HU Box
Cloud Storage Deployment and Operation at Humboldt-Universität zu Berlin

Karsten Asshauer, Enno Gröper

31 January 2017
Overview

- Statistics
- Our Deployment
- Join efforts!
Statistics!

![Graph showing space usage in HU Box from Sep 2015 to Jan 2017.](image)

- **Object count in thousands**
- **Used space in GiB**

![Graph showing users in HU Box from Sep 2015 to Jan 2017.](image)

- **Users**
- **Groups**
- **Libraries**
Seafile at Humboldt-Universität zu Berlin

- Seafile Professional (currently 6.x)
- Running on OpenStack with CEPH as storage backend
- Using single central database server
- Shibboleth for Authentication
  - Local users
  - Guests (DFN-AAI, eduGAIN) (in progress)
- Generic local guest accounts using invitation feature of Seafile (planned)
User management

- All users have to accept current terms and conditions (currently self-made solution)
- Shibboleth affiliation can be mapped to Seafile roles
- Apply guest role to all external users by default
- Including Shibboleth Embedded Discovery Service to select identity provider
- Clients use out-of-band authentication via tokens
  - By using authentication tokens you bypass central authentication servers
  - Token lifetime has to be limited
Cluster setup

Legend:
- Blue: network traffic
- Green: memcached traffic
- Purple: database traffic
- Black: storage traffic

Diagram:
- Loadbalancer HA pair
- Active node
  - memcached
  - Webserver
  - Shib-proxy
  - Fileserver
  - Seahub
- Node 2
  - Database
- Node n
  - Background node
    - File indexing
    - File preview
    - Email
- Ceph storage backend
Concerns?

- Backup and restore
  - Replication of CEPH objects into filesystem backend representation of Seafile (objects as files) in our SAN
  - Backup file representation to TSM (tape)
  - If ever needed: Restore from tape will take a *long time*

- Memcached high availability (HA)
  - Memcached cluster mainly to improve memory usage / share load, not for HA
  - Memcached cluster logic located in the different client implementations
  - Consistency issues could arise → active/passive using pacemaker?

- Deletion of user data: a users personal data vs. collaborative data (shared libraries)
Join efforts!

Goals:

- Join efforts
- Be flexible and extensible
- Get simple way to deploy production cluster (on Ubuntu)
Further collaboration

- seafile-python-api-scripts
  - Trying to establish collection of helpful scripts using python API of Seafile
- seafile-ceph-tools
  - Tools for ceph backend, like replication CEPH → filesystem or CEPH usage accounting
- Seafile Educational User Community (since 2015)
  - Meetings of academic users and Seafile developers (Seafile Ltd.)
  - Seafile Educational Community Forum (HU Moodle)
  - Next Seafile Workshop: September 2017, Université de Strasbourg, France
Thank you for your attention!

Karsten Asshauer - karsten.asshauer@cms.hu-berlin.de
Enno Gröper - enno.groeper@cms.hu-berlin.de

Feel free to contact us for further questions!