




KEEPER Service - Long term archiving tool for Max-Planck institutions on base of **Seafile**



Vladislav Makarenko, Max Planck Digital Library
<https://keeper.mpdl.mpg.de>
keeper@mpdl.mpg.de



History...

- The Max Planck Digital Library (MPDL) in Munich is a **central unit of the Max Planck Society** that **supports scientists** from all Max Planck Institutes with a broad portfolio of services in the fields of information provision, publication support and research data management.
- We got requests from several Max Planck Institutes to provide a service for research data archiving



Requirements

- Seamless integration into working environment
- Archiving data (10 years)
- Self-hosting
- Versioning
- Project landing pages and Catalogue
- Desktop and Web applications
- Browsing and simple search
- High scalability
- No restrictions on the data format and structure
- Trusted Data



Decided us to go with Seafile™

- Best performance (LBFS – algorithm & deduplication)
- Best stability
- Good scalability (clustering)
- Self-hosting
- Possibility of own implementations (open API)
- Possibility of core development (Seafile Ltd.)
- Easy to use User Interface, clients for all platforms



How we turn Seafire into a long term compliant archive – KEEPER Service



How we turn Seafire into a long term compliant archive – KEEPER Service

- Technical Infrastructure



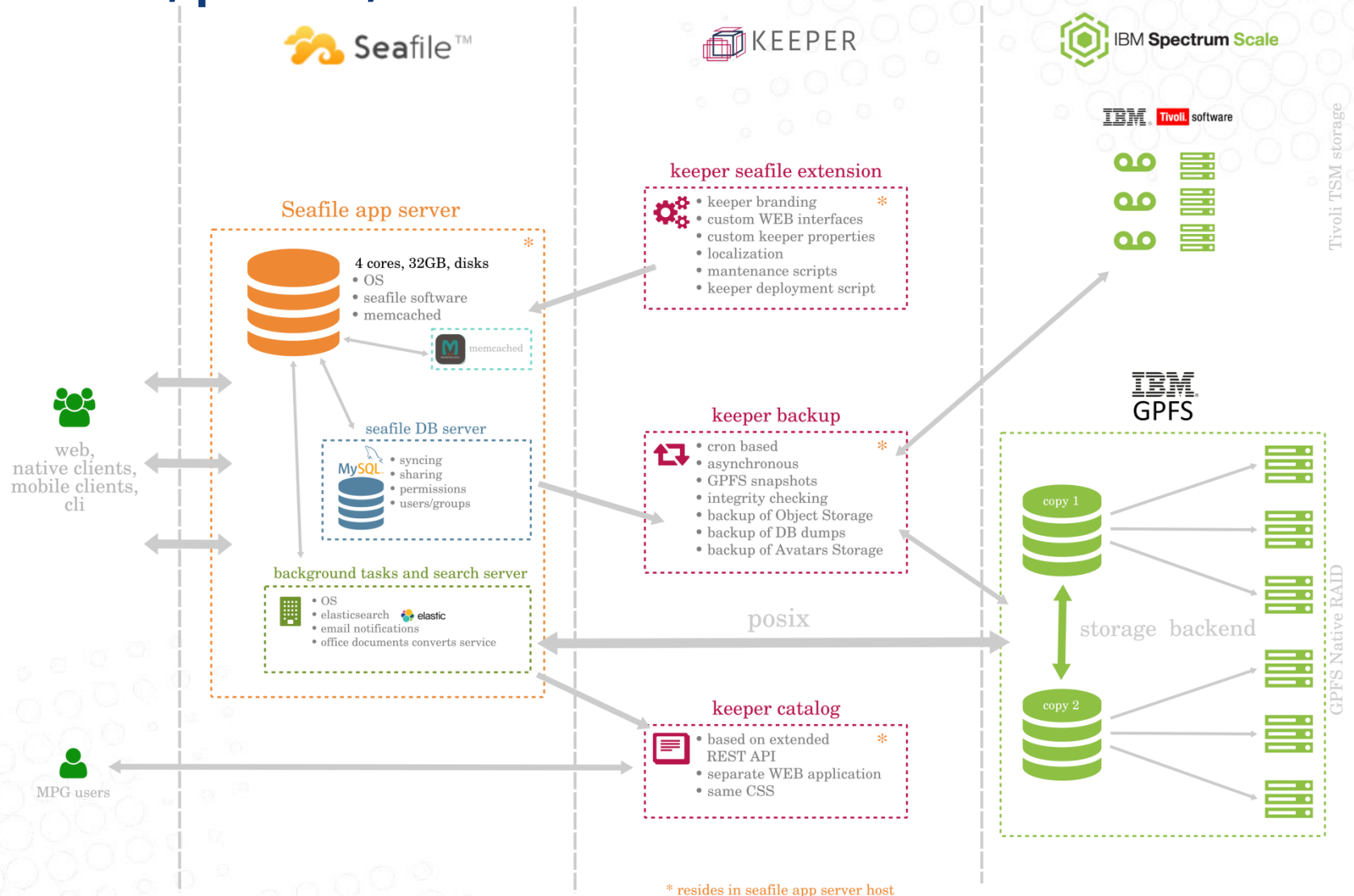
How we turn Seafire into a long term compliant archive – KEEPER Service

- Technical Infrastructure
- Functional extensions - Keeper ext

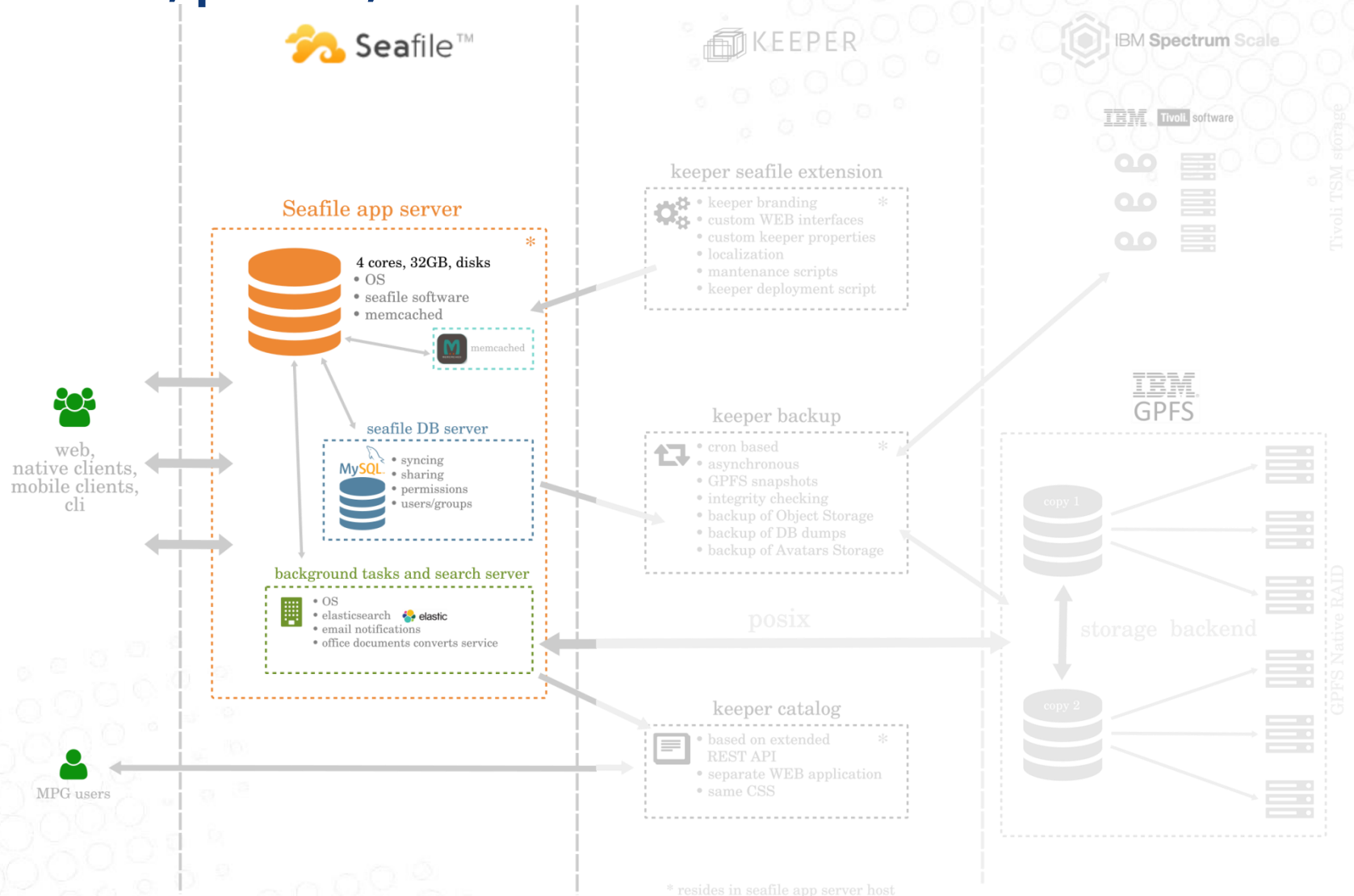
How we turn Seafire into a long term compliant archive – KEEPER Service

- Technical Infrastructure
- Functional extensions - Keeper ext
- Seafire Ltd. support

KEEPER, phase I, architecture

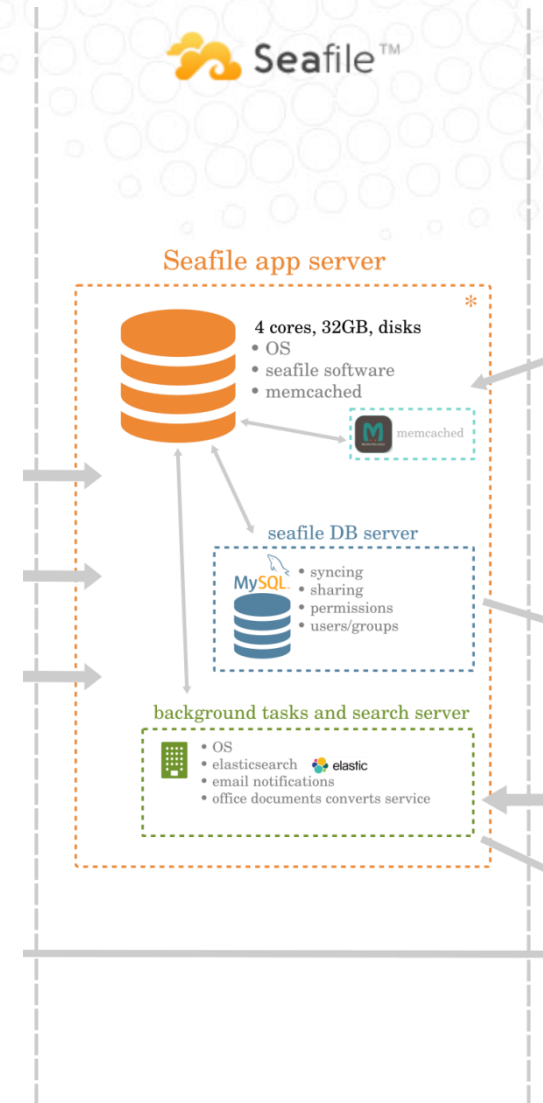


KEEPER, phase I, Seafile

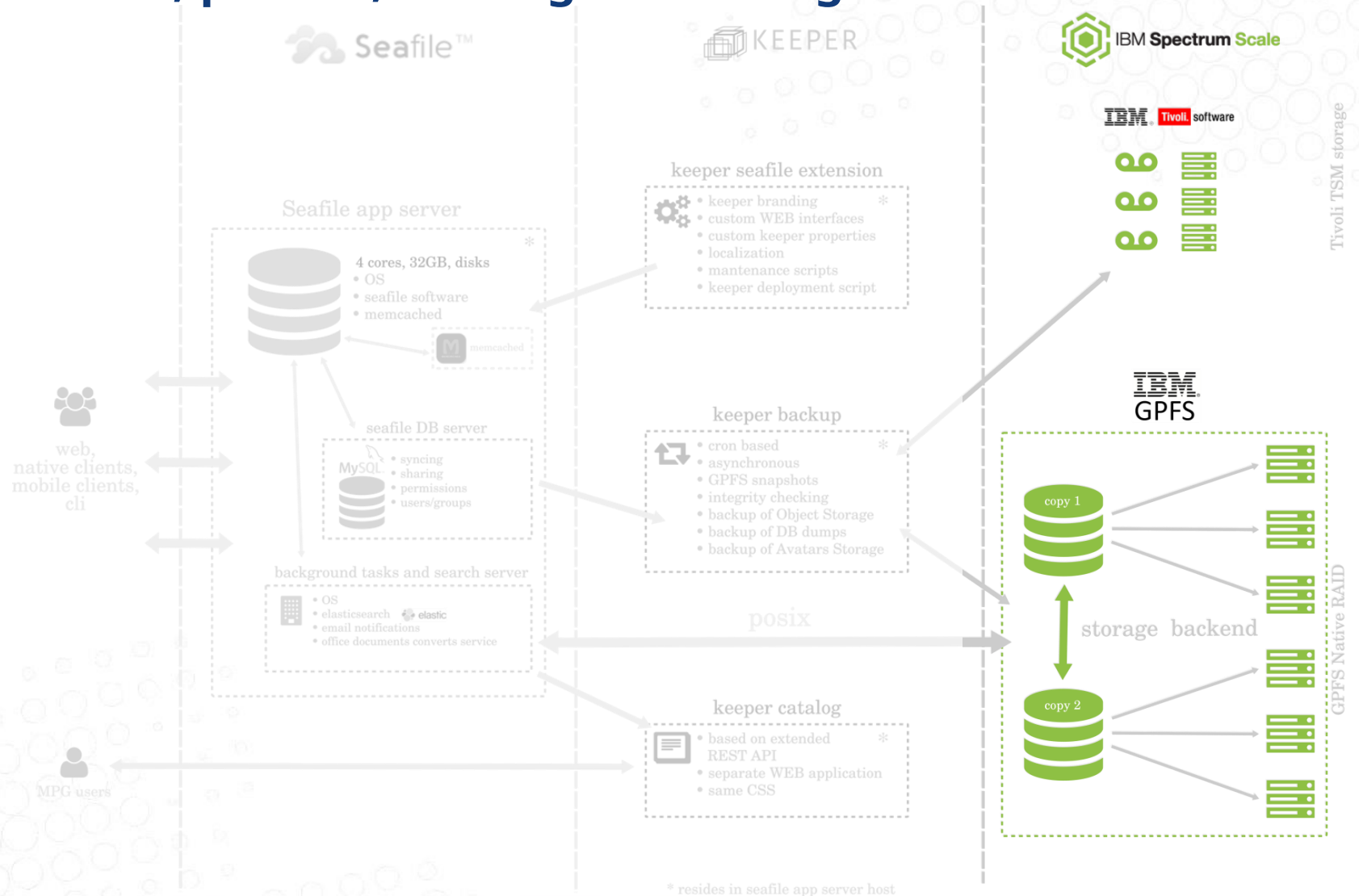


KEEPER, phase I, Seafile

- Seafile-pro 6.1.3
- nginx with https, fastcgi
- memcached
- MySQL
- Elasticsearch
- libreoffice for document previews
- ClamAV antivirus



KEEPER, phase I, hosting and storage backend



KEEPER, hosting and storage backend

Managed by the Max Planck Data and Computing Facility

Virtual Machines:

32GB RAM, 4 cores Intel Xeon(R) CPU E5-2680 v2 @ 2.80GHz

Object Storage:

IBM General Parallel File System, v 4.2

mirrored to ensure HA

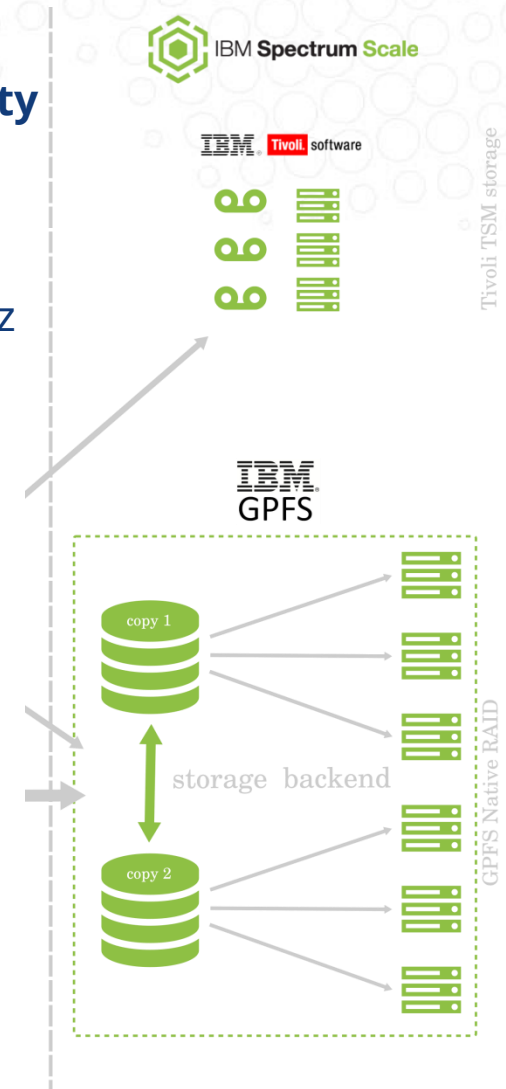
Available disk Space: 2x35 TB, max inodes: 8 Mlrd.

Backup:

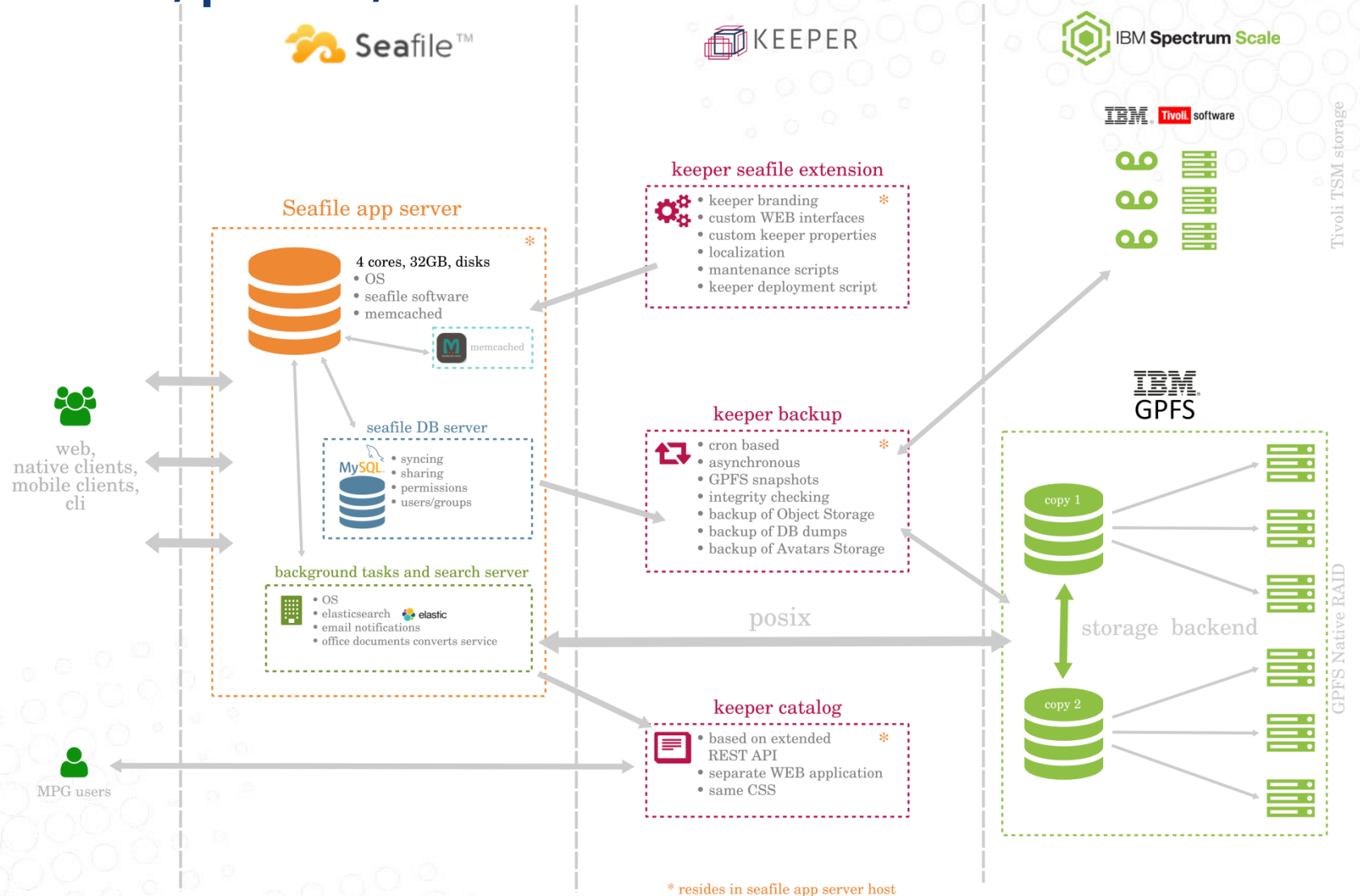
Tivoli TSM Manager

asynchronous daily incremental backup

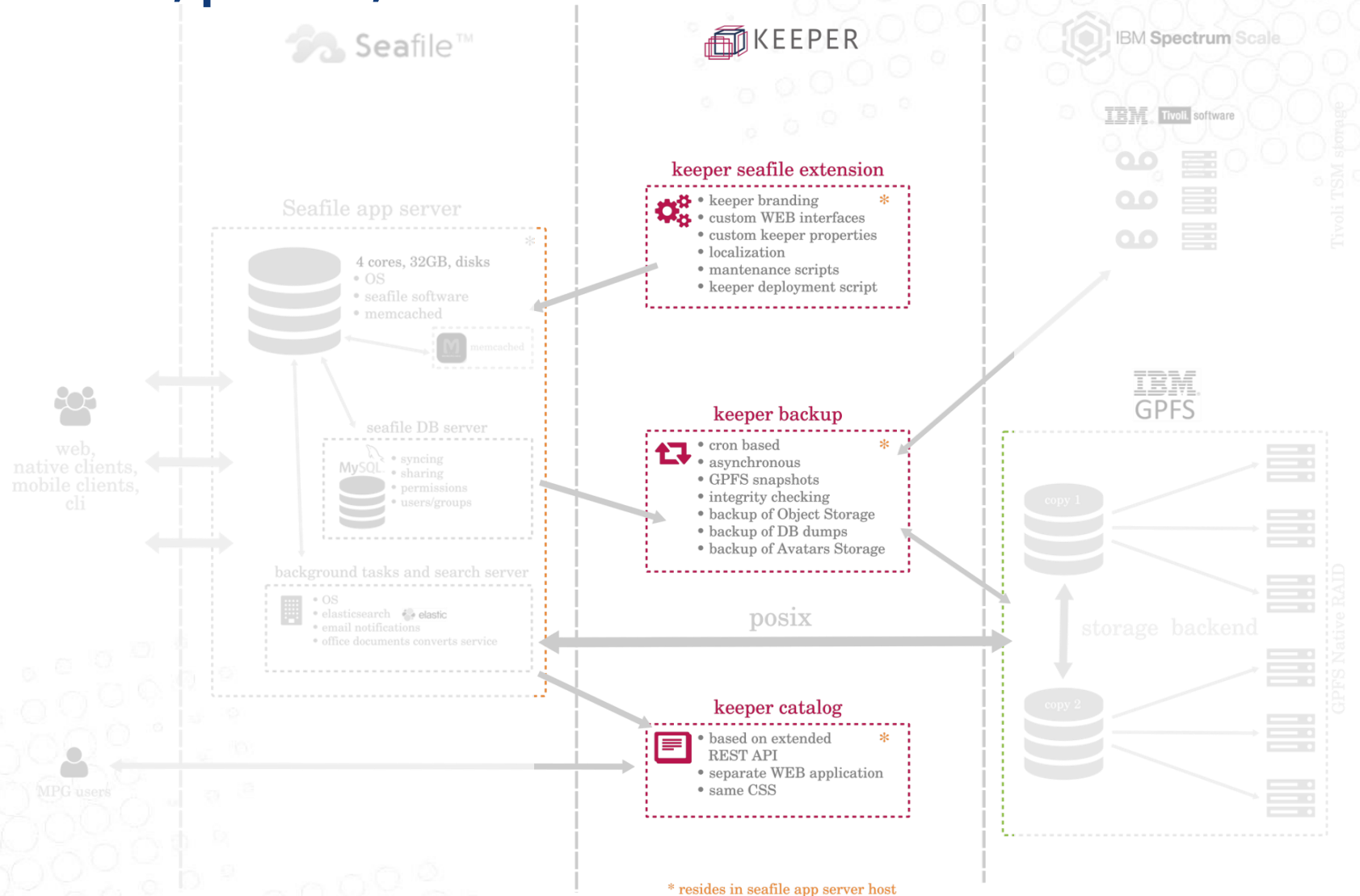
of DB & Object Storage



KEEPER, phase I, ext



KEEPER, phase I, ext



KEEPER ext

set of Seafile patches and KEEPER specific features:

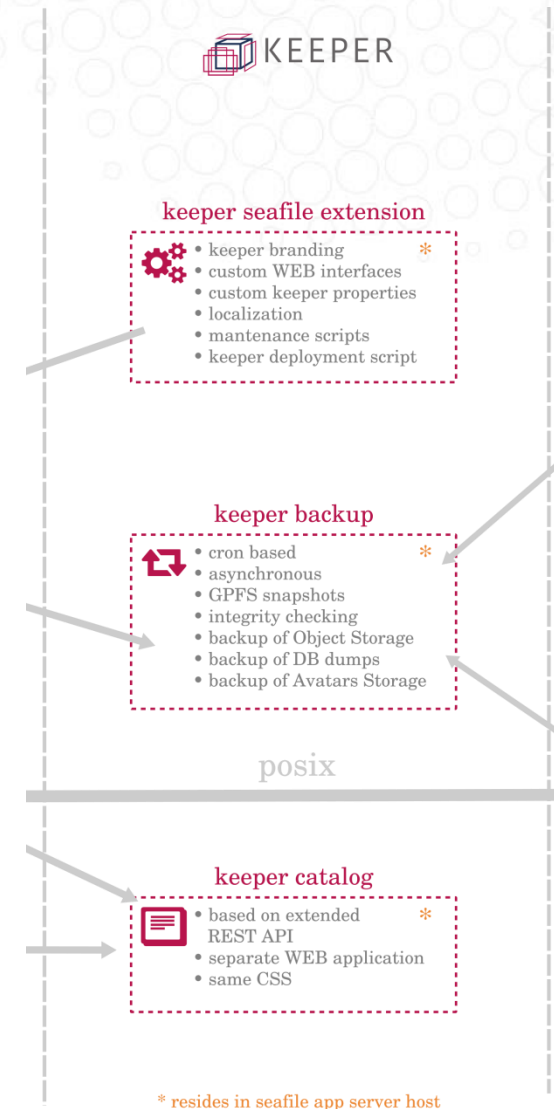
https://github.com/MPDL/KEEPER/tree/master/seafile_keeper_ext

Keeper branding

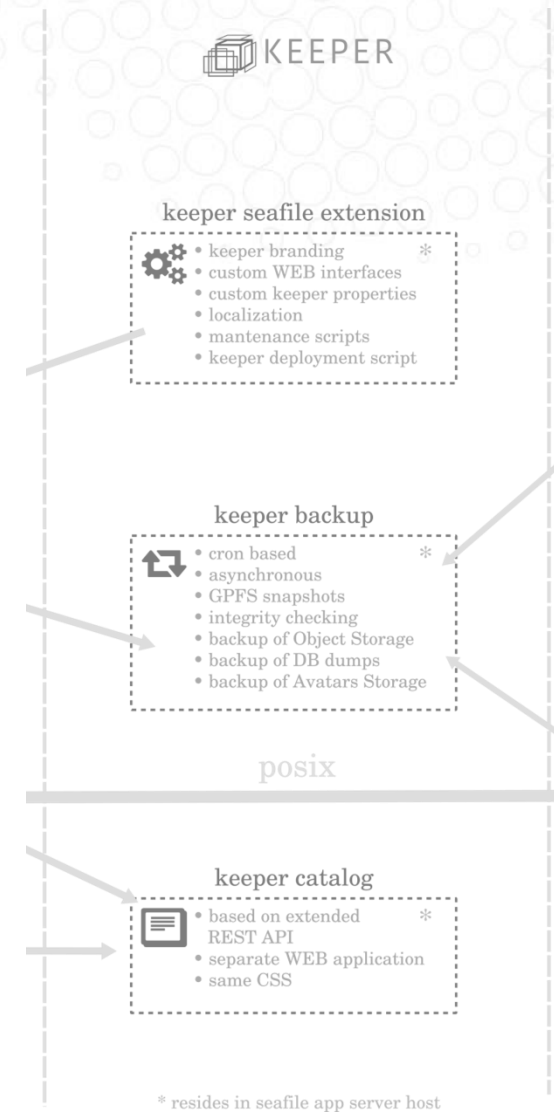
- Colors, Logo, Login Page, slight interface updates, footer, localizations, custom emails

Keeper specific features

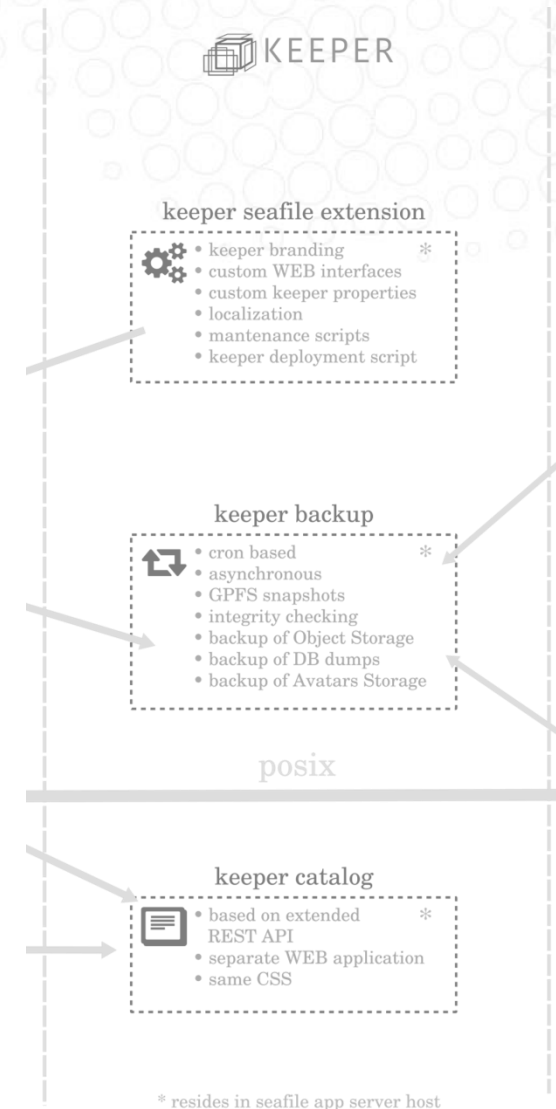
- Cared Data Certificate
- Project Catalogue
- MPG IP checking
- Auto-registration for MPG users



KEEPER ext, Cared Data Certificate



KEEPER ext, Cared Data Certificate

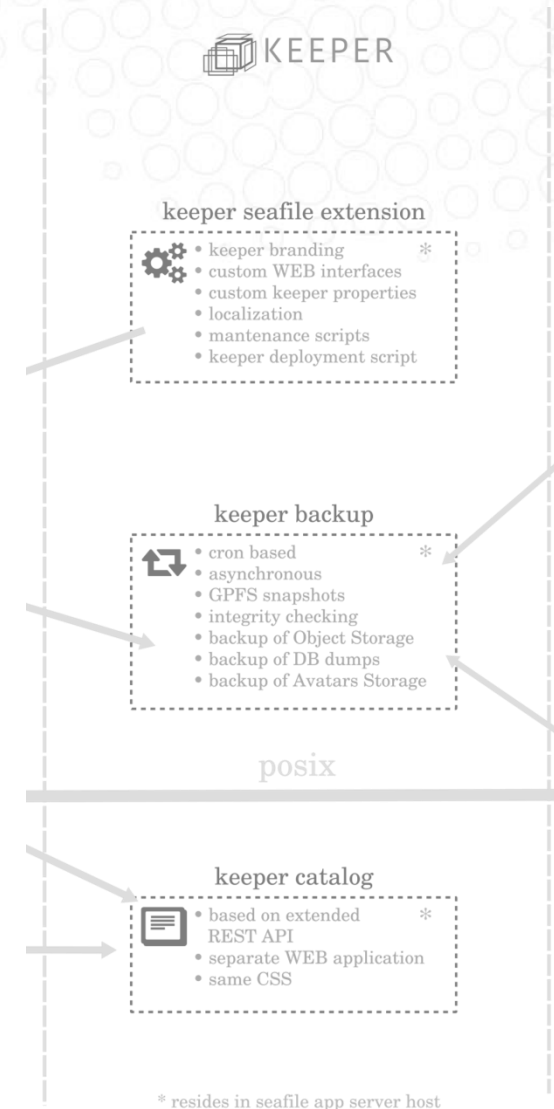


KEEPER ext, Cared Data Certificate



Archive Metadata

- Author
- Description
- Title
- Year
- DOI



KEEPER ext, Cared Data Certificate



Archive Metadata

- Author
- Description
- Title
- Year
- DOI



 KEEPER

keeper seafile extension

- keeper branding *
- custom WEB interfaces
- custom keeper properties
- localization
- maintenance scripts
- keeper deployment script

keeper backup

- cron based *
- asynchronous
- GPFS snapshots
- integrity checking
- backup of Object Storage
- backup of DB dumps
- backup of Avatars Storage

posix

keeper catalog

- based on extended REST API *
- separate WEB application
- same CSS

* resides in seafile app server host

KEEPER ext, Cared Data Certificate



October 12, 2017



No 25

CARED DATA CERTIFICATE

for the research project

Seafile as long term archiving tool for Max-Planck institutions

Makarenko, Vladislav; Max-Planck Digital Library

The Keeper presentation at Seafile Workshop 2017, on October 19th and 20th, 2017 in Strasbourg

Contact: makarenko@mpdl.mpg.de
<https://keeper.mpd.mpg.de>

*Max Planck Digital Library herewith certifies that the data of this research project is archived
[long term compliant]*

and according to the rules for data storage of the Max Planck Society.

The Max Planck Digital Library provides the KEEPER Service to support good scientific practice for Max Planck Researchers and their associates.



extension

g *
interfaces
properties
ripts
ent script

akup


*
s
ng
t Storage
umps
ars Storage

alog

led *
application

* resides in seafile app server host

KEEPER ext, Project Catalogue


English ▾

Institute

Alle

Institut A

Institut B

Institut C

Institut D

Institut E

Projektkatalog

Project in progress;
Contact: kleinfercher@mpdl.mpg.de

Project in progress;
Contact: gkhen@mpiwg-berlin.mpg.de

Project in progress;
Contact: knaus@mpdl.mpg.de

Project in progress;
Contact: hofmann@mpdl.mpg.de

me in test mit leerer bib
fjdshjkadh sdjfhjdhs dsjfhds dsjfhds hdjfhadh sdf
Contact: kleinfercher@mpdl.mpg.de

Post-traumatic sleep-wake disturbances are common after acute traumatic brain injury. Increased sleep need per 24 h and excessive daytime sleepiness are among the most prevalent post-traumatic slee...
measures in the acute phase after traumatic brain injury has so far not been addressed in a controlled and prospective approach. We therefore performed a prospective controlled clinical study to ex...
Contact: kleinfercher@mpdl.mpg.de

金属と金属の掛け合わせ。超耐熱合金のように特性が何倍にも強くなる組み合わせもあれば、反対に弱くなるものも。でも、「弱い=悪い」ではありません。弱いからこそ、できること。
金属と金属の掛け合わせ。超耐熱合金のように特性が何倍にも強くなる組み合わせもあれば、反対に弱くなるものも。でも、「弱い=悪い」ではありません。弱いからこそ、できること。
Contact: kleinfercher@mpdl.mpg.de

Project in progress;
Contact: knaus@mpdl.mpg.de

Project in progress;
Contact: kolter@mpdl.mpg.de

Project in progress;
Contact: knoth@mpdl.mpg.de

vorherige 1 2 3 4 5 6 7 8 9 10 11 nächste


What you need to know

About Keeper
Cared Data Commitment
Terms of Service

Desktop Client


Download the Keeper client for Windows, Linux and Mac

A service by

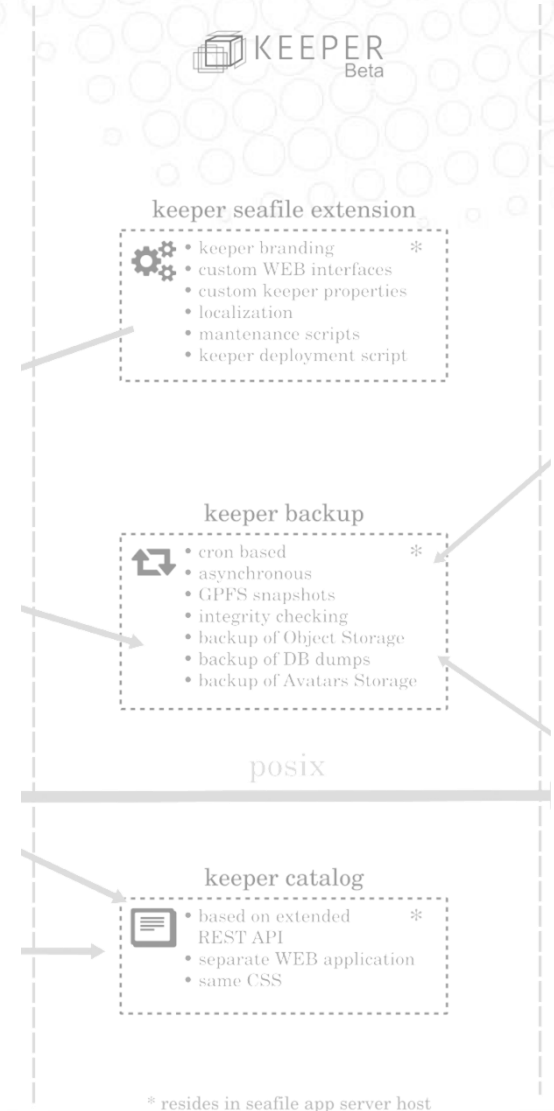
 MAX PLANCK digital library

Help
Keeper Support
Impressum

The software behind Keeper

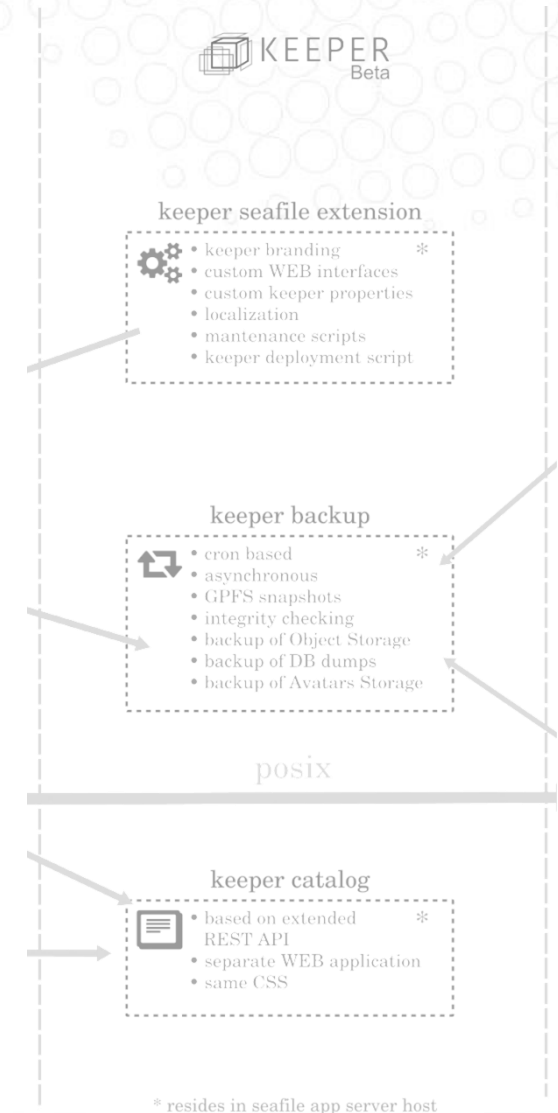


© 2016 Seafile
About Seafile



KEEPER ext, development

- PROD/QA/DEV environment, QA is very close to PROD (same for cluster)
- Build script, written in bash
- Release management in github + ZenHub
- [Mattermost](#) for team communication and notifications
- Releases are tested by MPDL QA group



KEEPER, Seafile core changes

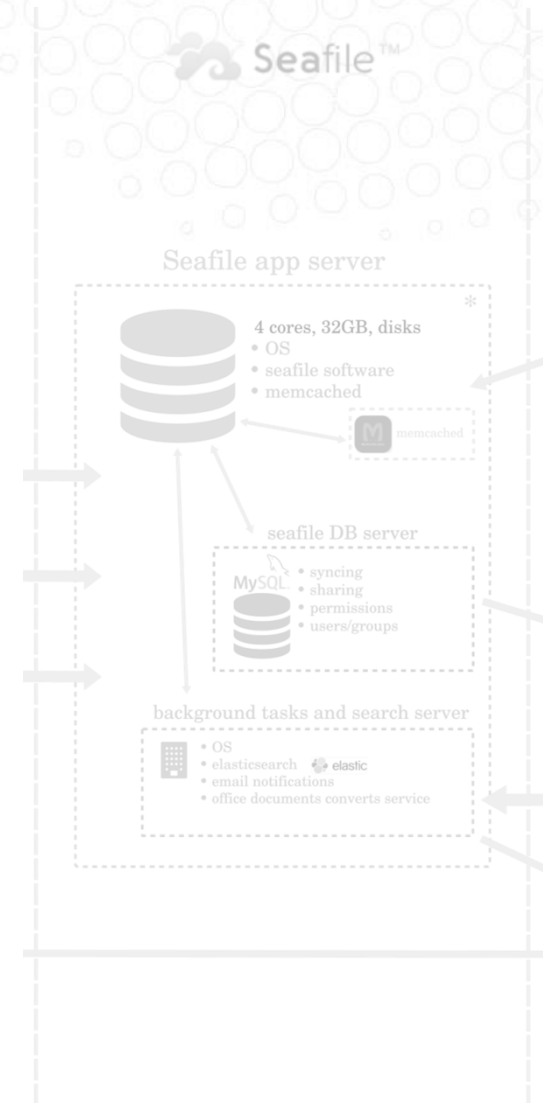
Snapshot Labels

The feature is used to mark certain library stands.

Depending on these labels, further functionalities can be offered, like search, browse, link to snapshots

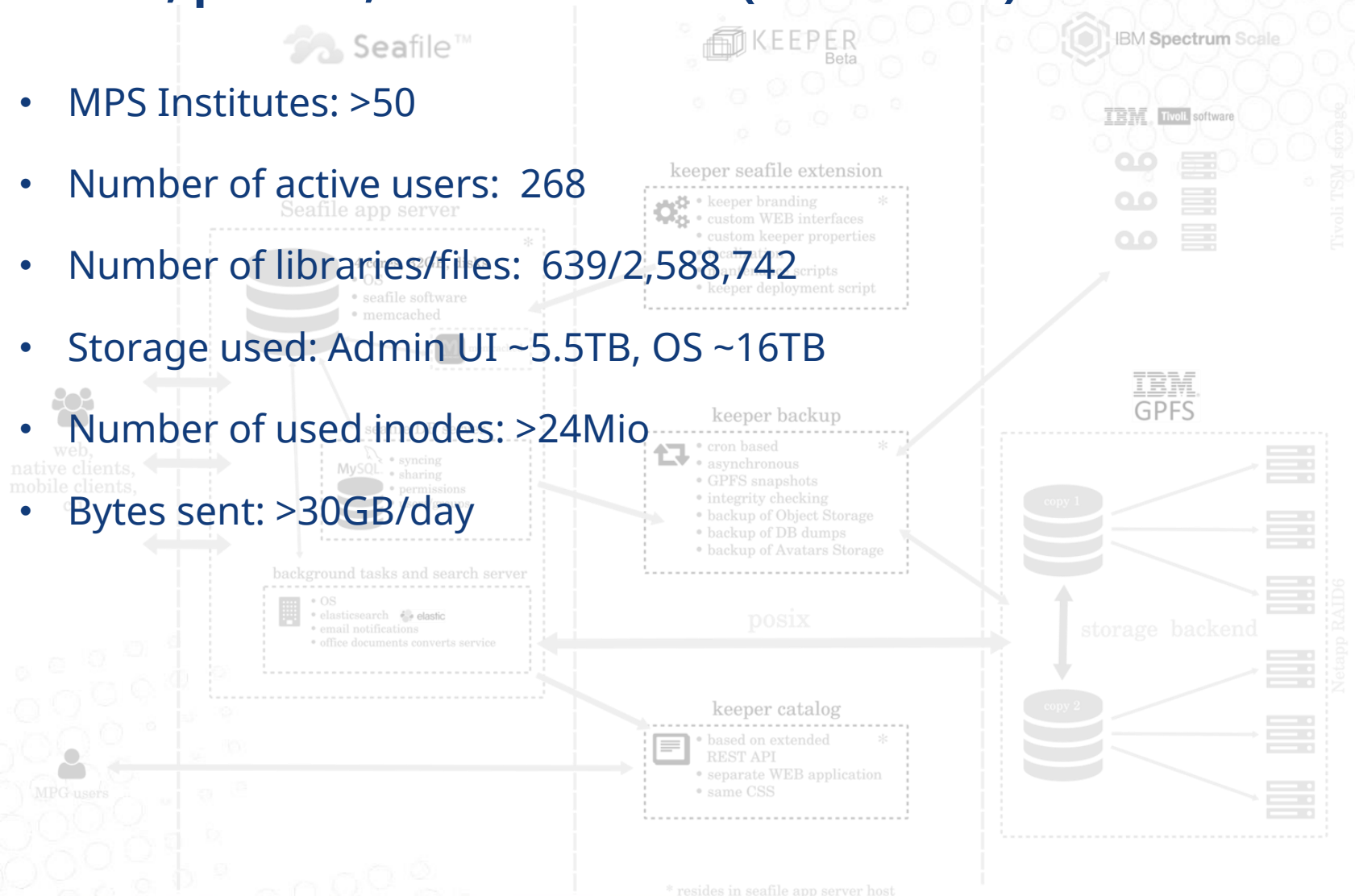
The Seafile REST API is respectively extended.

The feature is presented since seafile v6.2.x

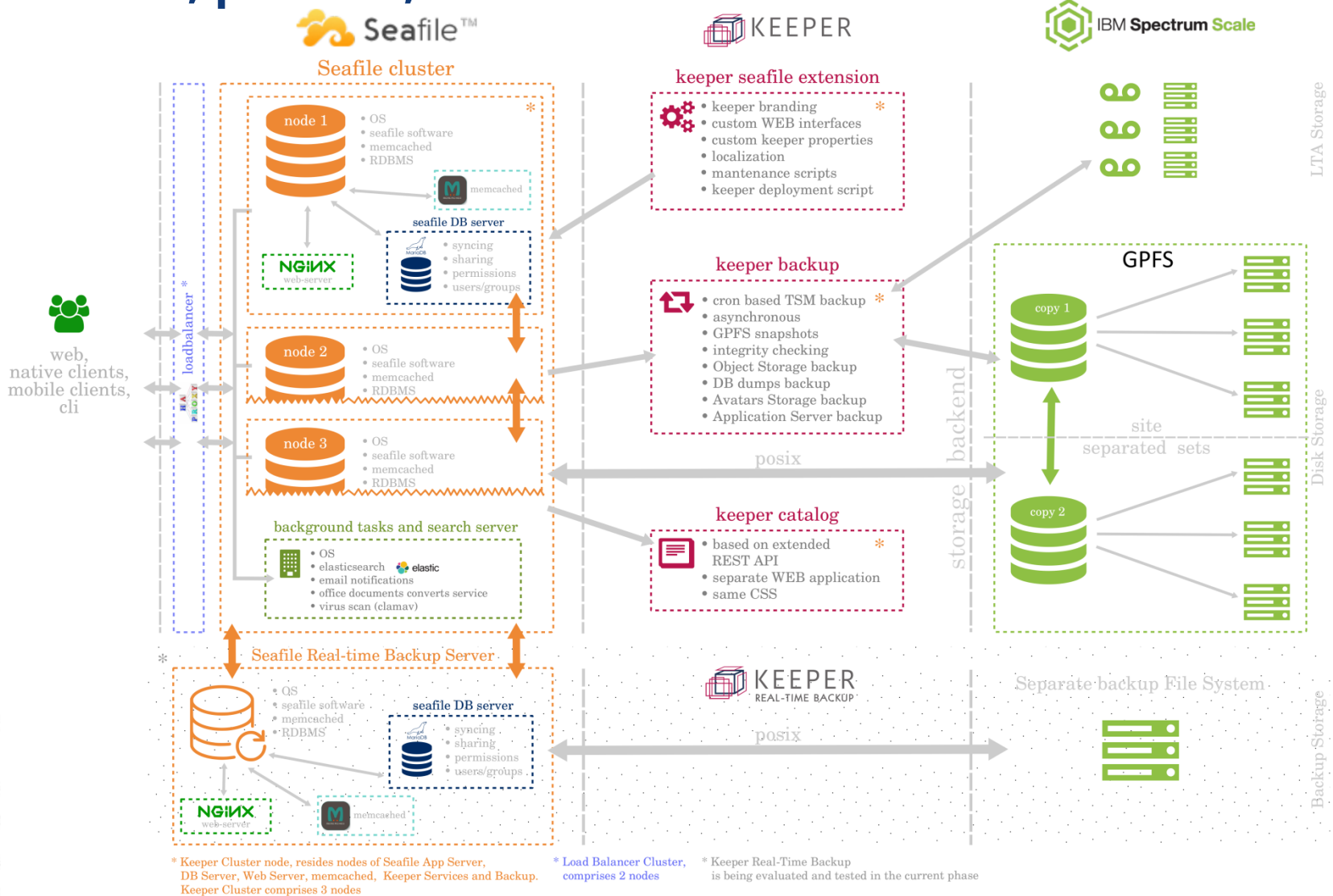


KEEPER, phase I, current status (26-01-2018)

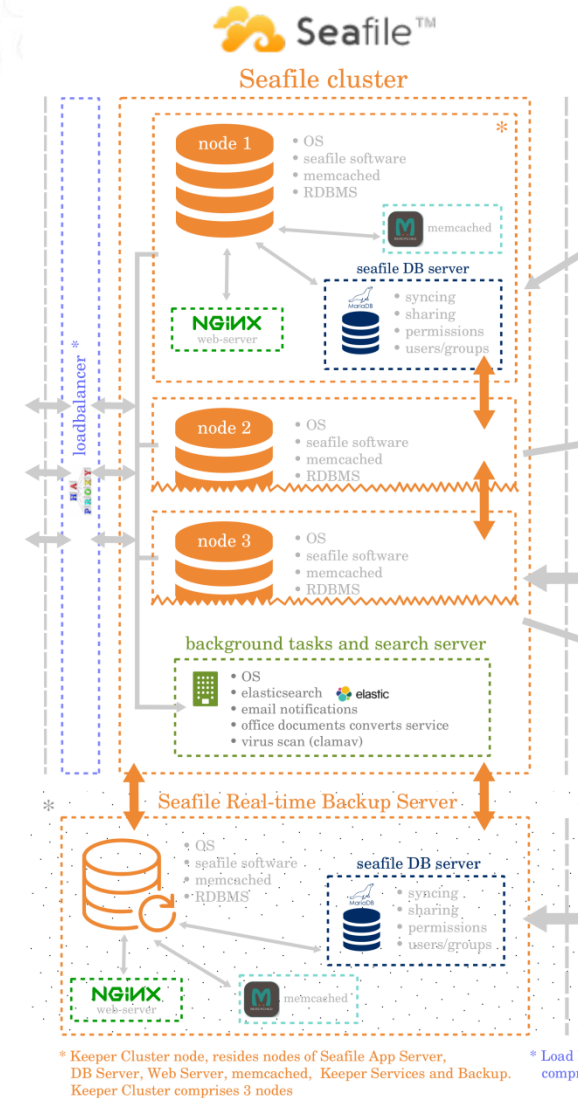
- MPS Institutes: >50
- Number of active users: 268
- Number of libraries/files: 639/2,588,742
- Storage used: Admin UI ~5.5TB, OS ~16TB
- Number of used inodes: >24Mio
- Bytes sent: >30GB/day



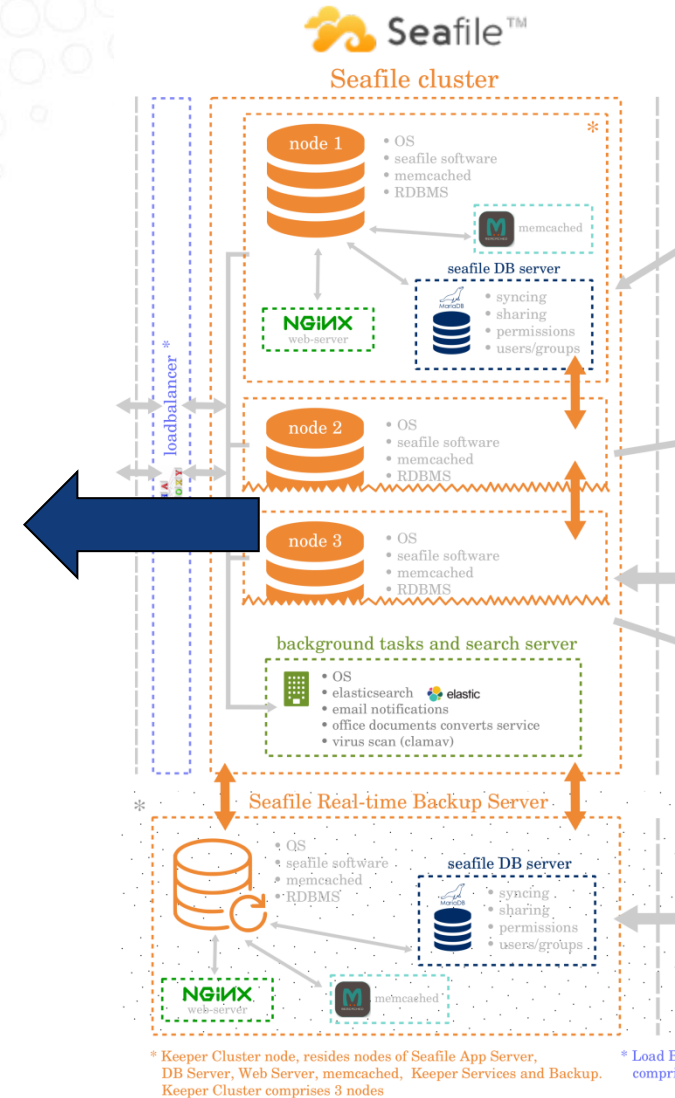
KEEPER, phase II, cluster



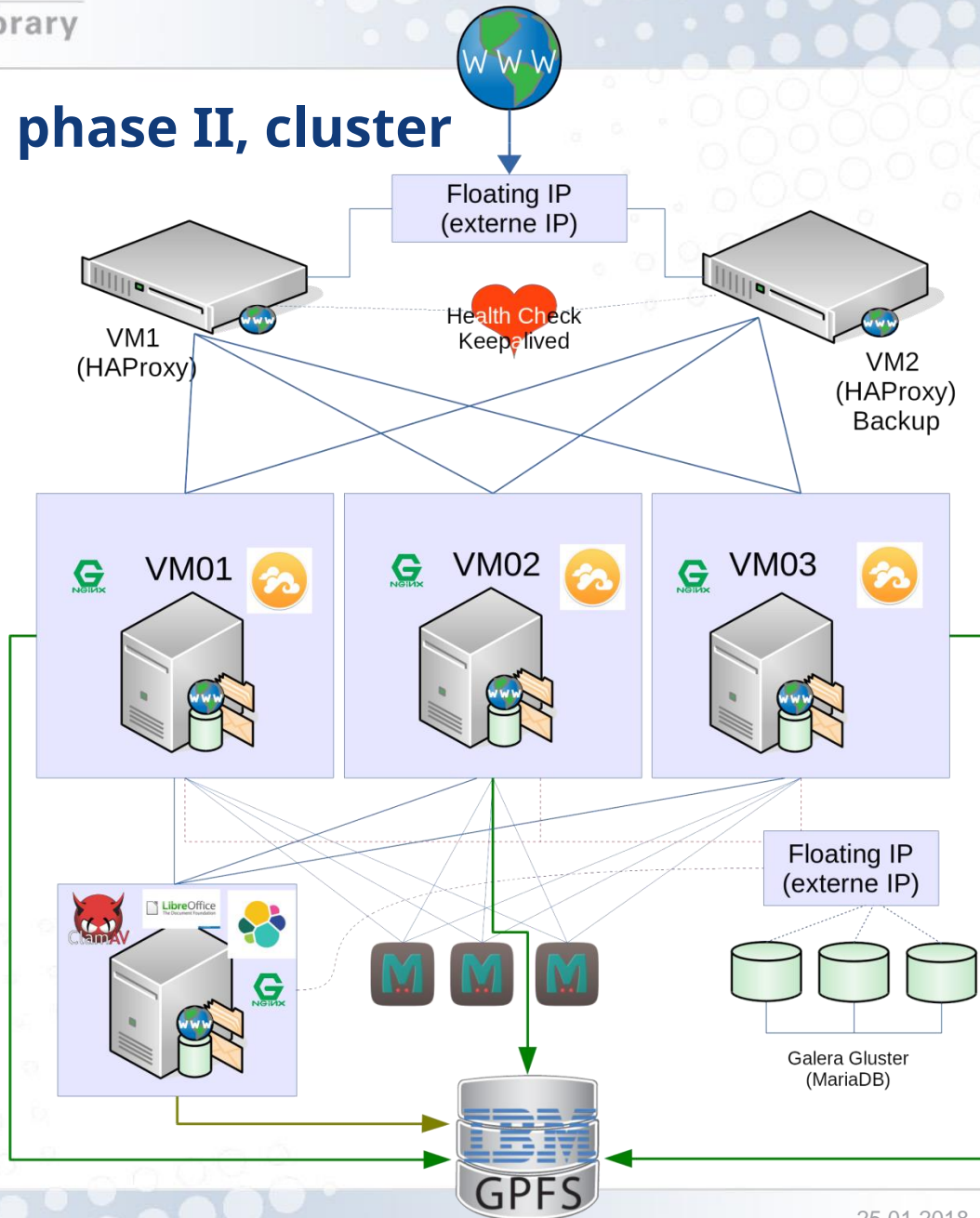
KEEPER, phase II, cluster



KEEPER, phase II, cluster



KEEPER, phase II, cluster



KEEPER, phase II, cluster

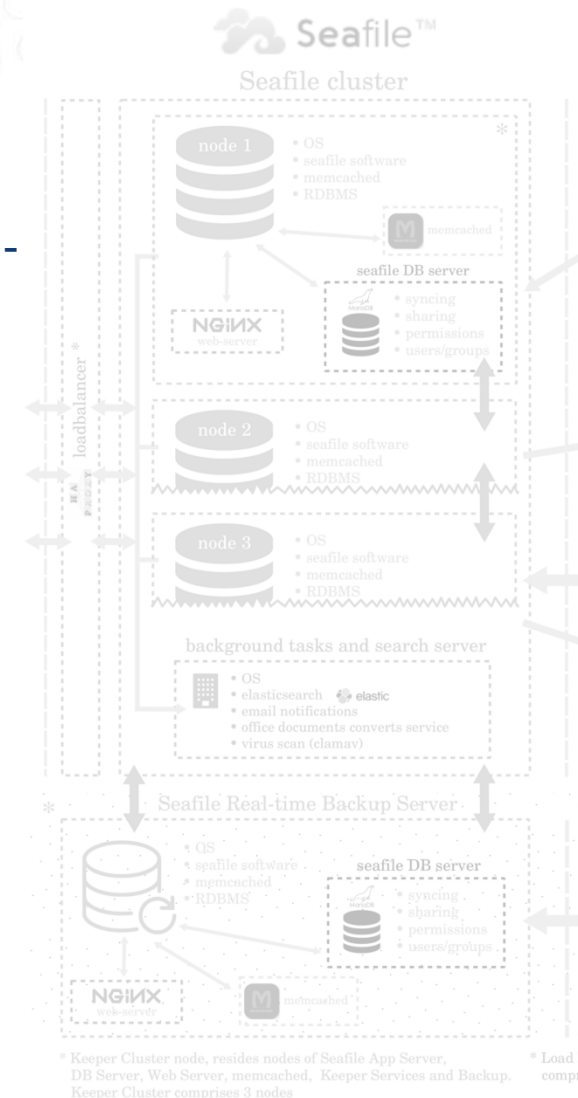
Hardware:

- 2 x Dell PowerVault MD3860f, 16G Fibre Channel, 4U-60 drive dense array, 42x8TB 7.2K RPM NLSAS 512e 3.5in Hot-plug Hard Drive, PI

Total: **672TB**

- 3 x Dell PowerEdge R730, 2 x Intel Xeon E5-2630 v4 (2.2GHz,10C/20T, 25MB Cache, 8GT/s QPI, 85W, Turbo, HT)

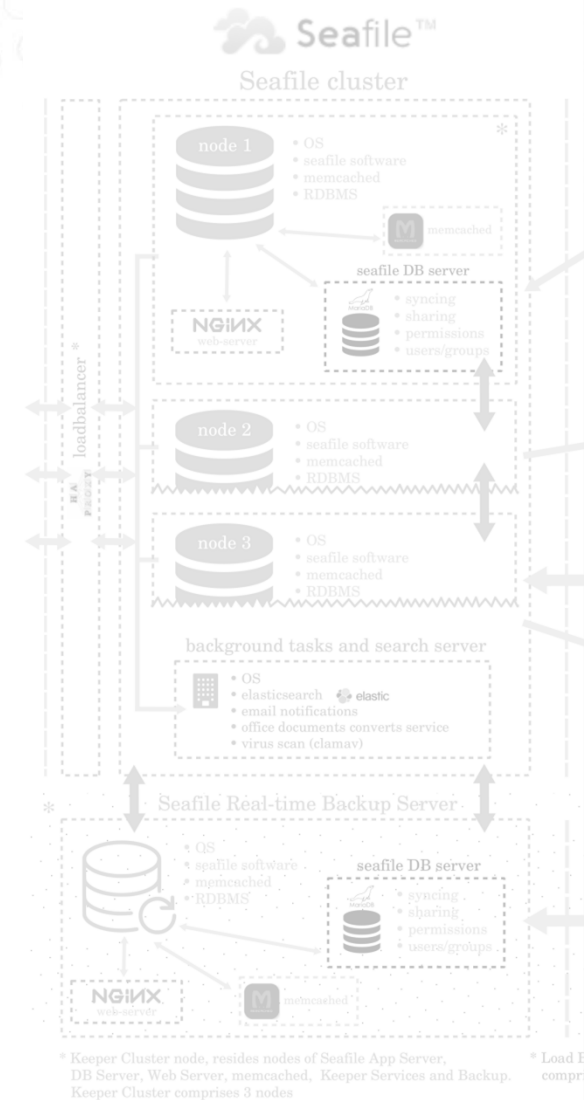
Total: **84 CPUs/768 GB RAM**



KEEPER, phase II, cluster

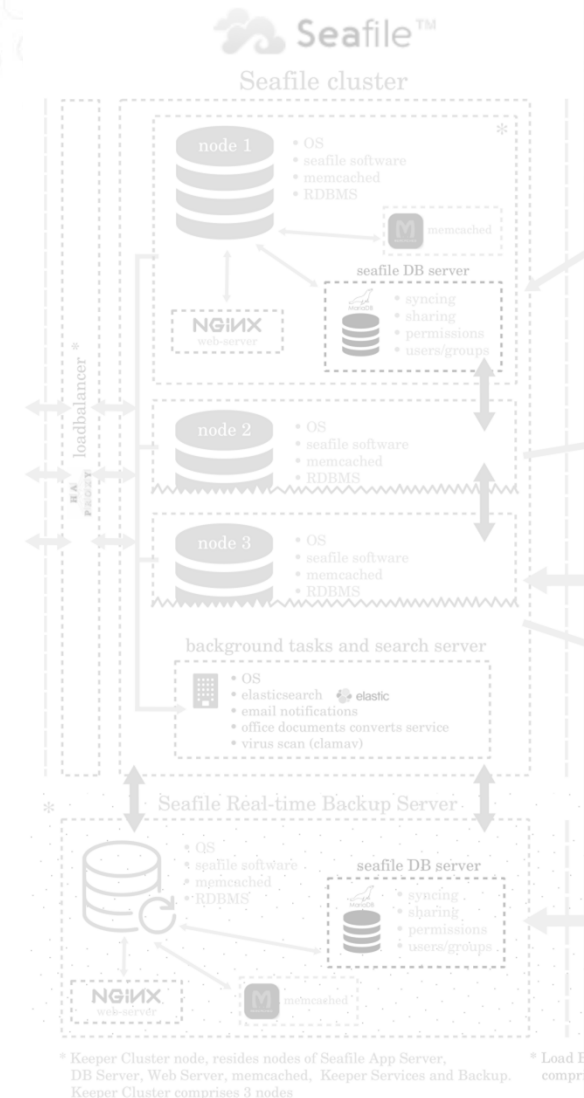
Cluster Management:

- HA Proxy + keepalived
- Puppet Server, management of nodes VMs
- Rsyslog Server
- Icinga, monitoring
- Kibana, log visualizations



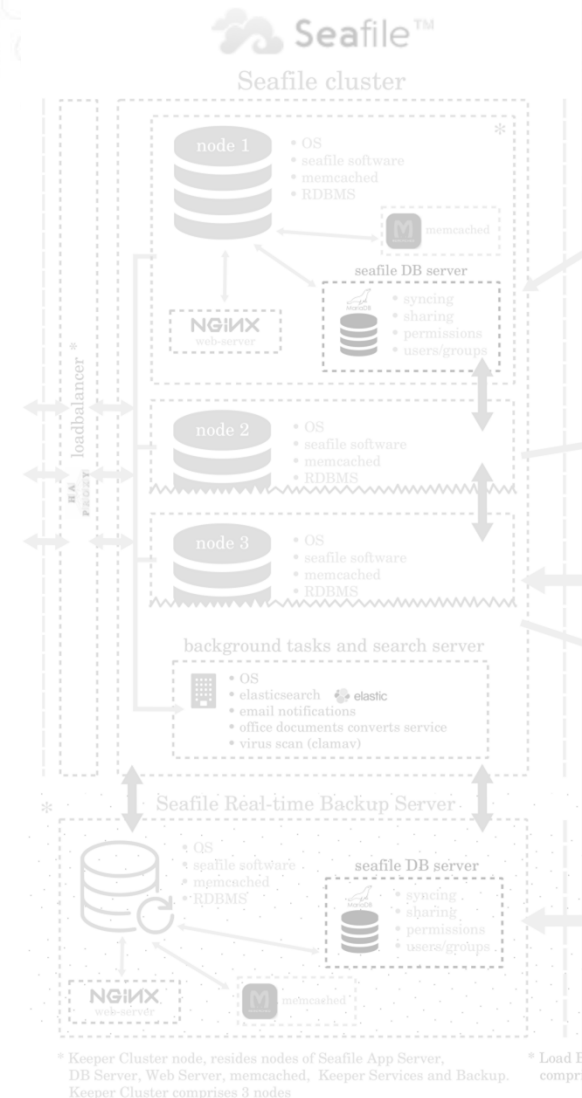
KEEPER, phase II, migration

- should be done in two steps (rsync + inc rsync)
- rsync should be parallelized
- considerable downtime is planned (a weekend)
- tested on QA infrastructure



KEEPER, next steps

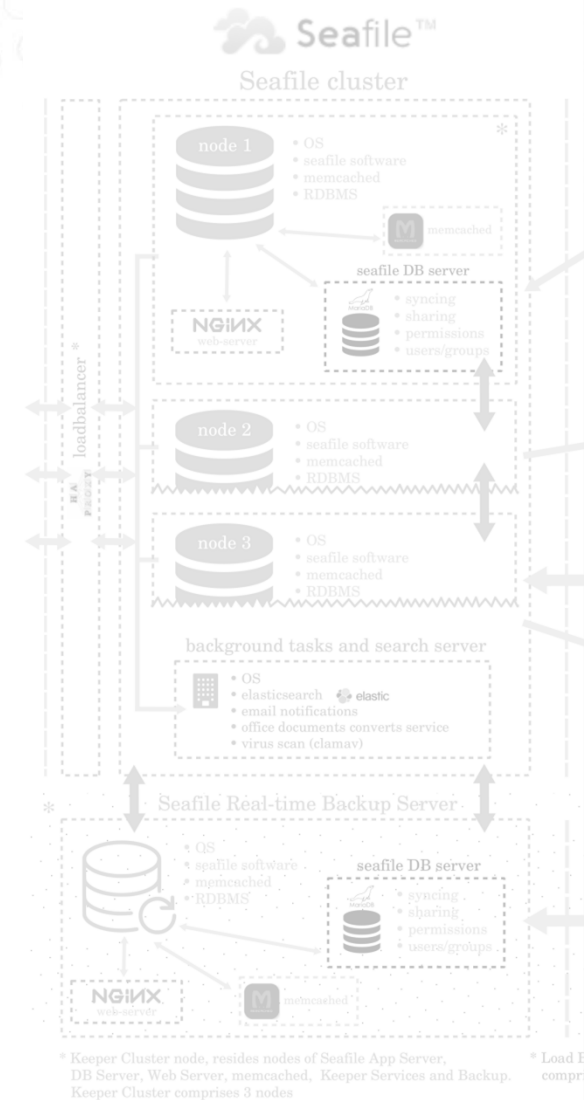
- Migration of MPIs Seafile instances into KEEPER
- Login with Shibboleth
- Real-Time Backup server
- OnlyOffice/Collabora Online
- Integration with MPG Services (like MPG SharePoint)
- Integration with external services like [Bernstein](#) (certification on the blockchain)
- Integration in internal workflows of MPIs (local drives with SeaDrive)
- Regular recovery tests on QA



KEEPER, next steps

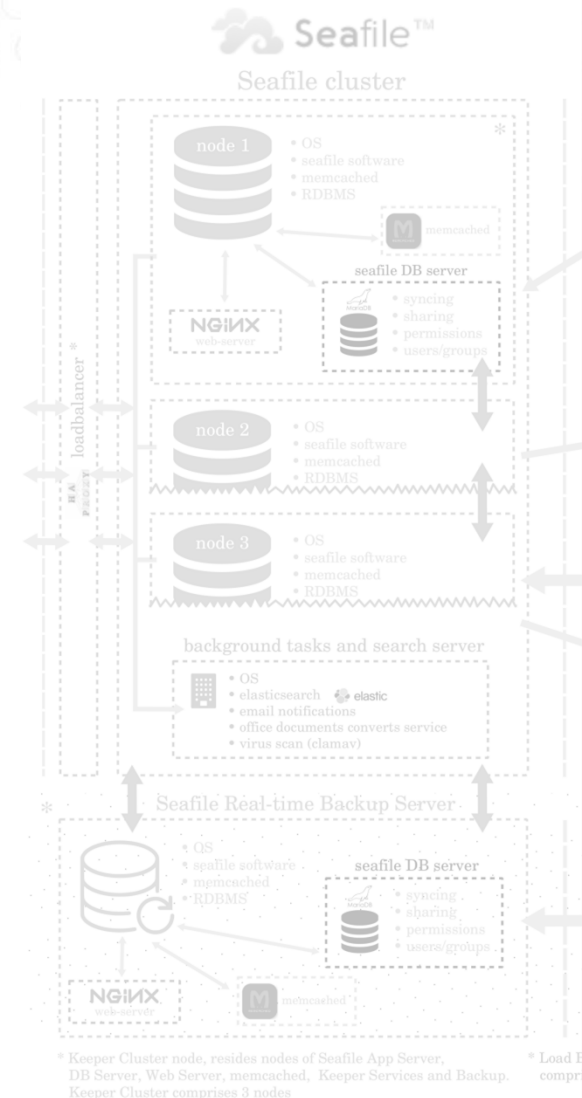
New KEEPER ext features

- Enhanced Project Catalogue
- Project Landing Pages
- Archive Metadata Form
- build script refactoring



KEEPER, timelines

- Project started: beginning of 2016
- Phase I: from October 2016 till now
- Phase II finalization and Max Planck wide roll out: beginning of 2018





THANK YOU!