Scaling Nextcloud for Universities
Insights from Managing 35 Instances for 50,000 Users
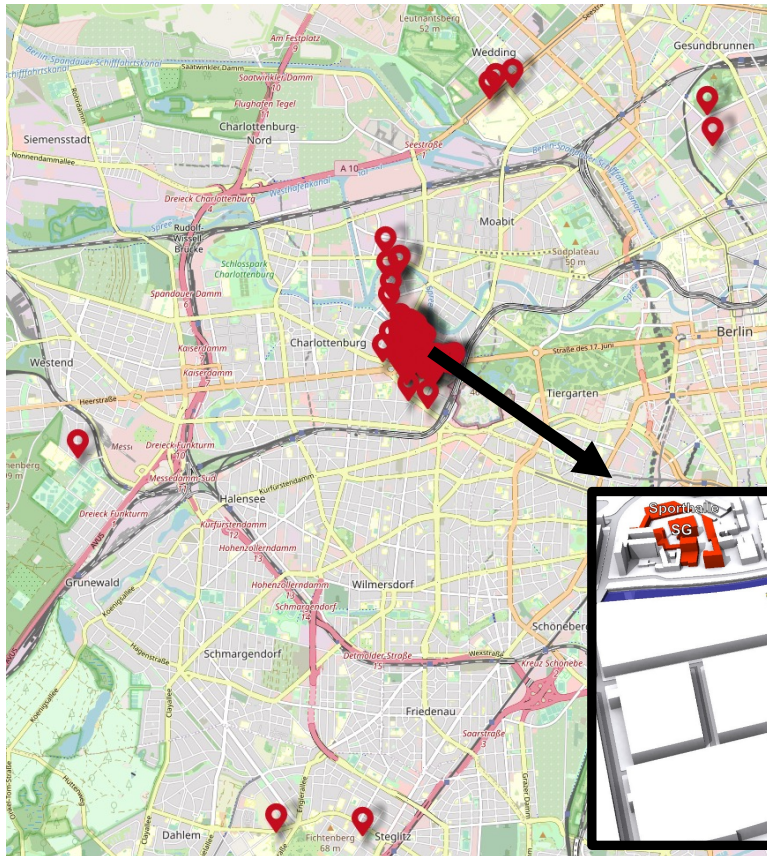
Dr. Thomas Hildmann | ZECM | Webinar February 16th, 2023

Agenda

1. Introduction

2. Motivation

3. Technical Architecture

4. Federation

5. Experiences in numbers

6. Conclusion

Introduction: Technische Universität Berlin

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

Staff: **7243**

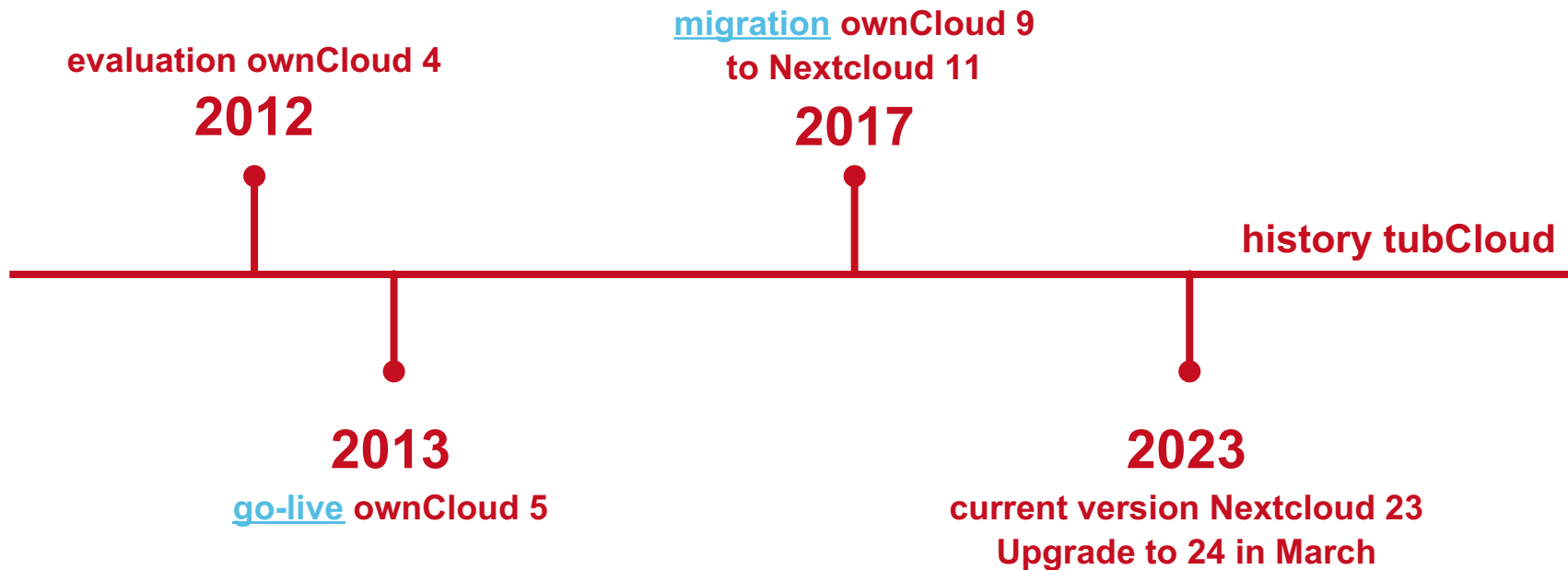**Funding** (2021):
State: **358.8 Mio EUR**
External: **204.4 Mio EUR**

Habilitations (2021): 13

Doctoral degrees (2021): 425

Area: **604,000 square meters**

Introduction: tubCloud (1)

evaluation ownCloud 4
2012

migration ownCloud 9
to Nextcloud 11
2017

history tubCloud

2013
go-live ownCloud 5

2023
current version Nextcloud 23
Upgrade to 24 in March

Introduction: tubCloud (2)

Statistics tubCloud

32,549 users

4,926 active (24h)

2,701 (1h)

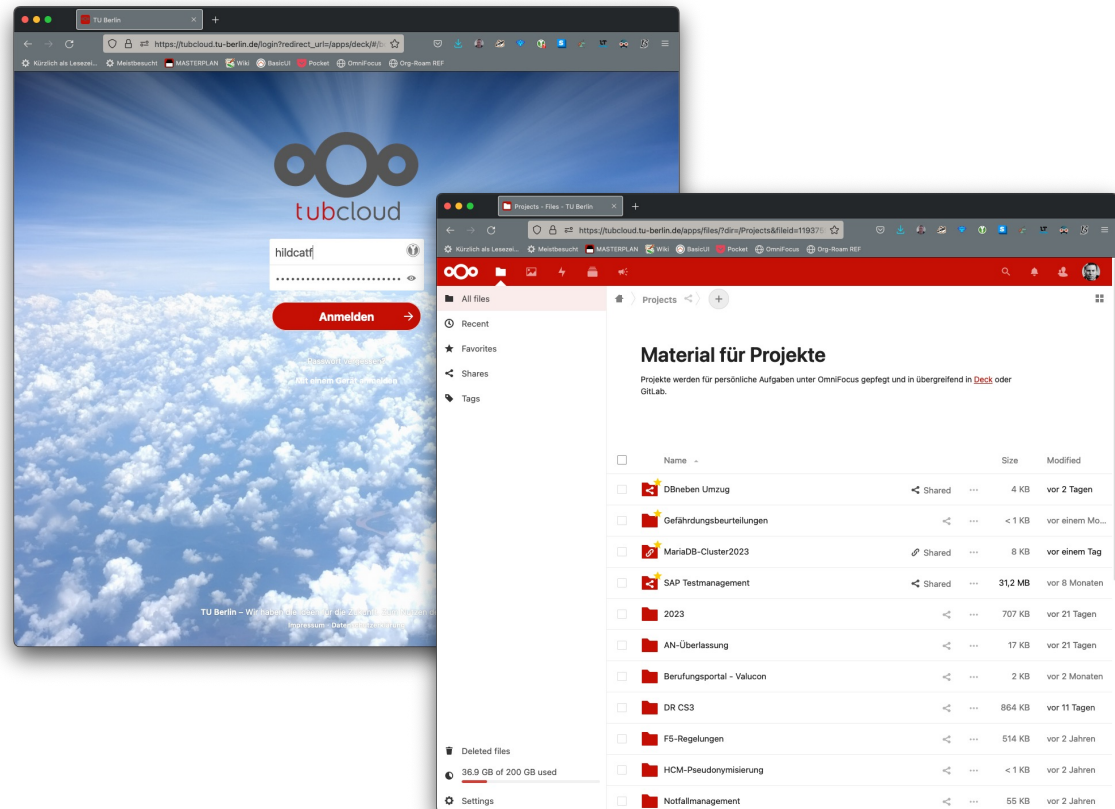**2,076 (5 mins)**

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, …
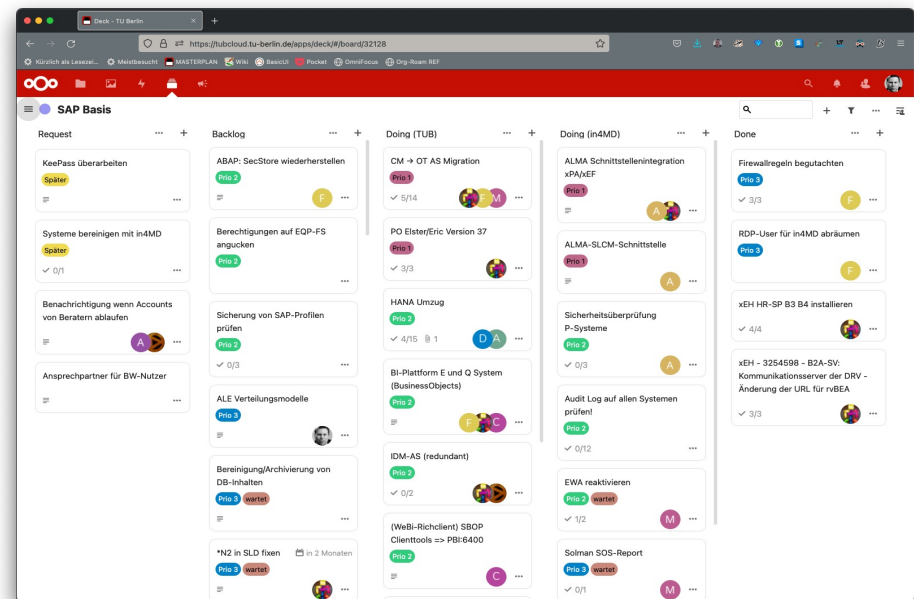
Motivation (1)

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
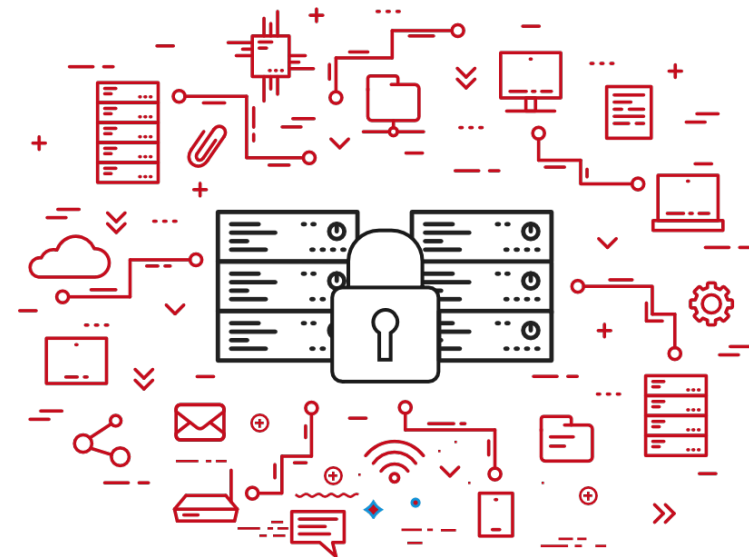
Motivation (2)

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
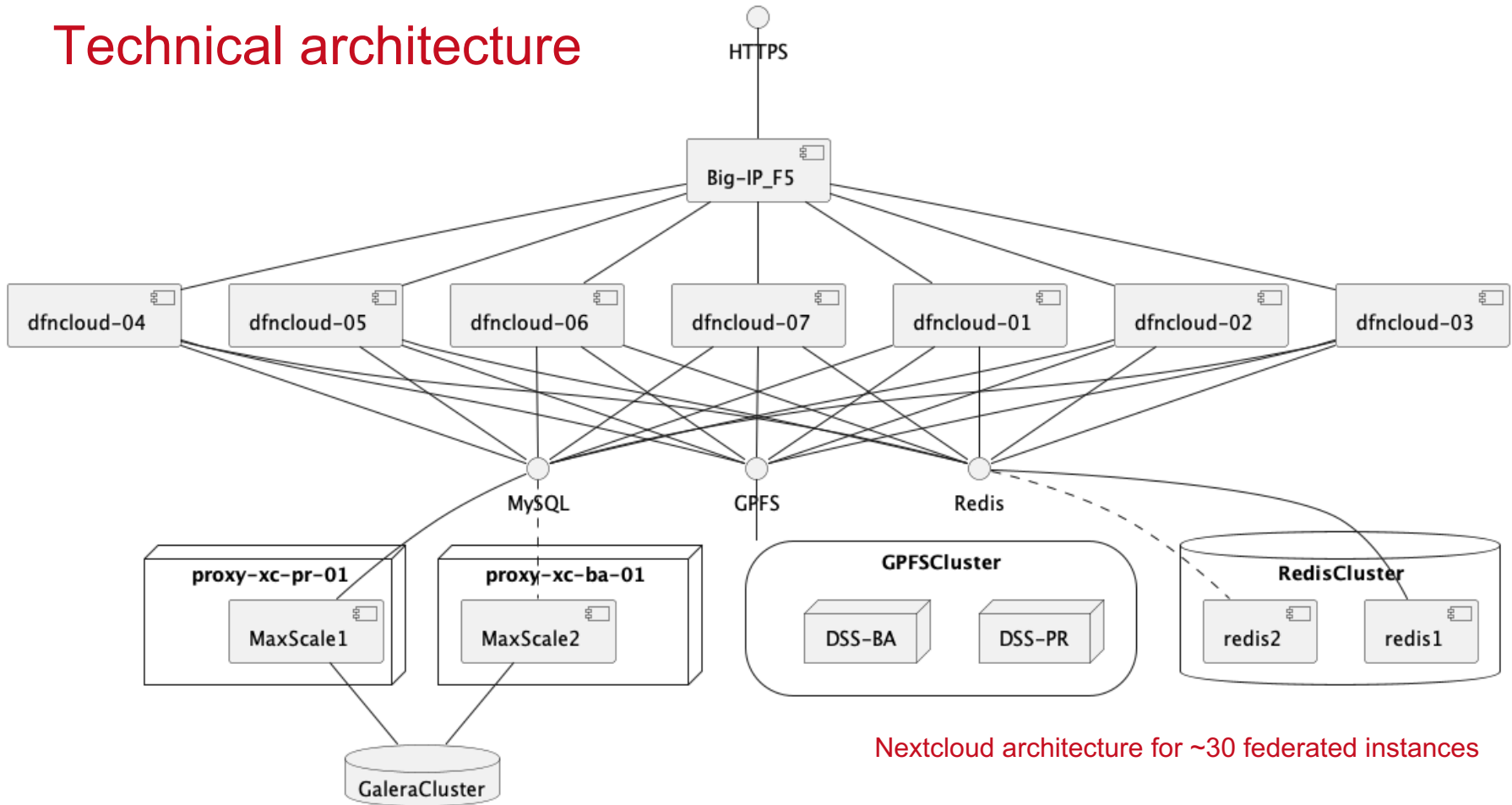
... and headache about 1,814 things that could go wrong

When we started we had ...

... to use workaround using our own group folders

... to report or fix some bugs.
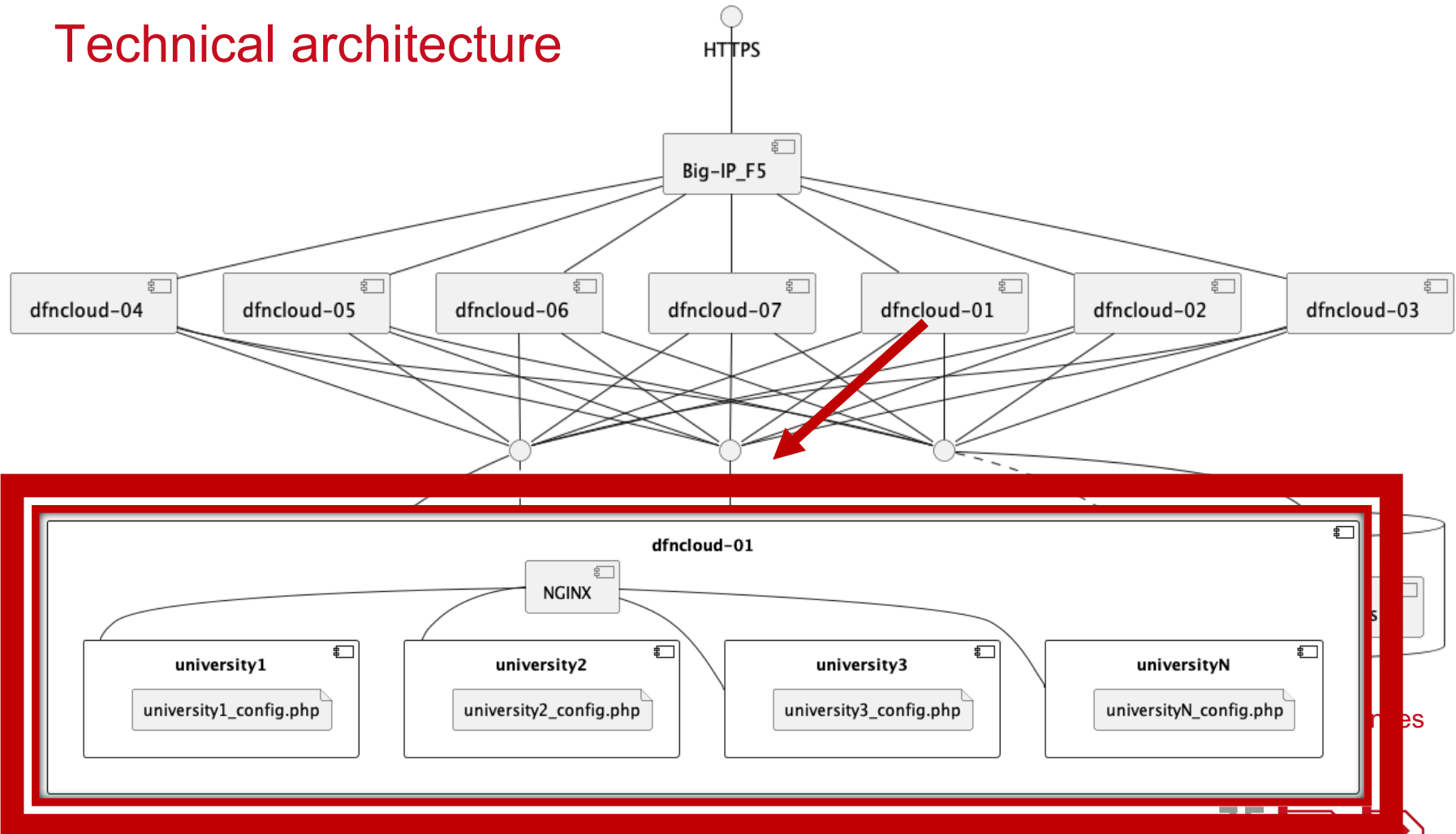
Universities have their own specific set of needs !!

Technical architecture

Nextcloud architecture for ~30 federated instances

Technical architecture

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
<?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',
...
```
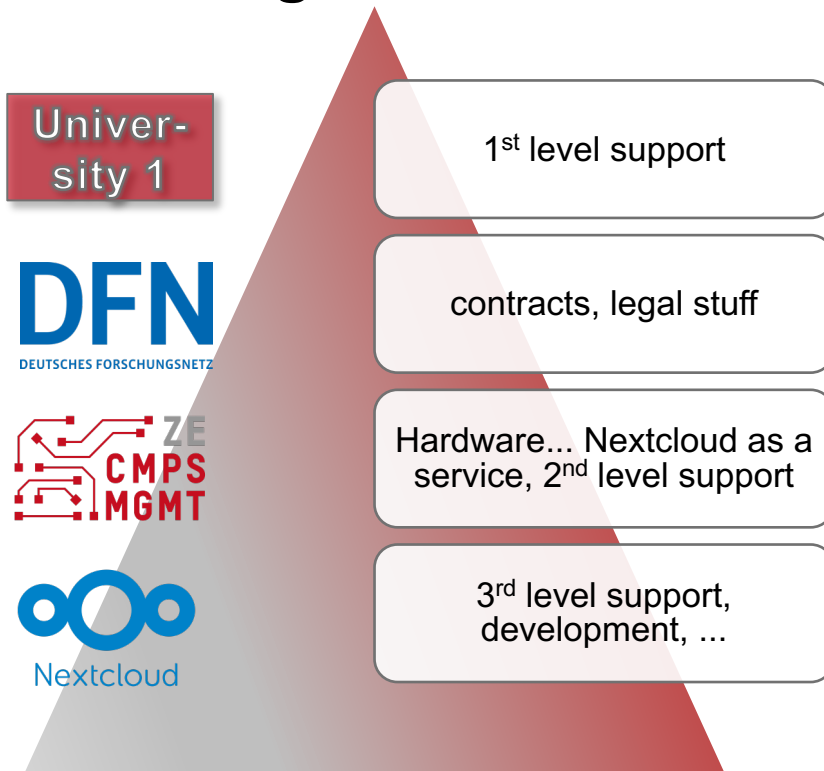
> Config from last slide

dfncloud.yaml:

```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'
```
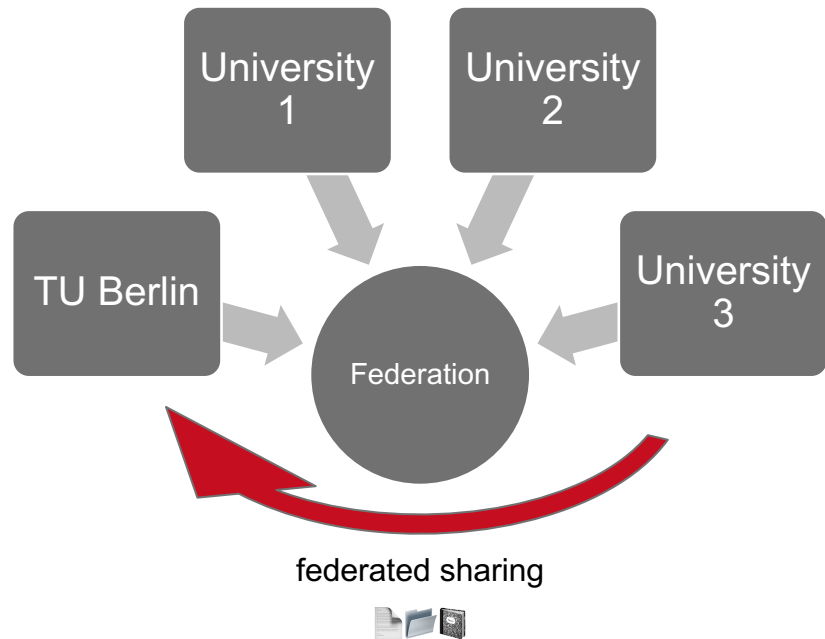
> Puppet control file to deploy instances

Federation

organisational

University 1

DFN
DEUTSCHES FORSCHUNGSNETZ

ZE CMPS MGMT

Nextcloud

1st level support

contracts, legal stuff

Hardware... Nextcloud as a service, 2nd level support

3rd level support, development, ...

technically

University 1

University 2

TU Berlin

University 3

Federation

federated sharing

Sizing and load

Frontends

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

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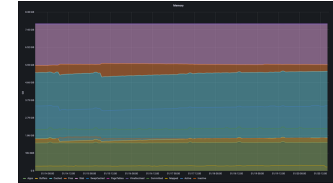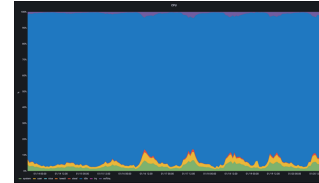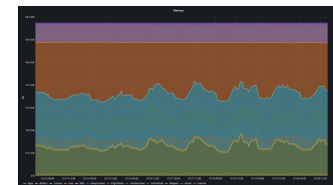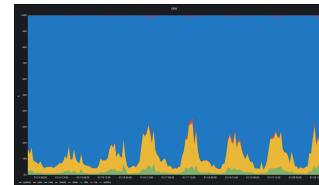
> 2+2 MaxScale proxy (active-passive): 16 cores, 8 GB RAM and 15 GB disk

> 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)
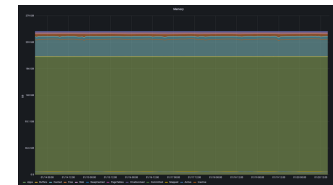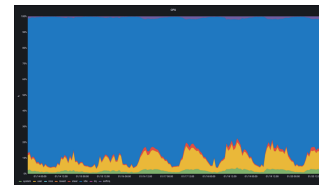
CPU   Memory

(Wo)manpower / human resources for operation

- 1 Admin fulltime for all 35 cloud instances

- 1 Database Admin for both cloud Galeras + an other 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.

- Teams we can delegate things / get services of:
    - Linux-Team: managing everything but the applications
    - Virtualization-Team: providing OpenStack, storage, hardware
    - Network-Team: routing, firewall, load-balancing, DNS, …
    - 1st level support: every organization has its own, we just do the 2nd level stuff

- Nextcloud
    - handling 3rd level issues, answering questions
    - stand-by when doing big upgrades
    - implementing needed features, fixing bugs and security issues
    - providing new versions opening new use cases and optimizing existing ones

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

Today Nextcloud is one of the three most important IT services at TU Berlin. We continuously get really good feedback from our users.

Driving a big Nextcloud installation is sometimes like driving a Formula One car with 300 km/h. Small mistakes have huge impacts. You have to teamwork with your team – this includes the Nextcloud support team.

Having everything up and running you can manage many cloud instances with only few admins.

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/JAD82NI3l2Y