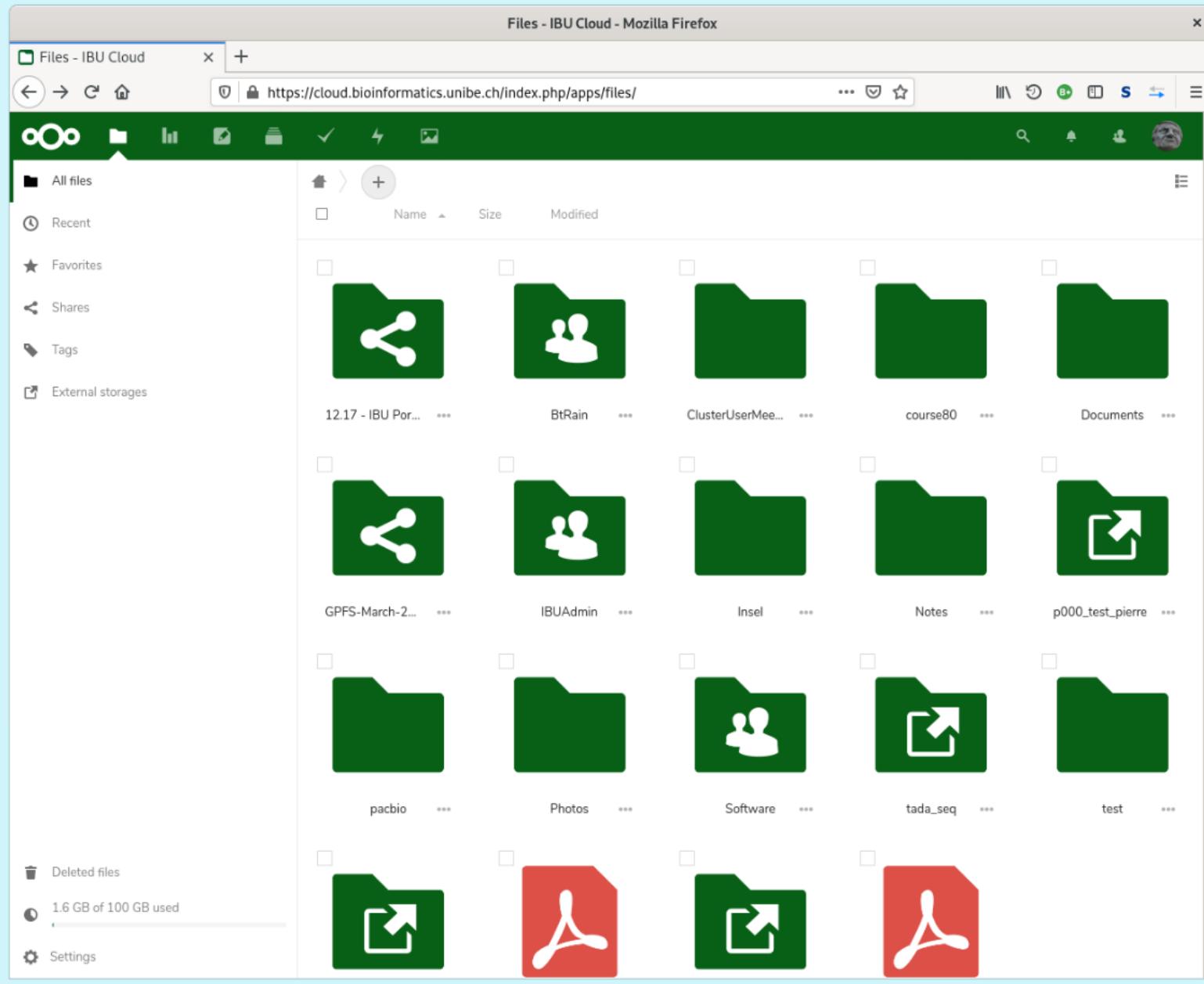
Environment is empty

Files pane:

Name	Size	Modified
.Rhistory	254 B	May 7, 2020,
CLUSTER		
p000		
p380		
R		

IBU Cloud

<https://cloud.bioinformatics.unibe.ch/>



openProjects

<https://projects.bioinformatics.unibe.ch/>

Proxmox VMs

binfsvpx01 - Proxmox Virtual Environment - Mozilla Firefox

binfsvpx01 - Proxmox

https://px01.binf.unibe.ch:8006/#v1:0:18:4::::5

You are logged in as 'root@pam' Documentation Create VM Create CT

PROXMOX Virtual Environment 5.4-13 Search

Server View

- Datacenter (IBU-MAIN)
 - binfsvpx01
 - 100 (app18)
 - 101 (des101)
 - 102 (app02)
 - 105 (app01)
 - 106 (app03)
 - 107 (des113)
 - 109 (binfsvapp13)
 - 111 (lic01)
 - 113 (app12)
 - 114 (app05)
 - 115 (binfldap01)
 - 117 (app17)
 - 119 (app09)
 - 120 (app10)
 - 121 (chef)
 - 123 (binfnfs02)
 - 125 (binfmonitor)
 - 129 (fs05)
 - 130 (app20)
 - 135 (app11)
 - binfsan01nfs (binfsvpx01)
 - local (binfsvpx01)
 - localdata-px01 (binfsvpx01)
 - localdata-px02 (binfsvpx01)
 - binfsvpx02
 - Apps
 - Test
 - allvms
 - critical

Tasks Cluster log

Datacenter

Health

Status

Nodes

Cluster: IBU-MAIN, Quorate: Yes

Online	2
Offline	0

Guests

Virtual Machines

Running	28
Stopped	3

LXC Container

Running	0
Stopped	0

Resources

CPU

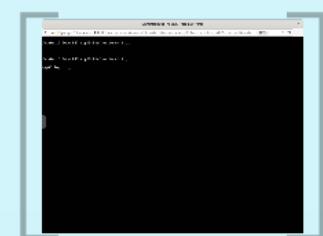
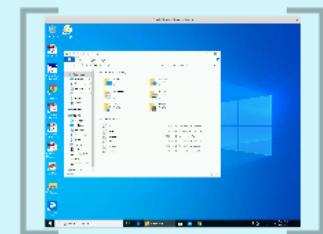
Memory

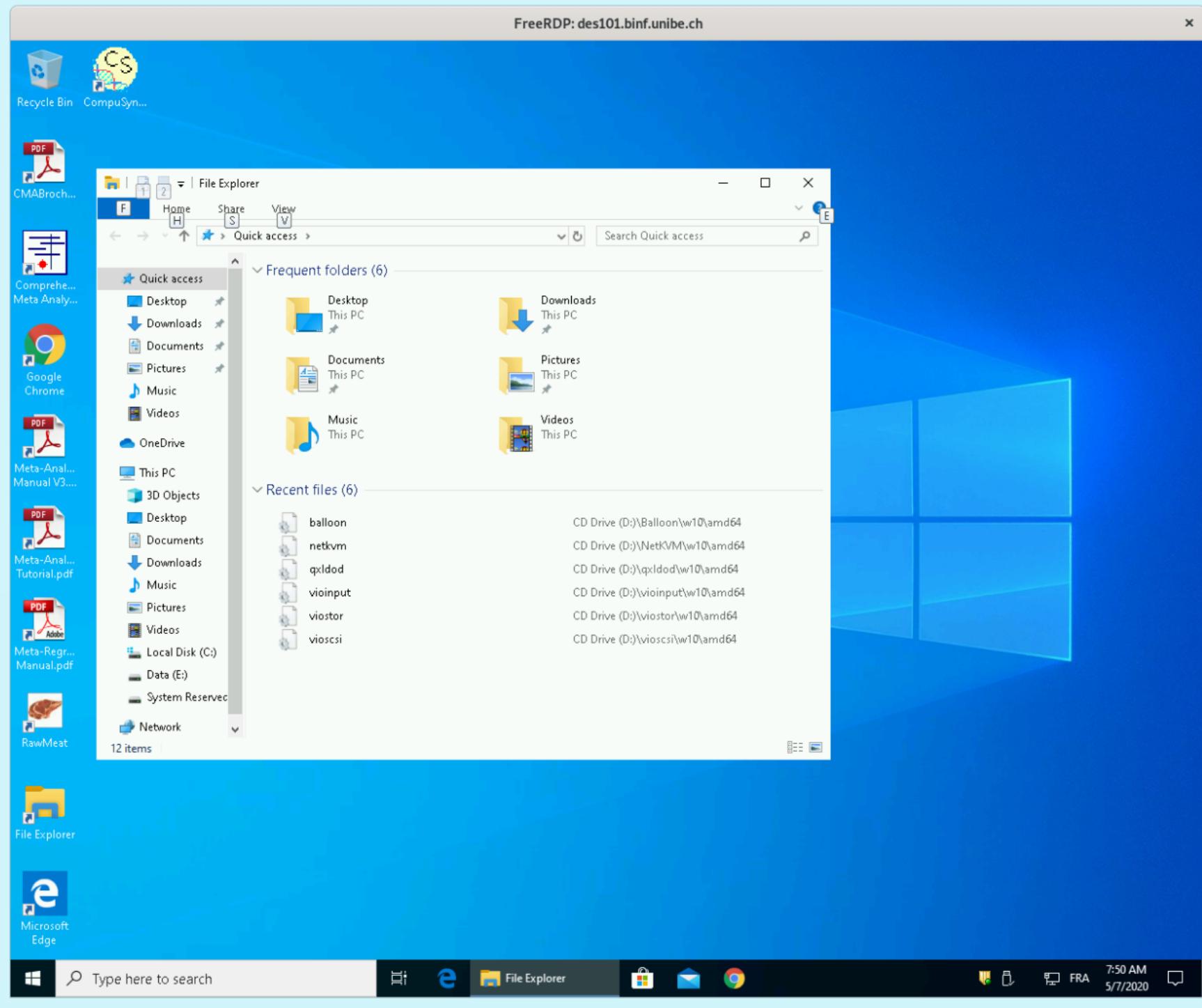
Storage

2% of 32 CPU(s)

30% 224.99 GiB of 755.64 GiB

39% 15.18 TiB of 39.18 TiB





QEMU (app08) - noVNC - Mozilla Firefox

x

https://px01.binf.unibe.ch:8006/?console=kvm&novnc=1&vmid=112&vmname=app08&node=binfservpx02&resize=off&cmd= 110% ... ☰ ☆

Ubuntu 18.04.4 LTS app08.binf.unibe.ch tty1

Ubuntu 18.04.4 LTS app08.binf.unibe.ch tty1

app08 login: _

HPC Linux cluster

Head node = entry point

ssh btrfservms01.unibe.ch

Cent OS 7

2 x 6 cores

64 GB RAM

1 TB /home

10 GBit/s Network uplink

SSH

Principles

**Secure channel over
an unsecured network**

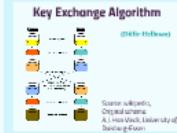


Cryptography

Symmetric cryptography

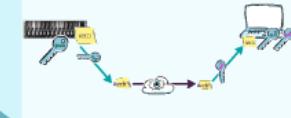
Goal: establish a secured channel
Shared secret

Key Exchange algorithm



**Asymmetric cryptography
public/private keys pair**

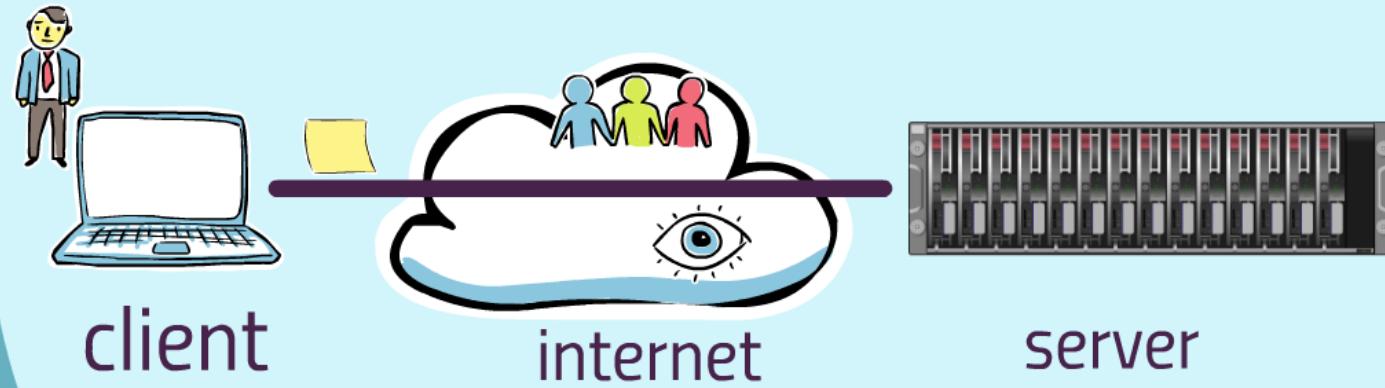
Server and User authentication



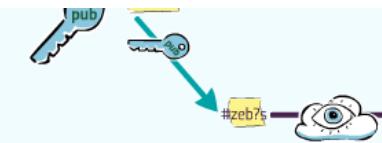
SSH Uses

- interactive sessions (shell)
- commands execution on server
- data transfer (scp, sftp)
- port forwarding

Secure channel over an unsecured network



SSH U



SSH Uses

- interactive sessions (shell)
- commands execution on server
- data transfer (scp, sftp)
- port forwarding

Cryptography

Symmetric cryptography

Goal: establish a secured channel

Shared secret 

Key Exchange algorithm



Key Exchange Algorithm

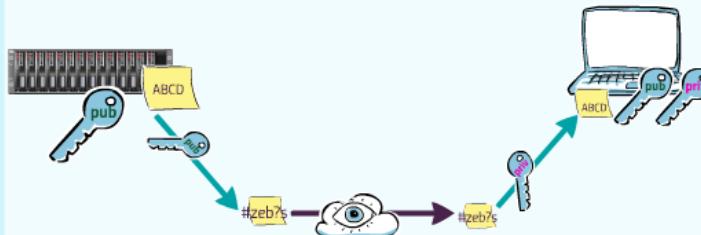
(Diffie-Hellman)



Source: wikipedia,
Original schema:
A.J. Han Vinck, University of
Duisburg-Essen

Asymmetric cryptography public/private keys pair

Server and User authentication



Uses

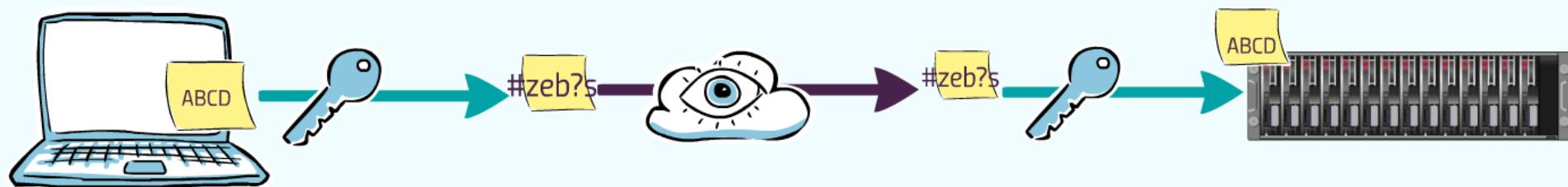
Symmetric cryptography

Goal: establish a secured channel

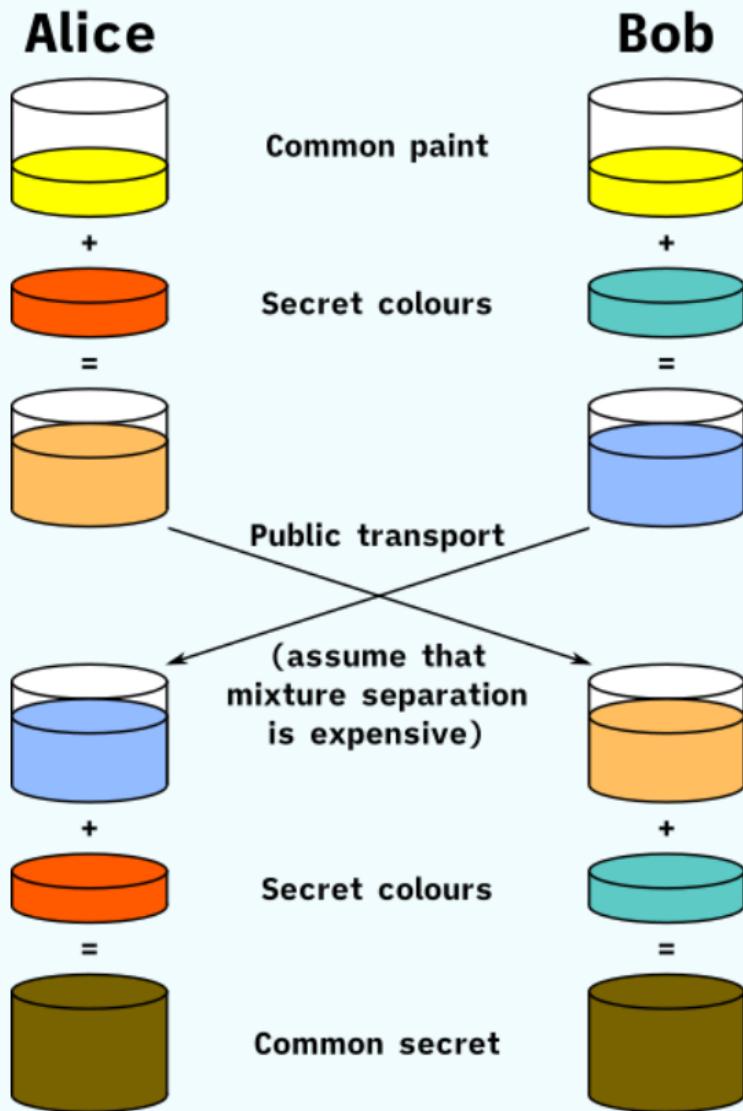
Shared secret



Key Exchange algorithm



Key Exchange Algorithm

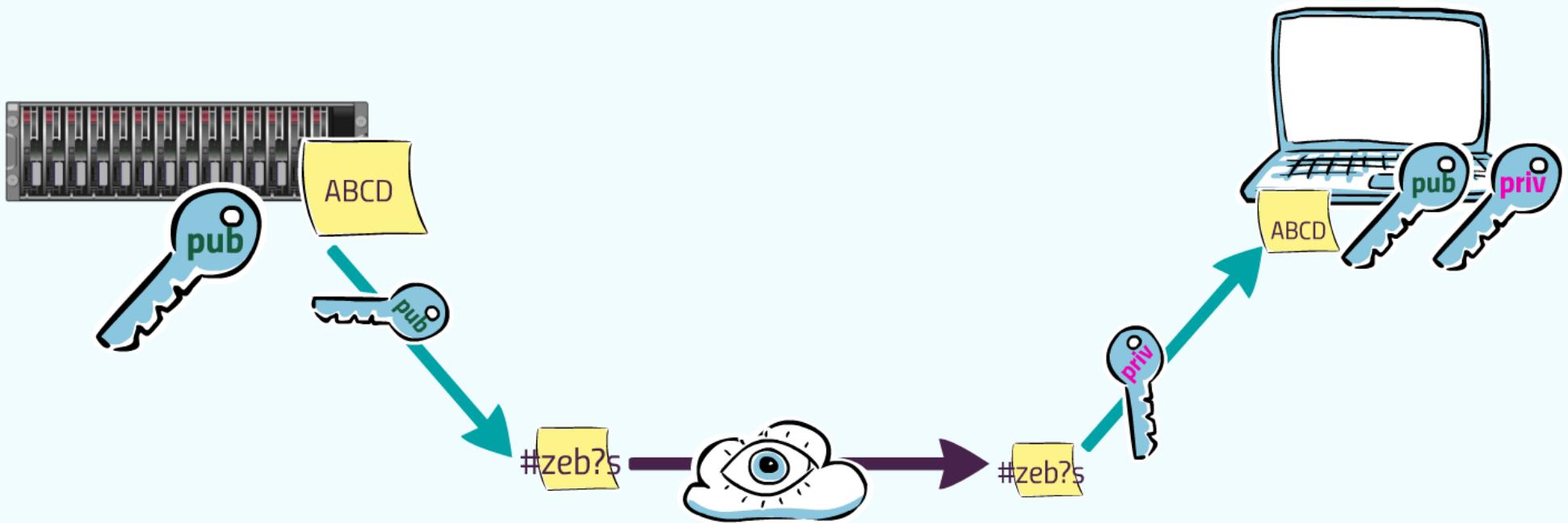


(Diffie-Hellman)

Source: wikipedia,
Original schema:
A.J. Han Vinck, University of
Duisburg-Essen

Asymmetric cryptography public/private keys pair

Server and User authentication



HPC Linux cluster

Head node = entry point

ssh btrfservms01.unibe.ch

Cent OS 7

2 x 6 cores

64 GB RAM

1 TB /home

10 GBit/s Network uplink

Data Storage

Total active ~ 400 TB

/home/*username*

/data/projects/p*nnn_abcd*

/data/users/*username*



File servers

/home 1 TB

ms01

/data 400 TB

fs07

/scratch

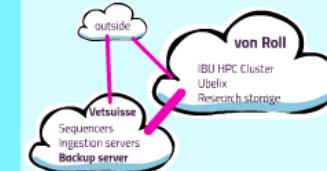
directory local to each node

during job execution: \$SCRATCH

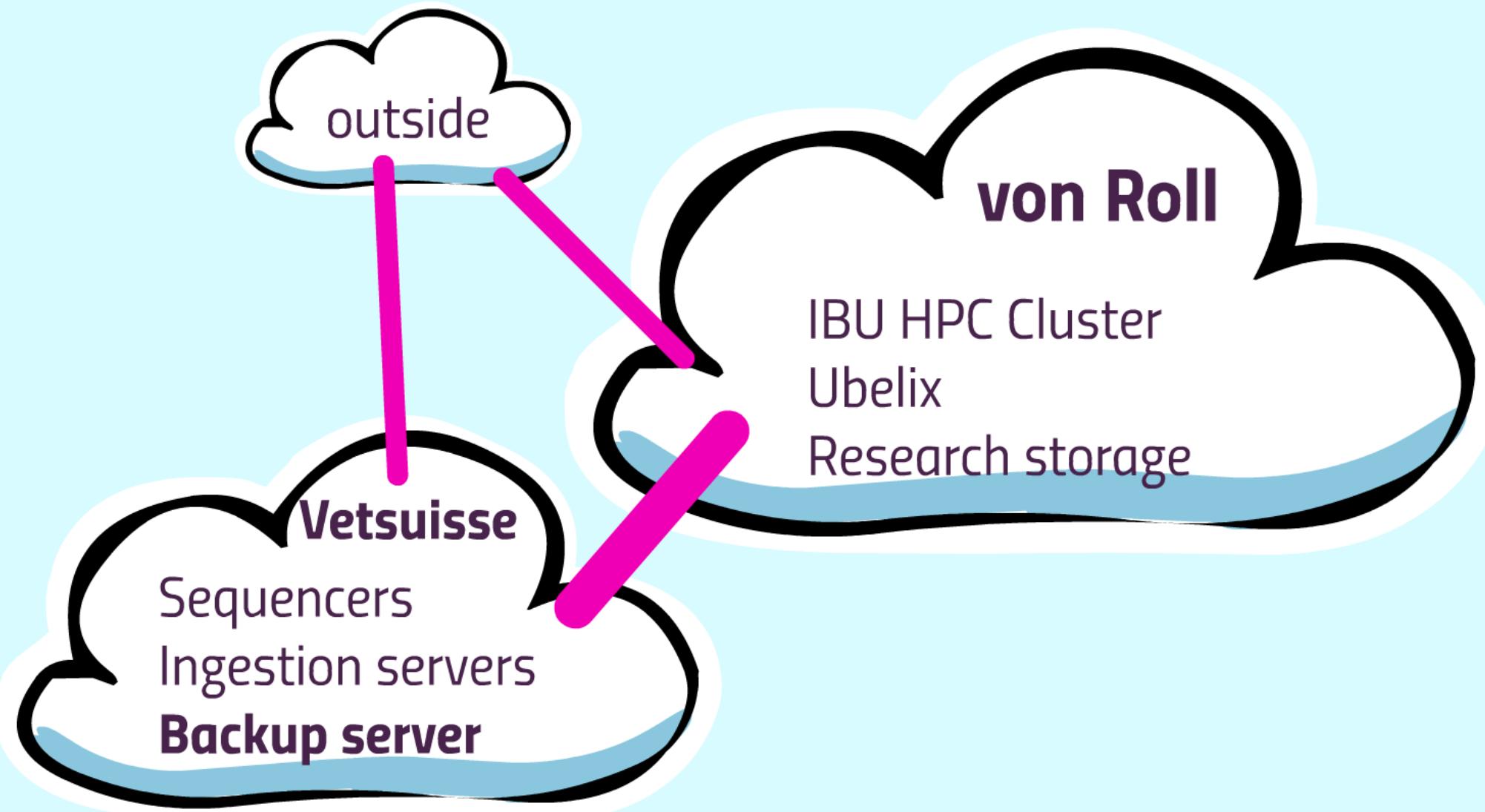
/scratch/172007

deleted after job completion

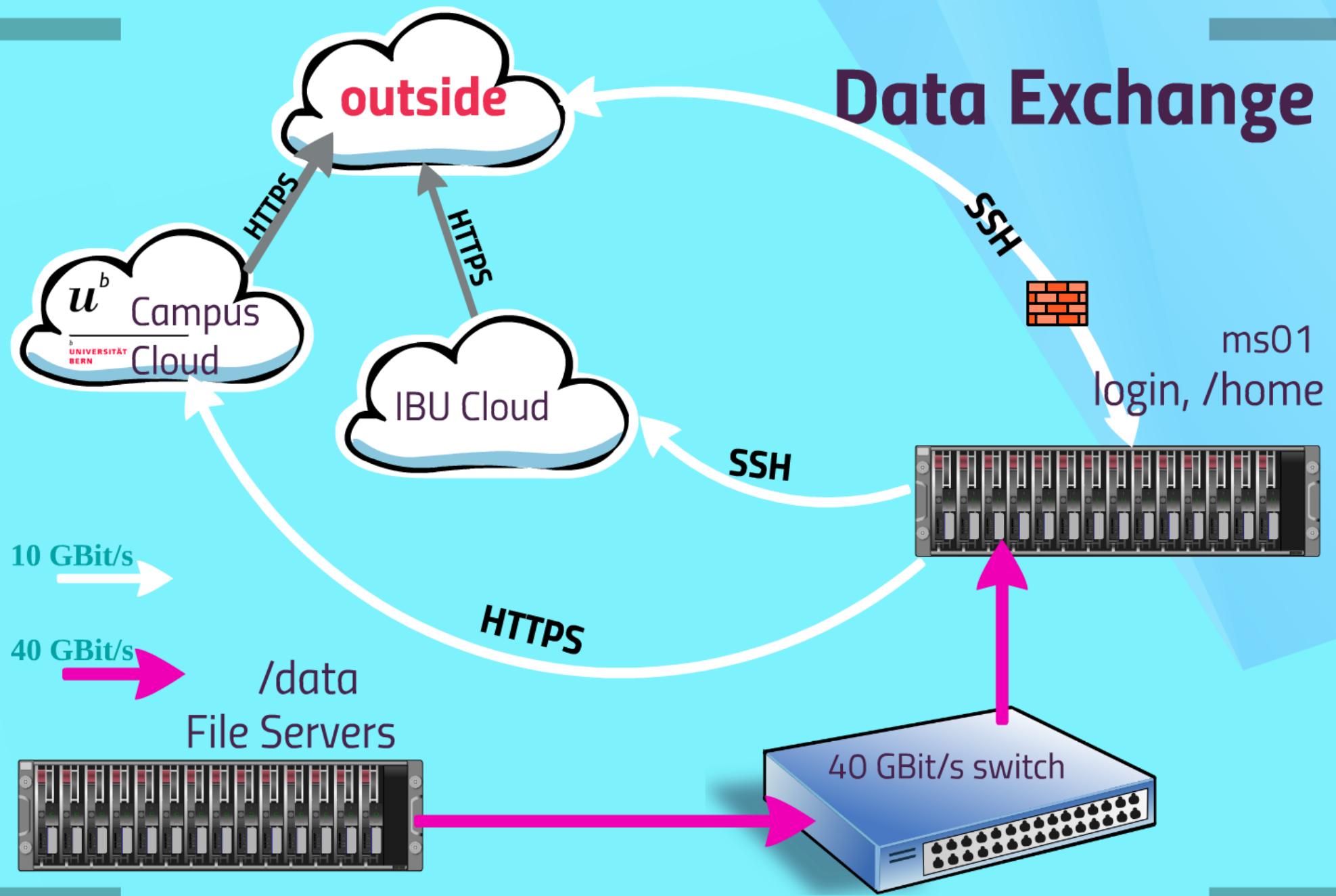
Backup

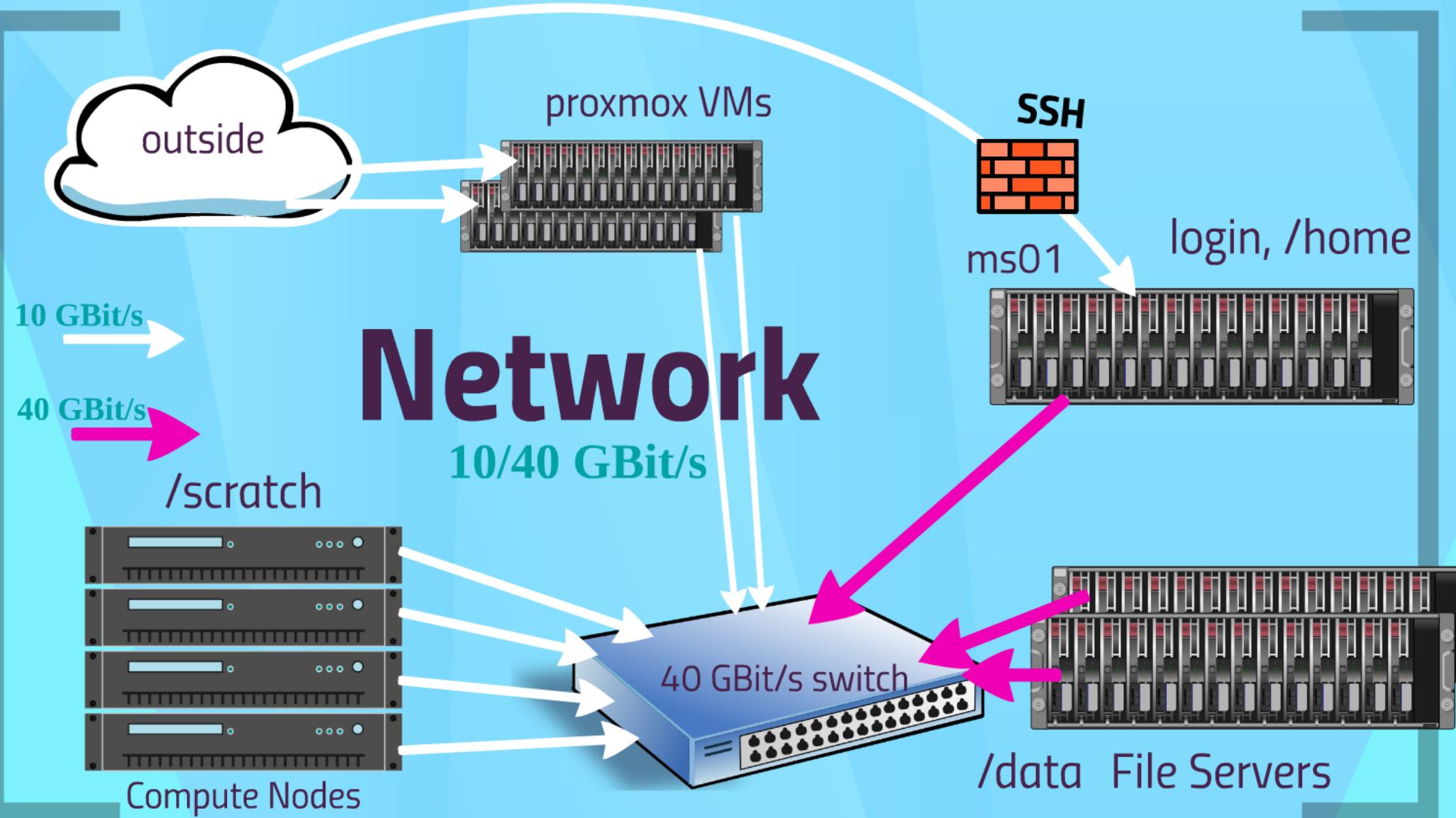


Backup



Data Exchange

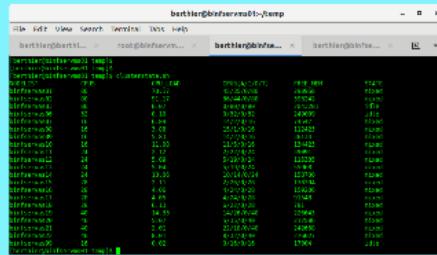




Compute nodes

binfo[01-34]

32 servers, 2048 cores



nodes	#cores	RAM	/scratch
as01-02	80	512G	8TB
as03	80	2T	11TB
as06	32	256G	5TB
as07-10	16	256G	7TB
as11-14	24	256G	11TB
as15-18	28	256G	7TB
as19-26	40	392G	7 - 9TB
as27-30	128	512G	9TB
as31-34	128	512G	3TB

berthier@binfervms01:~/temp

File Edit View Search Terminal Tabs Help

berthier@berthi... ×

root@binfervm... ×

berthier@binfse... ×

berthier@binfse... ×



[berthier@binfervms01 temp]\$

[berthier@binfervms01 temp]\$

[berthier@binfervms01 temp]\$ clusterstate.sh

NODELIST	CPUS	CPU_LOAD	CPUS(A/I/O/T)	FREE_MEM	STATE
binfervas01	80	70.17	45/35/0/80	298958	mixed
binfervas02	80	91.17	36/44/0/80	393242	mixed
binfervas03	80	0.07	0/80/0/80	2043281	idle
binfervas06	32	0.10	0/32/0/32	249009	idle
binfervas07	16	6.89	14/2/0/16	74547	mixed
binfervas08	16	3.88	15/1/0/16	112423	mixed
binfervas09	16	5.83	14/2/0/16	36134	mixed
binfervas10	16	11.08	11/5/0/16	134423	mixed
binfervas11	24	2.12	2/22/0/24	39091	mixed
binfervas12	24	5.09	5/19/0/24	115335	mixed
binfervas13	24	5.04	5/19/0/24	55968	mixed
binfervas14	24	10.06	10/14/0/24	158700	mixed
binfervas15	28	2.11	2/26/0/28	138394	mixed
binfervas16	28	4.06	4/24/0/28	159206	mixed
binfervas17	28	4.05	4/24/0/28	91548	mixed
binfervas18	28	6.13	6/22/0/28	781	mixed
binfervas19	40	14.33	14/26/0/40	226643	mixed
binfervas20	40	5.07	5/35/0/40	212586	mixed
binfervas21	40	2.01	22/18/0/40	242668	mixed
binfervas22	40	8.01	8/32/0/40	276027	mixed
binfervas99	16	0.02	0/16/0/16	17004	idle

[berthier@binfervms01 temp]\$

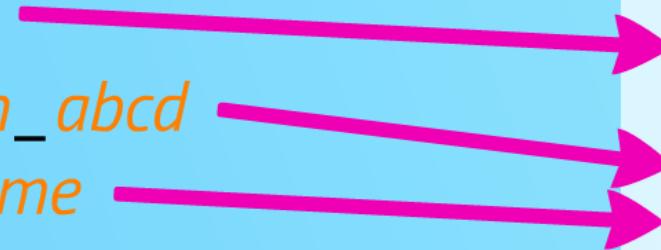
Data Storage

Total active ~ 400 TB

/home/*username*

/data/projects/p*nnn_abcd*

/data/users/*username*



File servers

/home 1 TB

ms01

/data 400 TB

fs07

/scratch

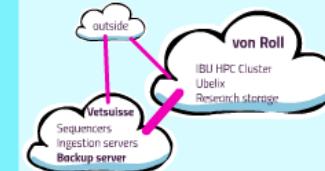
directory local to each node

during job execution: \$SCRATCH

/scratch/172007

deleted after job completion

Backup



/scratch

directory local to each node

during job execution: \$SCRATCH

/scratch/172007

deleted after job completion

Take home



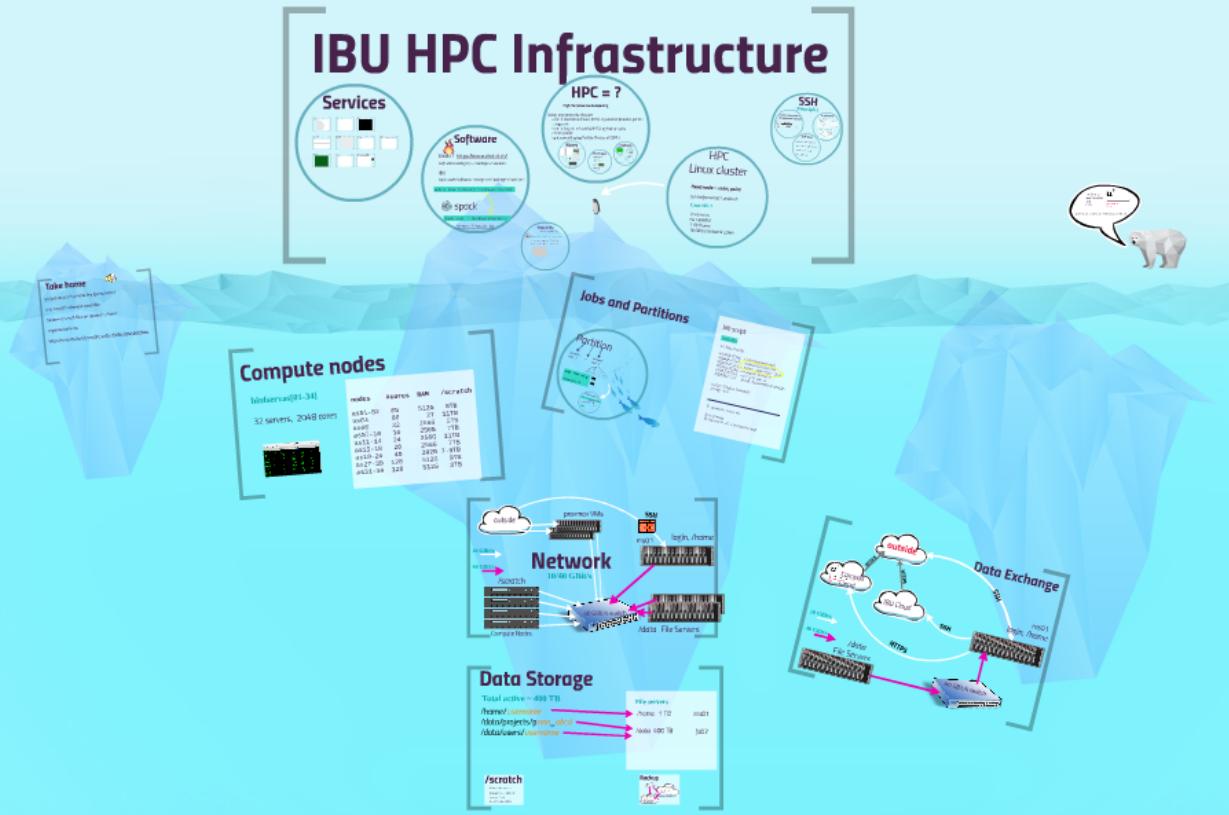
protect your ssh private key (passphrase)

use /scratch whenever possible

beware of small files on /projects, /home

organize backups

<https://projects.bioinformatics.unibe.ch/ibu-best-practices>



Interfaculty Bioinformatics Unit

Pierre Berthier, University of Bern, 26.10.2020