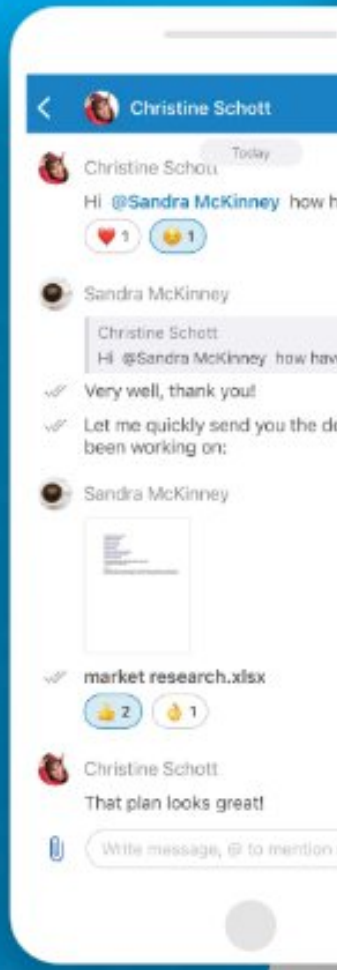
WEBINAR

Scaling Nextcloud for Universities

Insights from managing 35 instances for 50,000 users

Nextcloud



Nextcloud

Technische Universität Berlin

Nextcloud Hub 4

https://go.nextcloud.com/r/sH5

ZE CMPS MGMT

Webinars coming up

Collaborate and Contribute: A guide to joining Nextcloud's open source movement
Tuesday, 28 March 1PM CET

How to migrate to Nextcloud Hub
Thursday, 30 March 3PM CET

More webinars at nextcloud.com/events/

Nextcloud

Technische Universität Berlin

Webinar rules

- Please ask your questions in the chat
- We will answer questions **after each topic**
- Stating the name or the post's timestamp
- Webinar will be recorded and distributed

ZE CMPS MGMT

Nextcloud

Technische Universität Berlin

Thomas Hildmann

Head of the Department of Administrative, Infrastructure, and Application Services at Technische University of Berlin

ZE CMPS MGMT

Technische Universität Berlin

Scaling Nextcloud for Universities

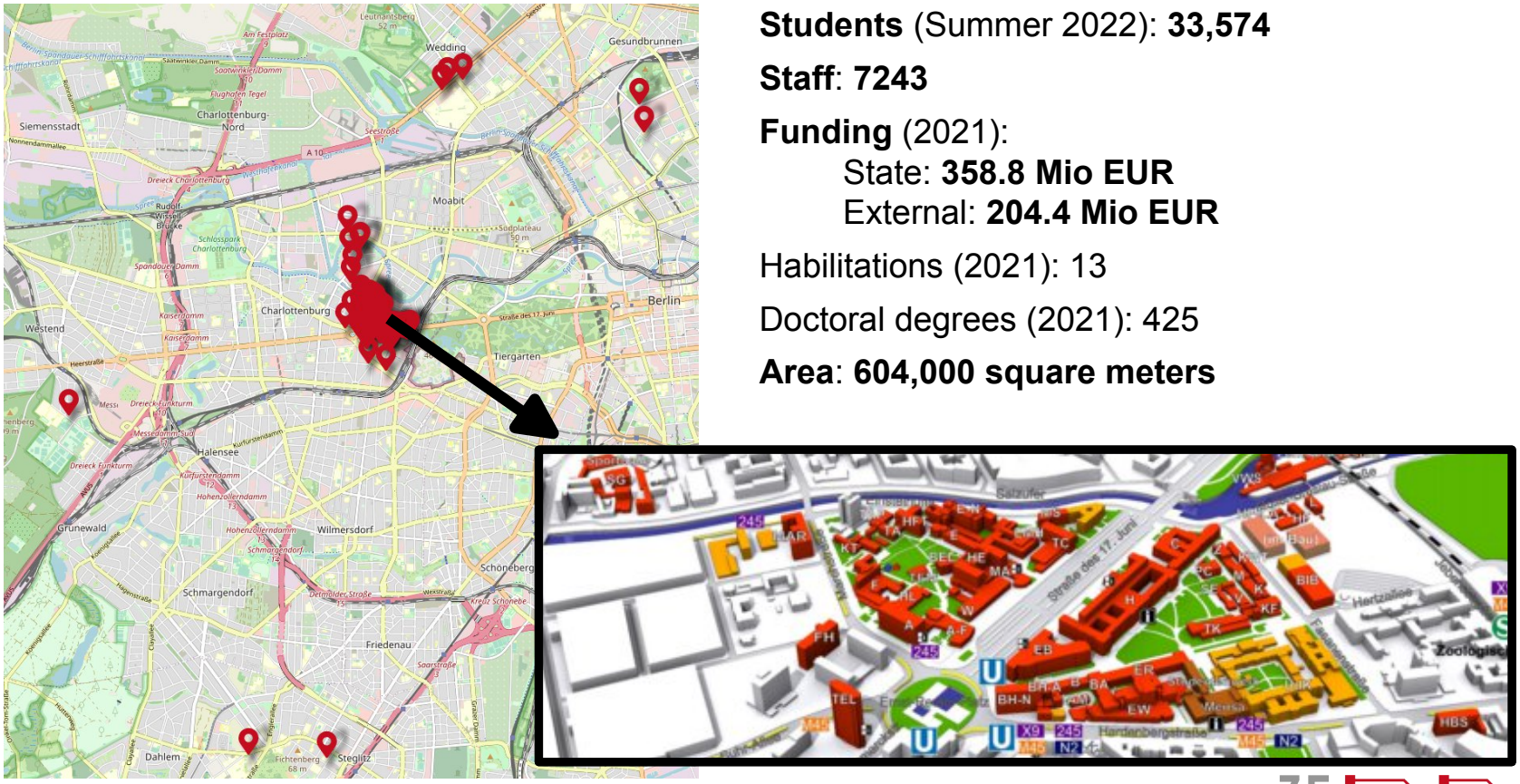
Insights from Managing 35 Instances for 50,000 Users

Dr. Thomas Hildmann | ZECM | Webinar March 22nd, 2023

Agenda

1. Introduction
2. Motivation
3. Technical Architecture
4. Federation
5. Experiences in numbers
6. Monitoring
7. Conclusion

Introduction: Technische Universität Berlin



Students (Summer 2022): **33,574**

Staff: **7243**

Funding (2021):

State: **358.8 Mio EUR**

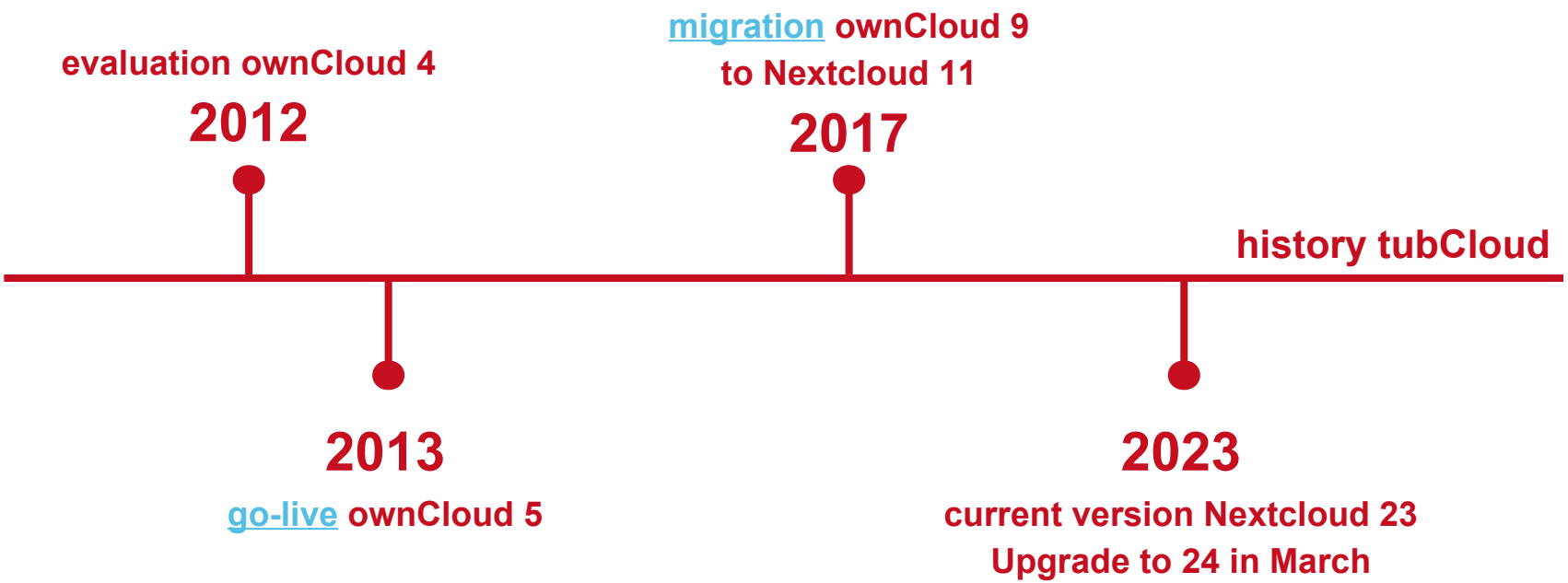
External: **204.4 Mio EUR**

Habilitations (2021): 13

Doctoral degrees (2021): 425

Area: **604,000 square meters**

Introduction: tubCloud (1)



Introduction: tubCloud (2)

Statistics tubCloud

32,549 users

4,926 active (24h)

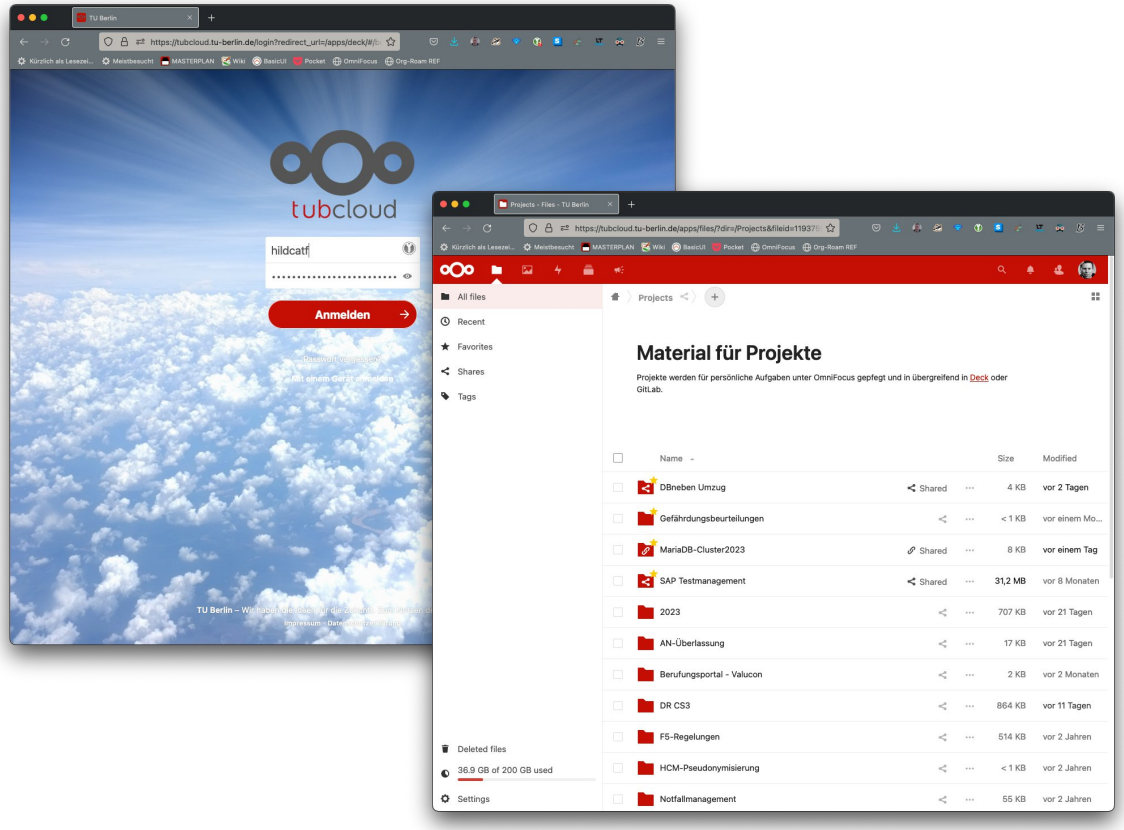
2,701 (1h)

2,076 (5 mins)

365,211,984 Files

using **279 TB**

average ~0,8 MB per file



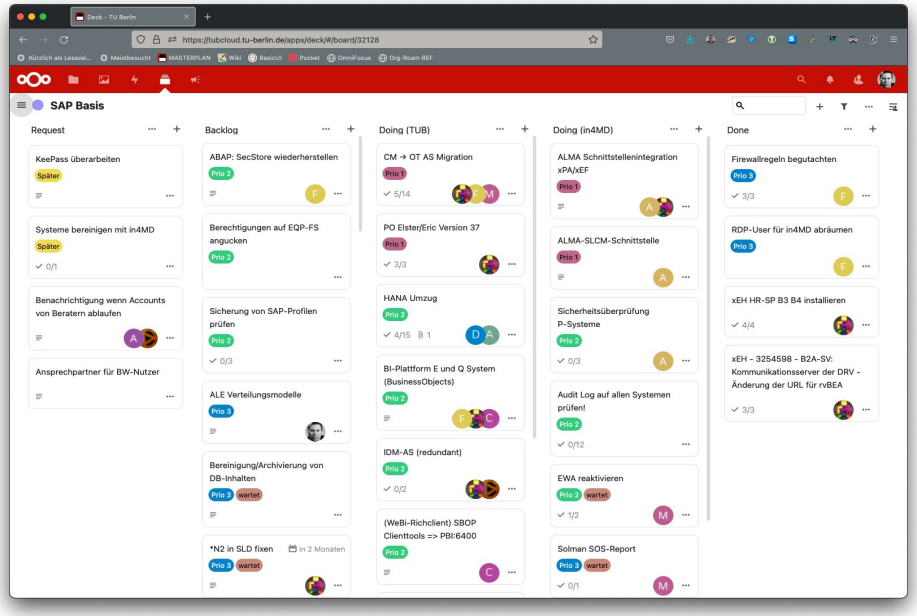
Quota tubCloud

students: 50 GB
staff: 200 GB
units: 500 GB

Introduction tubCloud (3)

Apps and Usecases tubCloud

- **Announcement Center**
  for maintenance windows, news, known issues
- **Deck**
  simple project planning, team meetings
- **ONLYOFFICE**
  collaborative office suite
- **Markdown Editor**
  for nearly everything



And many useful apps like video/audio player,
versions, deleted files, monitoring, …

Motivation

Why we started tubCloud

Motivation (1)

Reasons

Trust | our (research) data is the most valuable resource

⟹ trusted hoster (we) and trusted software (Nextcloud)

User Experience | customers are students, researches etc.

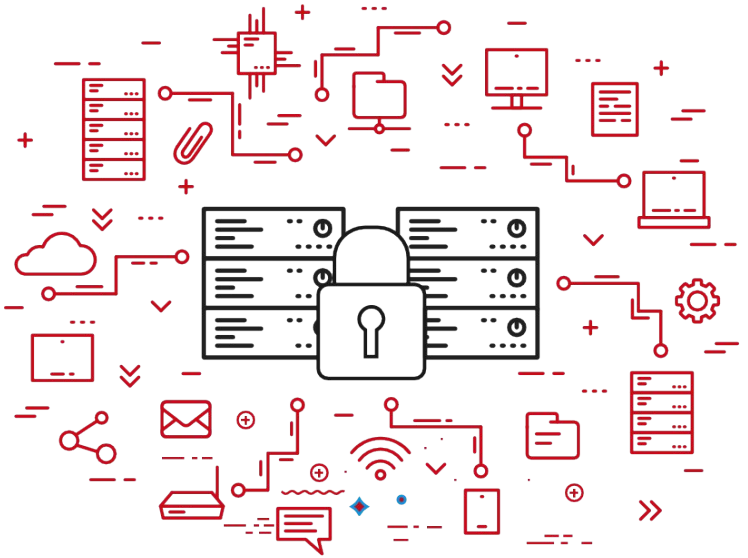
⟹ user-friendly and on every OS and device

Scalability | we've got lots of data and lots of users

⟹ scalability is very important

Integration | many IT systems we have to integrate / be integrated with

⟹ we need interfaces, openness, support



Motivation (2)

Challenges

When started we had...
... no sizing for our infrastructure
... no other big university we could ask
... no idea what our students and researches would do
... and headache about 1,814 things that could go wrong

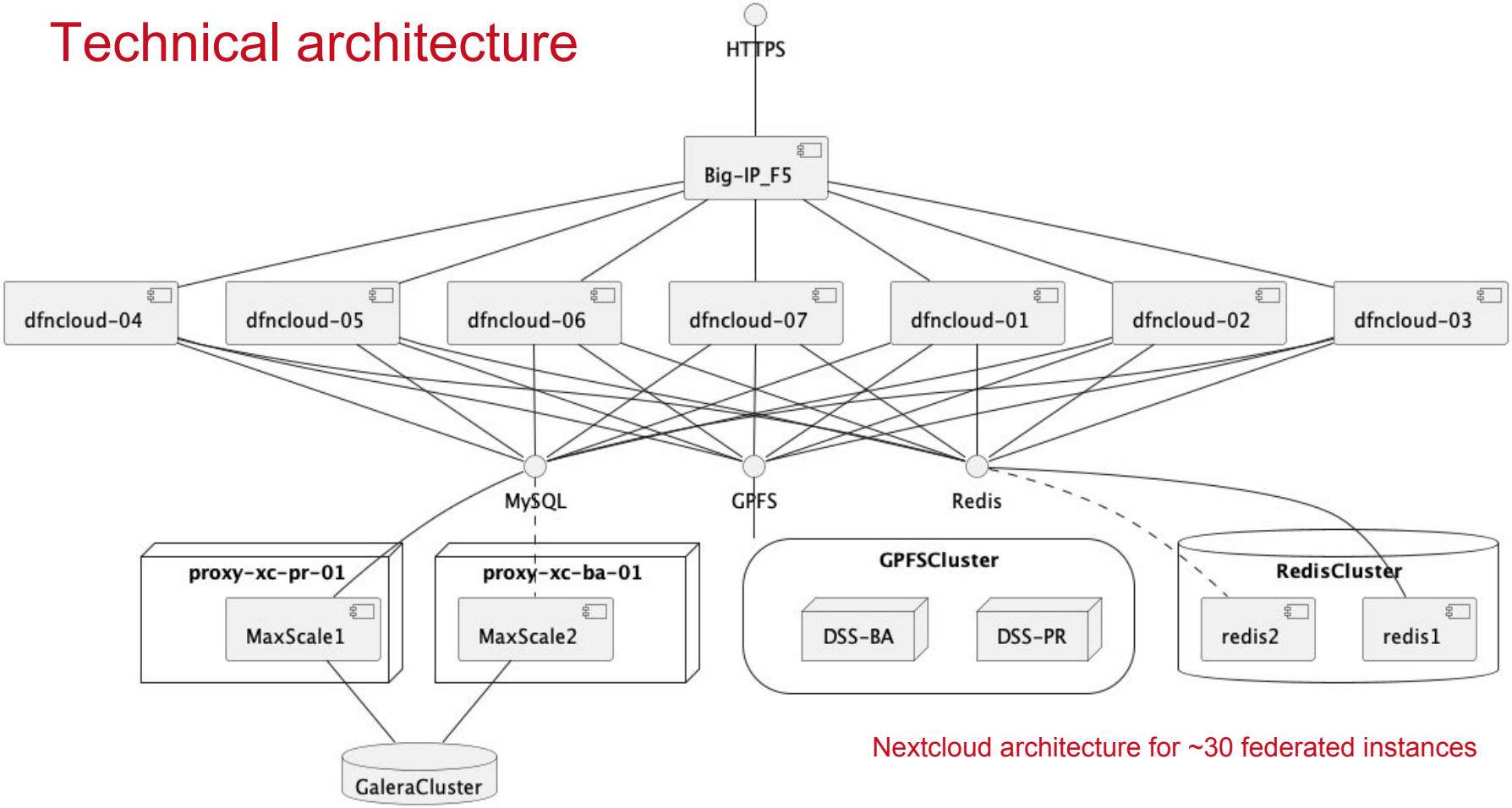
When we started we had ...
... to use workaround using our own group folders
... to report or fix some bugs.

Universities have their own specific set of needs !!

Technical Architecture

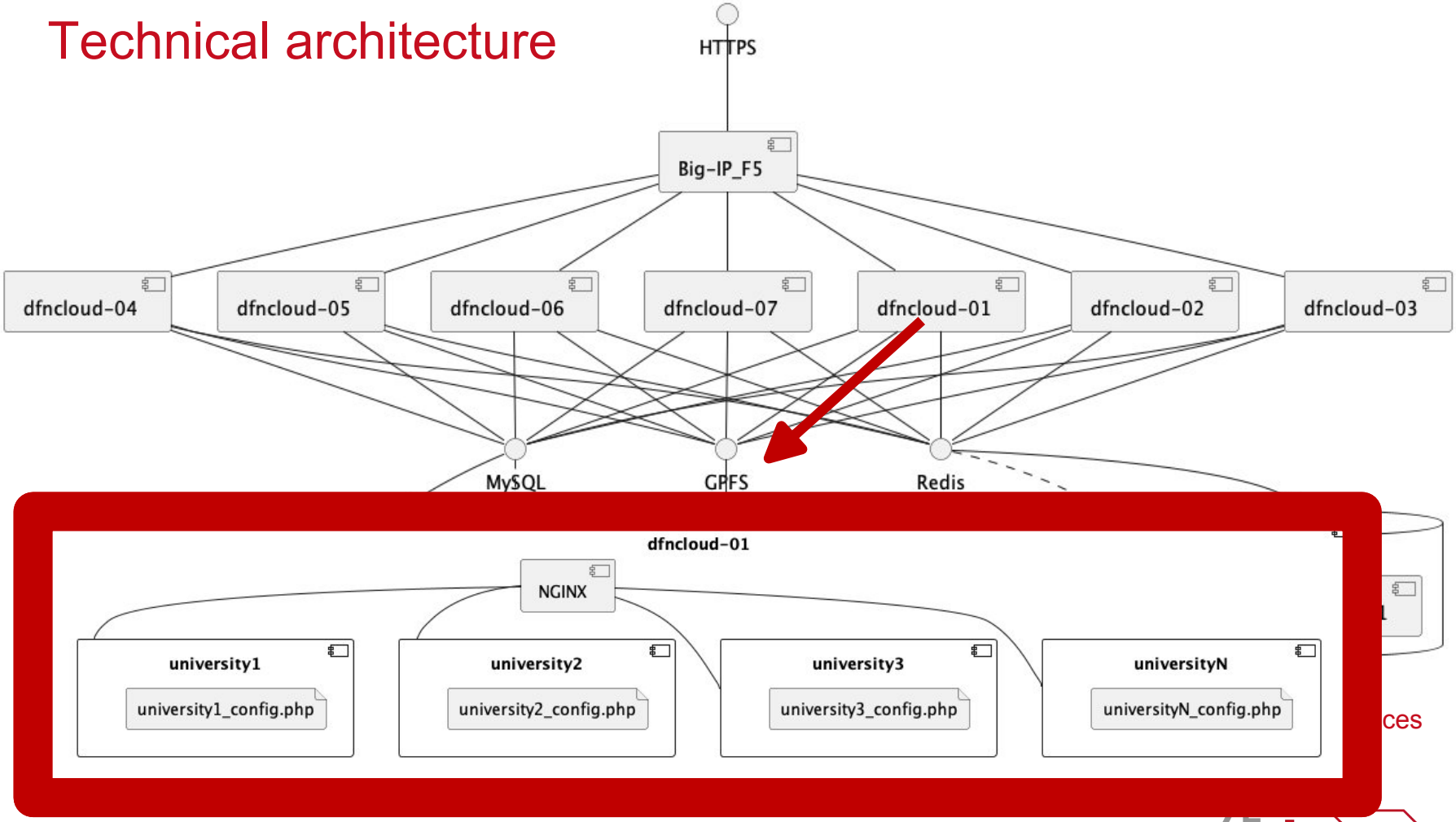
The construction plan of tubCloud

Technical architecture

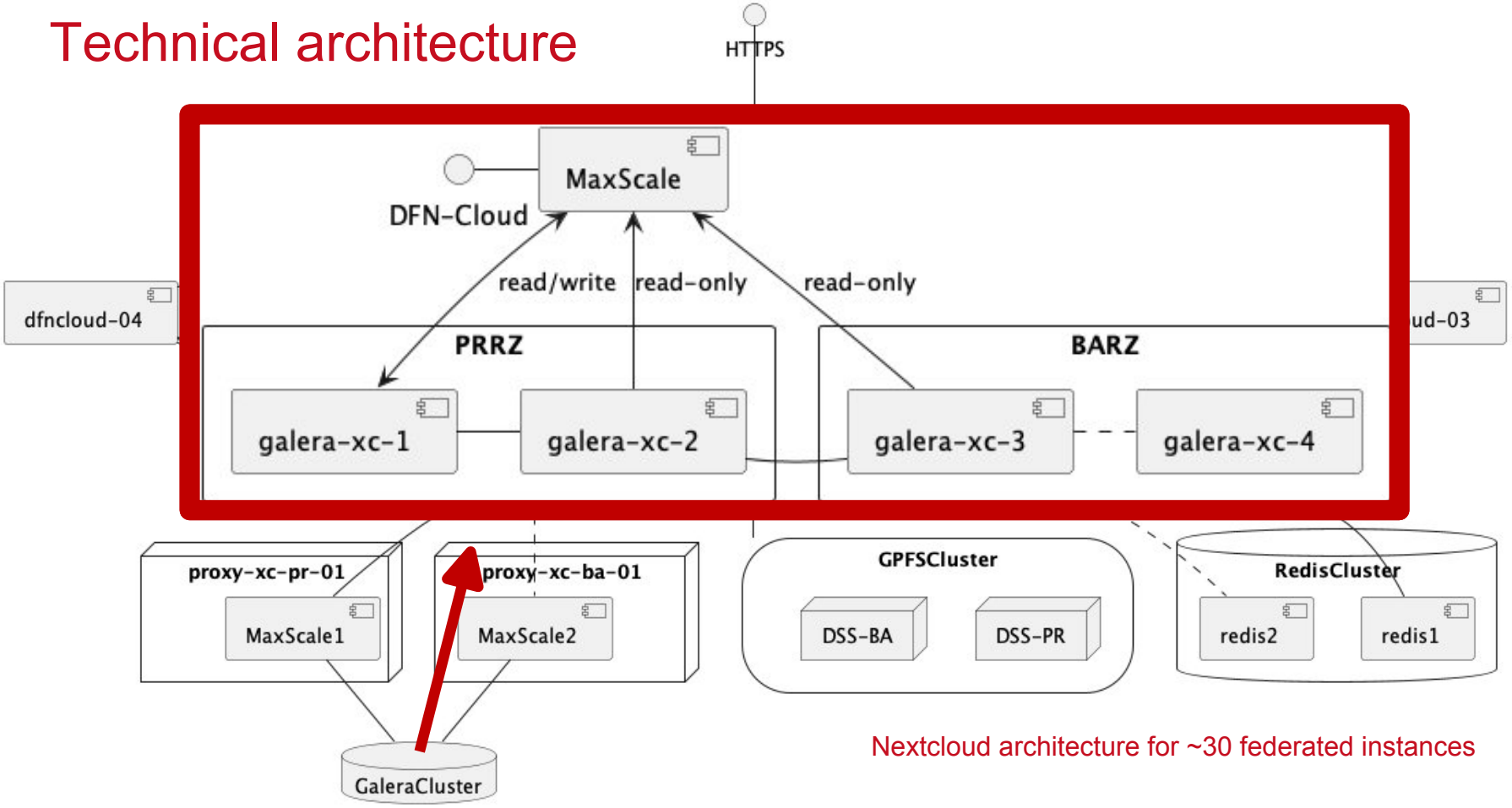


Nextcloud architecture for ~30 federated instances

Technical architecture

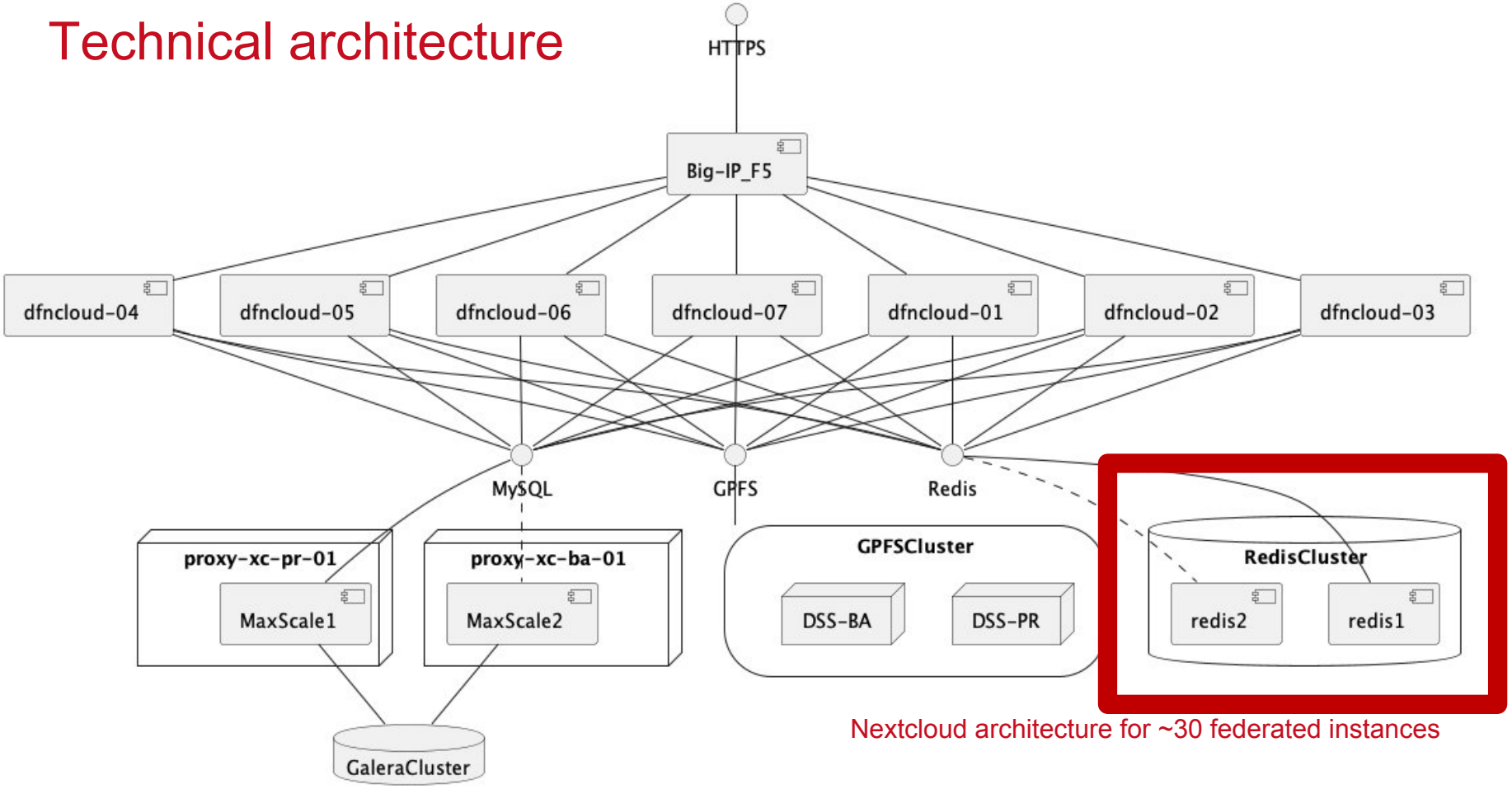


Technical architecture



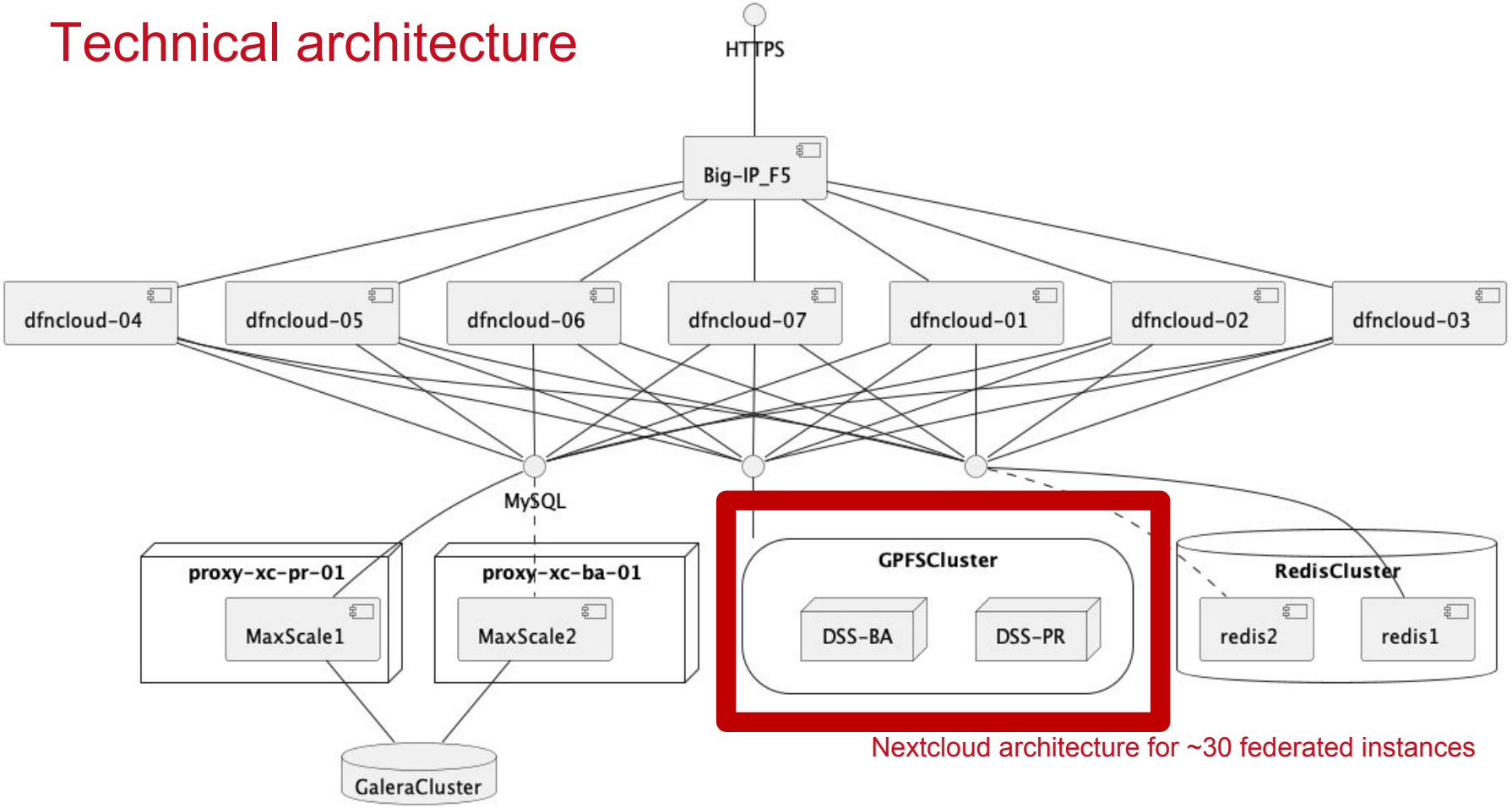
Nextcloud architecture for ~30 federated instances

Technical architecture



Nextcloud architecture for ~30 federated instances

Technical architecture



Nextcloud architecture for ~30 federated instances

Three config files for each cloud instance

mandant_xx-berlin.py:

```
dfn_id=xx
email='dfncloud-report@xxx-xxx.de'
mandant='xx Universität Berlin'
name_datenbank='xx_berlin'
gebuchter_speicherplatz=75000
gekaufte_lizenzen=11000
lizenzen_via_tu=0
startdatum='2016-01-01'
nxtcld=1
office_lizenzen=0
outlook_lizenzen=0
tu_intern=0
```

Quota, Reporting, Billing

config.php-xx_berlin:

```
<?php
$CONFIG = array (
 'datadirectory' => '/gpfs/xcloud/xx_berlin/data',
 'dbtype' => 'mysql',
 'dbname' => 'xx_berlin',
 'dbhost' => '10.176.1.131:4006',
 'dbtableprefix' => 'oc_',
 'dbuser' => 'xx_berlin‘,
...
```

Nextcloud Config file

dfncloud.yaml:

```
 xx_berlin:
   install_crons: true
   user: 'php-xx_berlin'
   uid: '1216'
   gid: '1216'
   www_root: '/var/www/cloud/xx_berlin'
   git_core_tag: 'v23.0.10-w-apps2'
```

Puppet control file to deploy instances

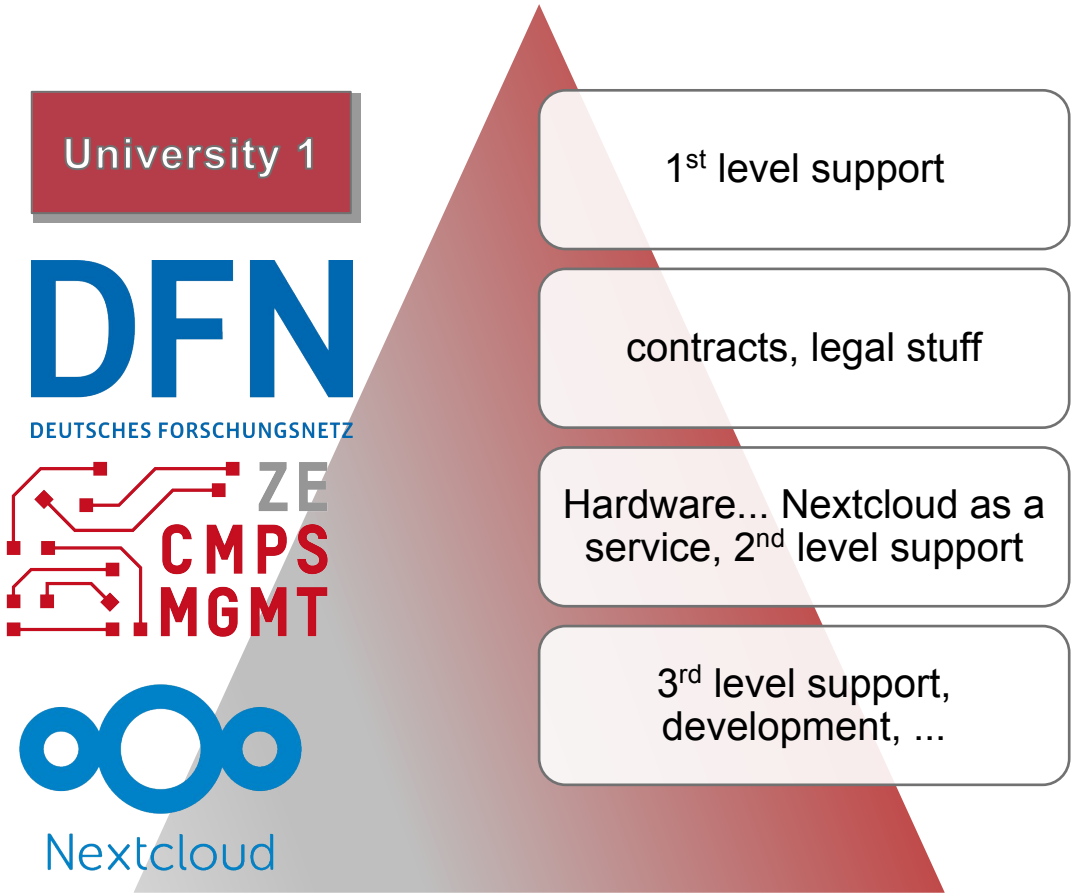
Technische Universität Berlin

Federation

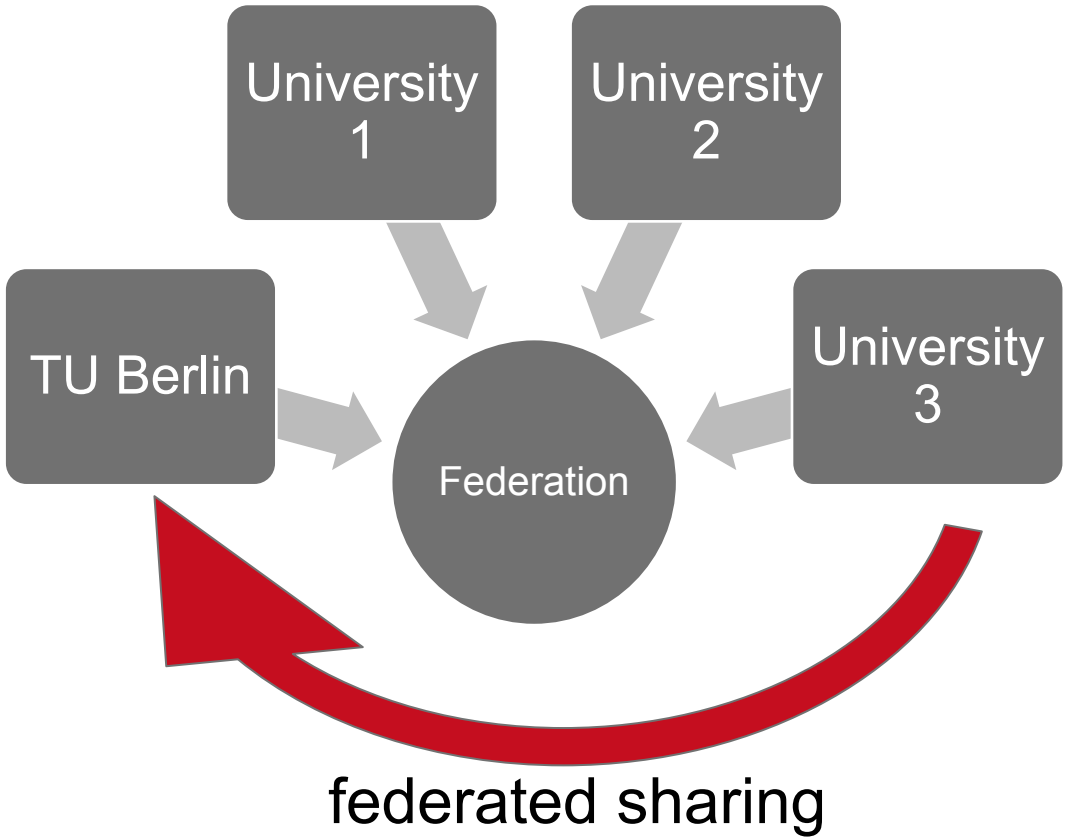
Borderless collaboration

ZE CMPS MGMT

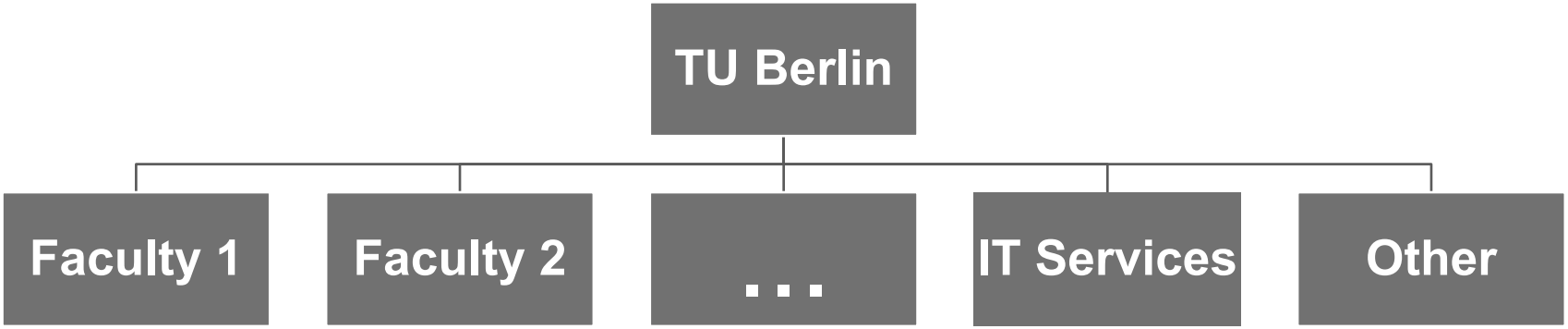
Federation: Organisational Perspective



Federation: Technical Perspective



Outlook: Global Scale = “internal federation“



- Idea is to divide the university into 7-8 smaller Nextcloud instances (35.000 / 7 = 5.000 users and 43 million files per instance)
- Most of shares will still be local because most files are shared within one faculty.
- Our IT services instance can be updated and the new version can be tested in advance.
- Shorter downtimes per organisational unit but more work for administrators.

Outlook: Shibboleth (SAML) logon

Advantages

- Better UX with Single Sign On
  (some day with other universities, e.g. guest students)
- Compatible with Moodle (uses Shibboleth authentication)
- Unified web authentication (avoid phishing)
  with unified 2FA integration (one 2FA to rule them all)
- Less LDAP/AD traffic (uncoupled base systems)
- Just users in system that have logged in at least once

Disadvantages

- Users that never logged in can not be found / shared with.

Technische Universität Berlin

Experiences

in numbers...

ZE CMPS MGMT

Sizing and load

Frontends

6 for tubCloud, 7 for DFN-Cloud: 16 cores, 64 GB RAM, 95 GB disk, Ubuntu 20 LTS

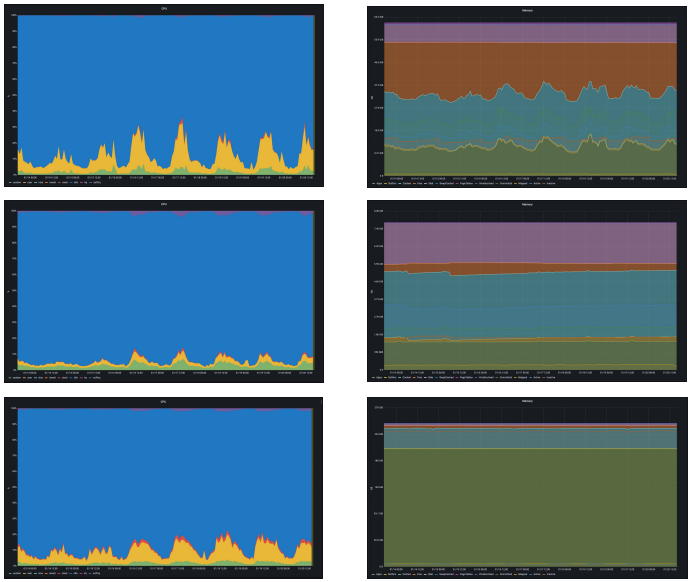
Database

2+2 MaxScale proxy (active-passive): 16 cores, 8 GB RAM and 15 GB disk

4+4 Galera nodes (3 active) tubCloud: 16 cores, 256 GB RAM, 256 GB NVME

Filesystem

GPFS cluster filesystem with 1,7 PB (used by all 35 clouds, 70% full)



CPU Memory

(Wo)manpower involved for tubCloud (1)

- 1 Admin fulltime for all 35 cloud instances
- 1 Database Admin for both cloud Galeras + another Galera and Primary-Secondary cluster
- ½ Person leading, coaching, communicating and connecting doing the admin job for one of the above when not available
- 2 times a year: 1 person some hours for billing etc

(Wo)manpower involved for tubCloud (2)

tubCloud services	Nextcloud services
Our teams providing services for tubCloud	handling 3rd level issues and answering questions
Team Linux managing everything but the applications	standing-by when doing big upgrades
Team Virtualization providing OpenStack, storage, hardware	implementing needed features, fixing bugs and security issues
Team Network routing, firewall, load-balancing, DNS, …	providing new versions opening new use cases and optimizing existing ones
Team Support every organization organizes their own 1st level support and we do the 2nd level support	

Technische Universität Berlin

Monitoring

Keep everything up and running

ZE CMPS MGMT

What to monitor?

Application

- Storage available in data directory
- User: total, active
- Largest db tables (records / size)
- Number of shares
- Requests by User-Agent

Webserver

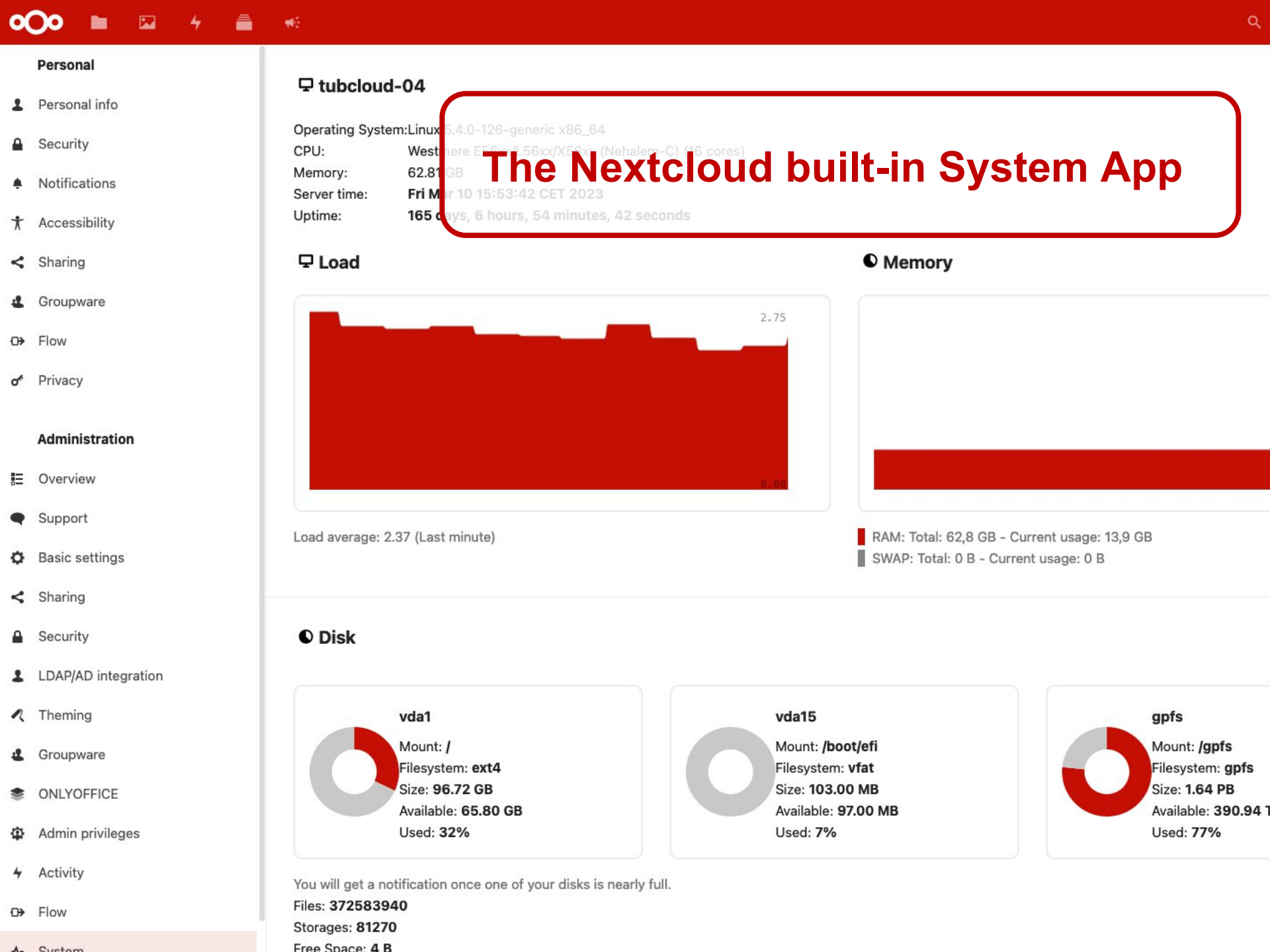
- Requests per second
- Bandwith
- Response Time
- Status of Web server workers
- Number of workes

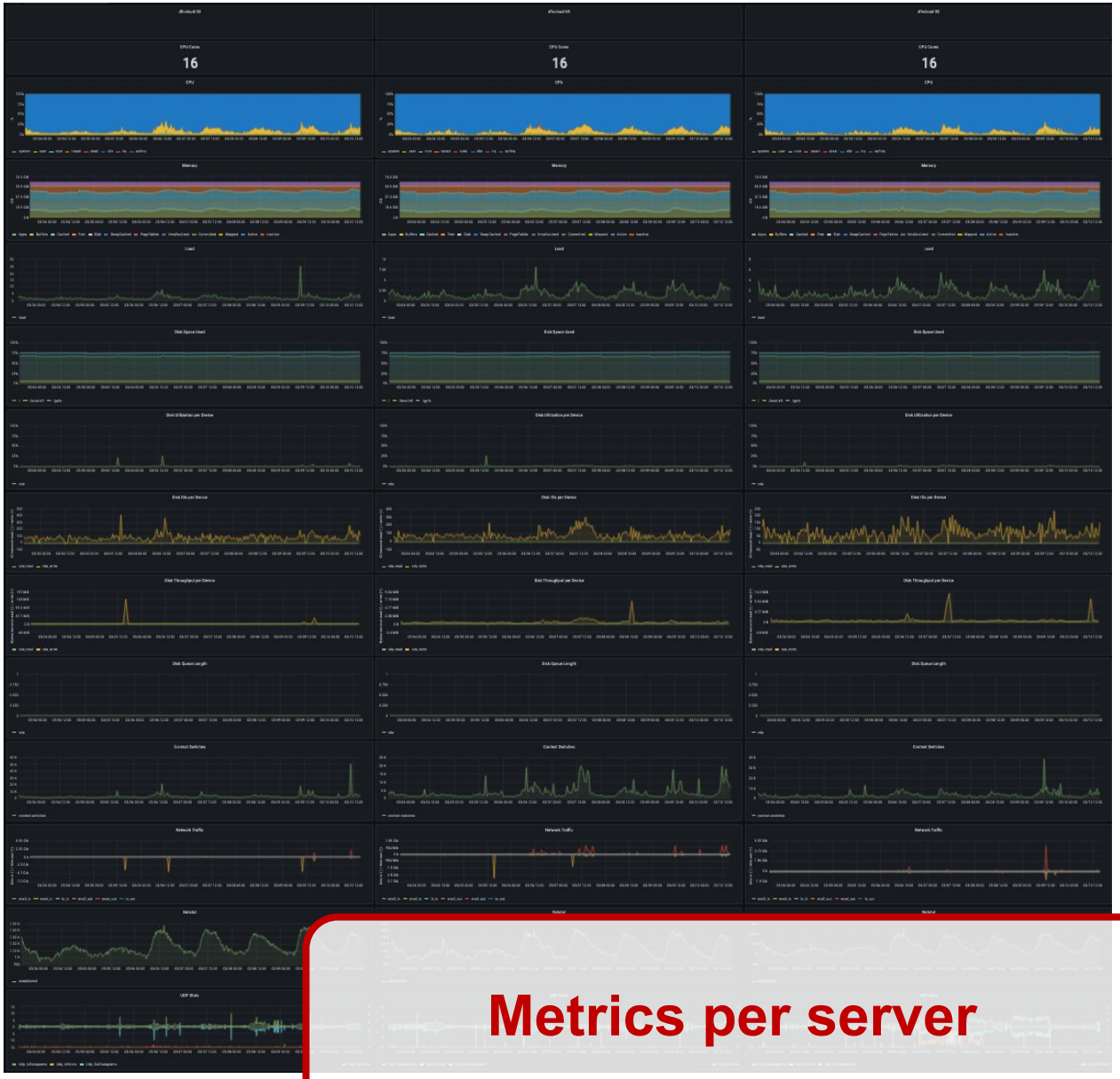
Database

- Number of queries per second
- Slow queries (> 2 sec)
- InnoDB buffer pool memory usage
- Number of connections

Storage

- IPOS
- Avarage wait time (read/write)
- Can I write/read from Filesystem?


Personal
Personal info
Security
Notifications
Accessibility
Sharing
Groupware
Flow
Privacy
Administration
Overview
Support
Basic settings
Sharing
Security
LDAP/AD integration
Theming
Groupware
ONLYOFFICE
Admin privileges
Activity
Flow
tubcloud-04
Operating System: Linux
CPU:
Memory:
Server time: Fri
Uptime: 165
The Nextcloud built-in System App
Load
2.75
0.00
Load average: 2.37 (Last minute)
Memory
RAM: Total: 62,8 GB - Current usage: 13,9 GB
SWAP: Total: 0 B - Current usage: 0 B
Disk
vda1
Mount: /
Filesystem: ext4
Size: 96.72 GB
Available: 65.80 GB
Used: 32%
vda15
Mount: /boot/efi
Filesystem: vfat
Size: 103.00 MB
Available: 97.00 MB
Used: 7%
gpfs
Mount: /gpfs
Filesystem: gpfs
Size: 1.64 PB
Used: 77%
You will get a notification once one of your disks is nearly full.
Files: 372583940
Storages: 81270



Technische Universität Berlin

Metrics per server using Grafana

- CPU / Load
- Memory (RAM)
- Storage (free / IOPS)
- Network Traffic

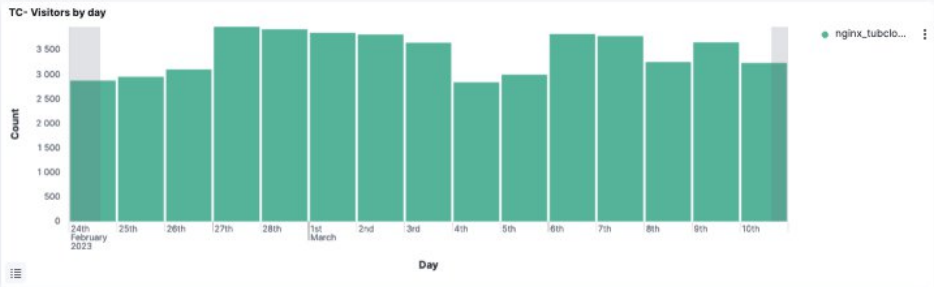
Metrics per server

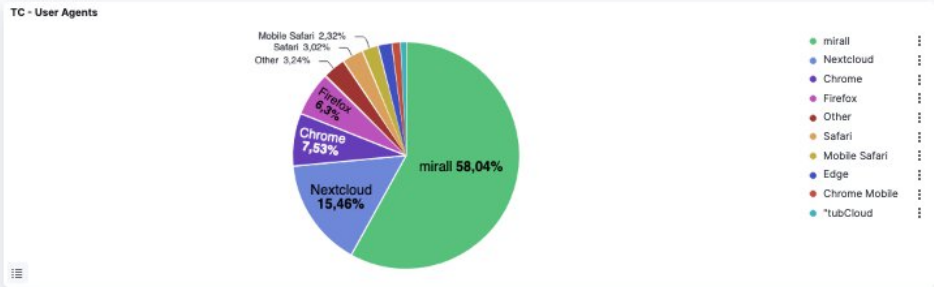
ZE CMPS MGMT

Dashboards with Log Data

Technische Universität Berlin

Dashboards with Log data generated by Kibana



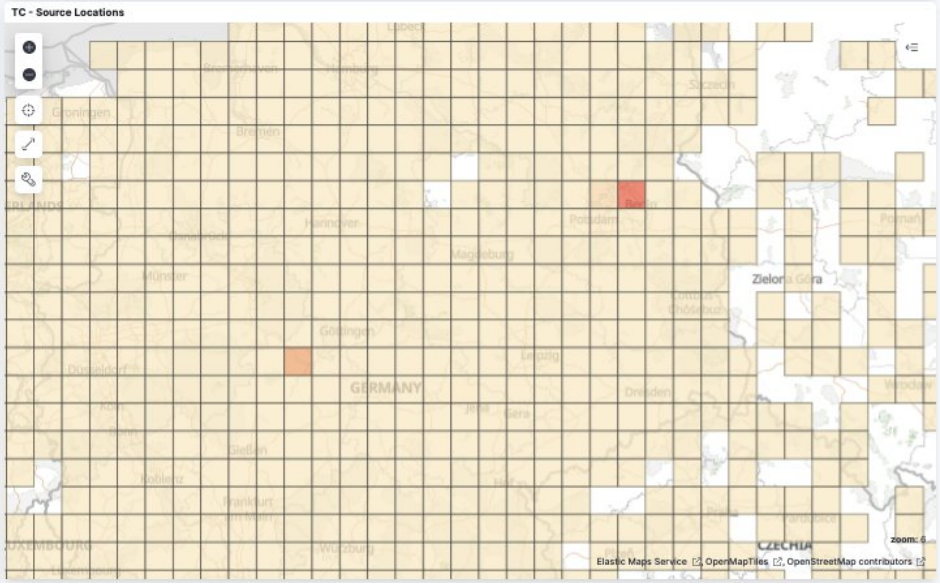


TC- Old Clients

Export

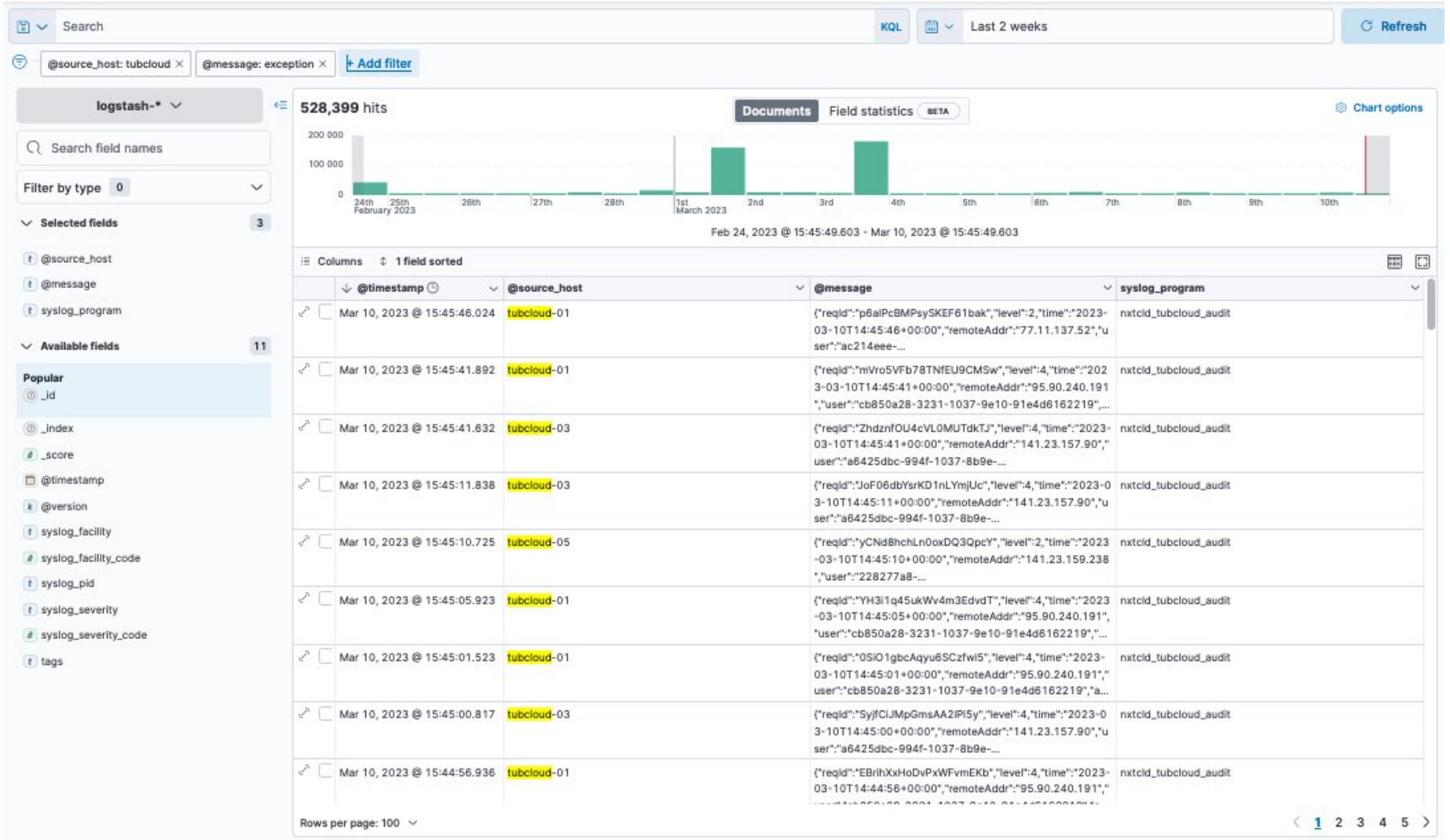
UserID	# of requests
[redacted]	366 233
[redacted]	298 539
[redacted]	44 873
[redacted]	27 166
[redacted]	17 925
[redacted]	13 377
[redacted]	12 452
[redacted]	11 743
[redacted]	11 216
[redacted]	9 809
38	38

1 2 3 4 >



Debugging by finding Log Peaks

Technische Universität Berlin



Technische Universität Berlin

Lessons Learned

Tackling obstacles, implementing solutions

ZE CMPS MGMT

Lessons Learned (1)

Network

⟹ Optimize for latency not for throughput

⟹ Only inter-database and inter-filesystem traffic is really critical

⟹ Don‘t underestimate base-services like DNS, LDAP/AD

Webservers

⟹ We like NGINX. If you like Apache, it works as well.

⟹ 6 frontends can handle more than 30 small to big sized instances

⟹ We use a dedicated frontend for cron jobs and administration.

⟹ LDAP with 2,000+ active users we install local LDAP cache servers.

Lessons Learned (2)

Virtualization

⟹ Be ready for many connections and lots of inter-everything traffic!

⟹ CPU-power is important for network traffic not for the services.

Databases

⟹ Big cloud installations need big DB machines = enough RAM

⟹ At least the oc_filecache table has to fit into InnoDB cache when handling 2,000+ concurrent users.

⟹ DB machines have local SSD storage

⟹ MariaDB and Galera are great

⟹ MaxScale makes live easier

⟹ Performance depends on the DBMS so take care in optimizing, monitoring, …

Support

⟹ Use support for Nextcloud and MariaDB

Conclusion

Synthesizing Insights for Success

Conclusion

1 Today **Nextcloud** is one of the three **most important IT services** at TU Berlin.

2 You have to teamwork with your team, this includes the Nextcloud support team.

3 Having a solid infrastructure setup makes it possible to manage many cloud instances with only few admins.

QA Closing

Remaining questions to address

Used images in presentation

Image on slide 8 (coffee) is a photo by Tyler Nix on Unsplash
https://unsplash.com/@nixcreative
https://unsplash.com/photos/wVexcTg7oXY

Image on slide 16 (formula 1 car) is a photo by Kenny Leys
https://unsplash.com/@kennyleys
https://unsplash.com/photos/JAD82NI3I2Y