The cloud storage service bwSync&Share at KIT

*Alexander Yasnogor, Nico Schlitter, Andreas Petzold*

@CERN, Workshop on Cloud Services for File Synchronisation and Sharing
The bwSync&Share is a sync and share solution installed and operated at KIT.

- It is a on-premises alternative to existing public cloud services.
- Available for students and scientists in the state of Baden-Württemberg, Germany.
- Authentication and authorization of users is based on SAML 2.
- The service is running on PowerFolder software.
  - Windows, Linux, MacOS, Android, iOS, WEB versions are available.
What we have done

- Evaluation of available software solutions (2012-2013)
  - Running multiple test installations to get user feedback and operations experience
  - Candidates: Druva InSync, Grau Data, OwnCloud, PowerFolder, TeamDrive

- Decision criteria:
  - Software functions (shibboleth…)
  - Operation factors (scalibility, sustainability, support)
  - User feedback
  - Experience from the test installation
  - Operation, license, support costs

- Start as official service at 01.01.2014 using PowerFolder 9
Service architecture

**Web Layer**
- F5 BIG LTM8400 Load balancer
- Apache+Shibboleth SP

**Application Layer**
- Apache+Shibboleth SP
- PowerFolder Node 1
- PowerFolder Node 2
- PowerFolder Node N

**Data Layer**
- MySQL Galera cluster
- LSDF Storage

**Other Components**
- SAML A&A
- Mobile access
- WEB access
- Applicaiton access
- KIT Firewall
- Peer to Peer Sync.
Service architecture

- **Web layer:**
  - 1 failover/load balancer F5 BIG LTM8400
  - 2x online, 1x standby
    - VM: 4GB RAM, 4 cores Xeon CPU X5650@2.67GHz
    - OS: Debian, SW: Apache(proxy balancer,ajp), Shibboleth SP

- **Application layer:**
  - 3x online
    - VM: 24GB RAM, 4 cores Xeon CPU X5650@2.67GHz
    - OS: Debian, SW: PowerFolder server v. 9

- **Data layer**
  - MySQL Galera cluster
    - 3x online
  - LSDF Storage
  - IBM SONAS
Authentication & Authorization

- Defined two kinds of accounts:
  - Members of universities/colleges (shibboleth)
  - Invited external people (external)

- Ways to create an account:
  - Invitation to a folder by an existing user (shibboleth, external)
  - Self registration (only shibboleth)

- Authentication & Authorization via Shibboleth (DFN-AAI)
  - Initial registration always via Web registration of bwIDM
  - AA by WebSSO in Web-interface
  - AA by ECP profile in desktop/mobile application

- Authentication & Authorization via Database
  - Initial registration always via Web registration of bwIDM
  - AA by local email/password
Service usage (Accounts)

- 9123 accounts (12.11.14)

- Potential:
  - 450 000 eligible users
    - 9 universities
    - >50 colleges
  - 55 000 expected users
Service usage (Logins)

- 1153 unique logins per day (6.11.14)
- 2624 unique logins in October
Service usage (Quotas)

- Default shibboleth user quota: 10 GB
  - Can be extended by user request
- Default external user quota: 0 GB

Quota usage statistics

- <1MB: 34.84%
- 1MB-10MB: 10.37%
- 10MB-100MB: 15.13%
- 100MB-1G: 20.02%
- 1GB-10GB: 19.24%
- >10GB: 0.40%
Service usage (Files)

- ~4.8 TB total data (snapshots, metadata, archive)
- ~2.7 TB of users’ data
- ~3.7 M of users’ files

99.86% files < 58.75MB
Outlook

- Feature integration
  - Account merging
  - Online editing
  - Download center improvements by adding Linux client repository

- Enlargement of the service usage
  - Service advertising
  - Integration of remaining colleges in BW
  - Integration of new educational institutions through the new DFN Cloud initiative across Germany

- Thinking about
  - End-to-End user data encryption
## Service summary

<table>
<thead>
<tr>
<th>Category</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Status:</td>
<td>Prod</td>
</tr>
<tr>
<td>Number of users (current, target):</td>
<td>9123, 55 000</td>
</tr>
<tr>
<td>Default and Maximum quota:</td>
<td>Default:10GB, Maximum: 100GB</td>
</tr>
<tr>
<td>Linux/Mac/Win user ratio:</td>
<td>Unknown</td>
</tr>
<tr>
<td>Desktop clients/Mobile Clients/Web access ratio:</td>
<td>Unknown</td>
</tr>
<tr>
<td>Technology:</td>
<td>PowerFolder, Galera DB, F5 solution, SAML</td>
</tr>
<tr>
<td>Target communities:</td>
<td>Baden-Württemberg students and scientists</td>
</tr>
<tr>
<td>Integration in your current environment (examples):</td>
<td>LSDF storage, bwIDM, DFN-AAI Federation, VMWare ESX, MySQL Cluster</td>
</tr>
<tr>
<td>Risk factors:</td>
<td>No single point of failure</td>
</tr>
<tr>
<td>Most important functionality:</td>
<td>Syncing, sharing</td>
</tr>
<tr>
<td>Missing functionality (if any):</td>
<td>online editing (coming soon)</td>
</tr>
</tbody>
</table>
User feedback

- What additional functionality would you like to see?
  - “…Sharing not only for users in the state, but also beyond, ie by entering any email address…”
  - “…Lock files during editing for other users…”
  - “Viewing Excel and Word in the browser”
  - “Anonymous links publishing and sharing folders”
  - “Full text search”
  - “Encryption”
Thanks for attention!
Questions?

Karlsruhe Institute of Technology
Alexander Yasnogor
Hermann-von-Helmholtz-Platz 1
76344 Eggenstein-Leopoldshafen, Germany

Email: alexander.yasnogor@kit.edu