File Syncing Technology Advancement in Seafile -- Drive Client and Real-time Backup Server

Johnathan Xu
CTO, Seafile Ltd.

What is Seafile?

VS

Seafile is a fast and reliable Open Source file sync & share solution

What can Seafile do?

- Comprehensive cloud storage features
- Fast and reliable file sync
- High performance, light weight
- No maintenance headache

Used by about 20 educational institutes world wide

THE UNIVERSITY OF BRITISH COLUMBIA
McMaster University
UNIVERSITATIS ISLANDIAE SIGILLUM
Greenland
Iceland
Canada
Hudson Bay
Northwestern Passages
Oslo
Stockholm
Helsinki
Turun yliopisto
University of Turku
Tallinn
Baltic Sea
Riga
North Sea
Edinburgh
Glasgow
United Kingdom
Isle of Man
Dublin
Ireland
Manchester
Liverpool
London
Copenhagen
Denmark
Lithuania
Vilnius
Amsterdam
Netherlands
Brussels
Belgium
Luxembourg
Germany
Leibniz Universität Hannover
PSNC
Poland
Warsaw
Prague
CAMPUS CONDORCET
Paris–Aubervilliers
Inria
UNIVERSITÉ DE STRASBOURG
Université d'Auvergne
France
Switzerland
Austria
Vienna
Slovakia
Budapest
Hungary
Universität Zürich
Zagreb
Romania

Seafile Drive Client

How do we use cloud storage today?

- File Syncing
  - Pros
    - Automatically sync files across devices
    - Access anywhere: web, mobile, desktop
    - Easy to share files
    - Access offline
  - Cons
    - 1TB cloud space vs. 256GB local SSD drive
    - Selective sync is not intuitive to setup
    - “Why do I need to replicate all those files?”

How do we use storage before?

- Windows Share/Samba/NFS
  - Pros
    - Intuitive: access remote files like local files
    - Access files on-demand, no need to replicate
    - Is this how “cloud storage” should look like?
  - Cons
    - Cannot access offline
    - No web and mobile access
    - Slow in WAN
    - (Almost) no way to share files

“The best of both worlds”?

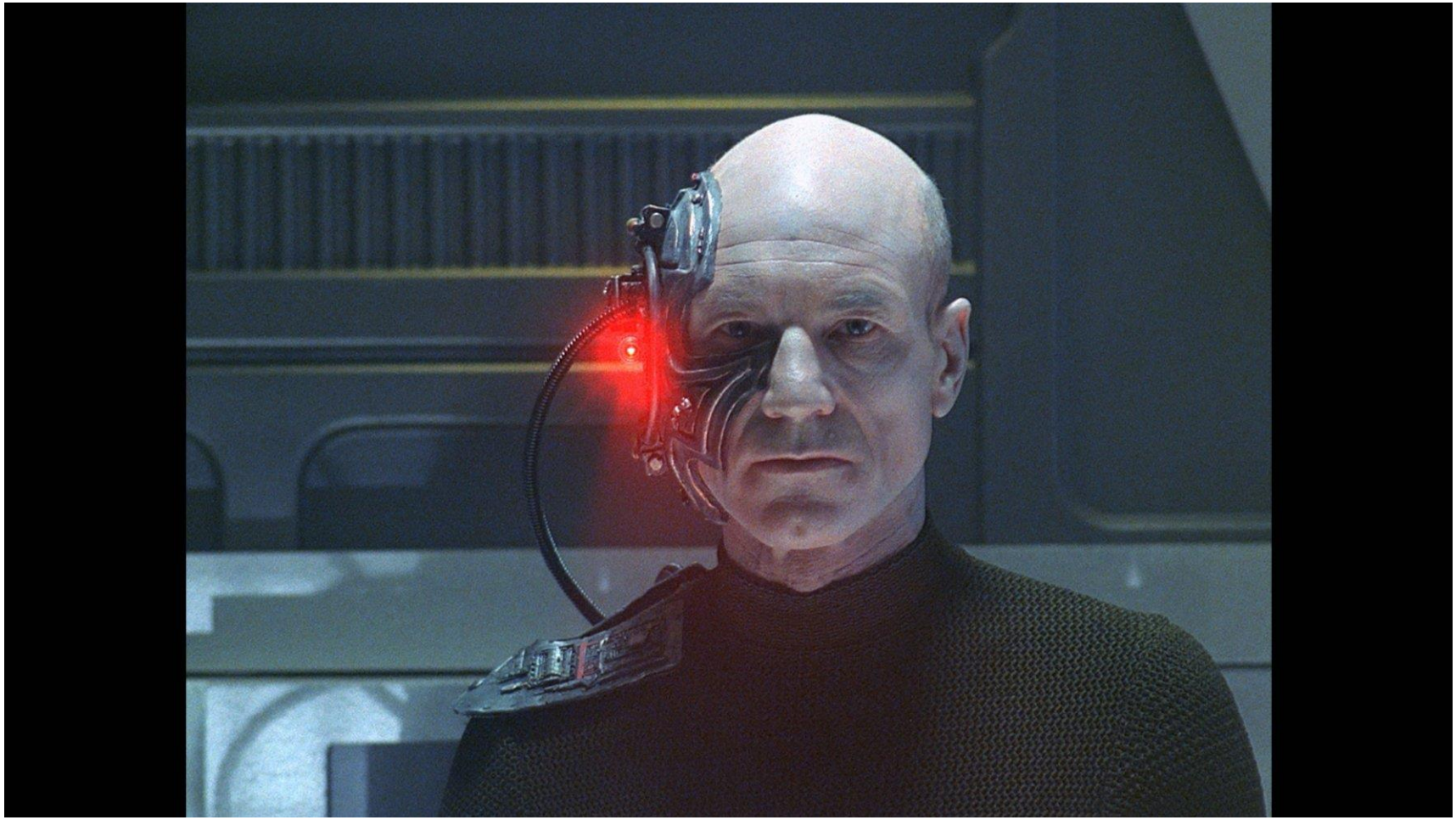
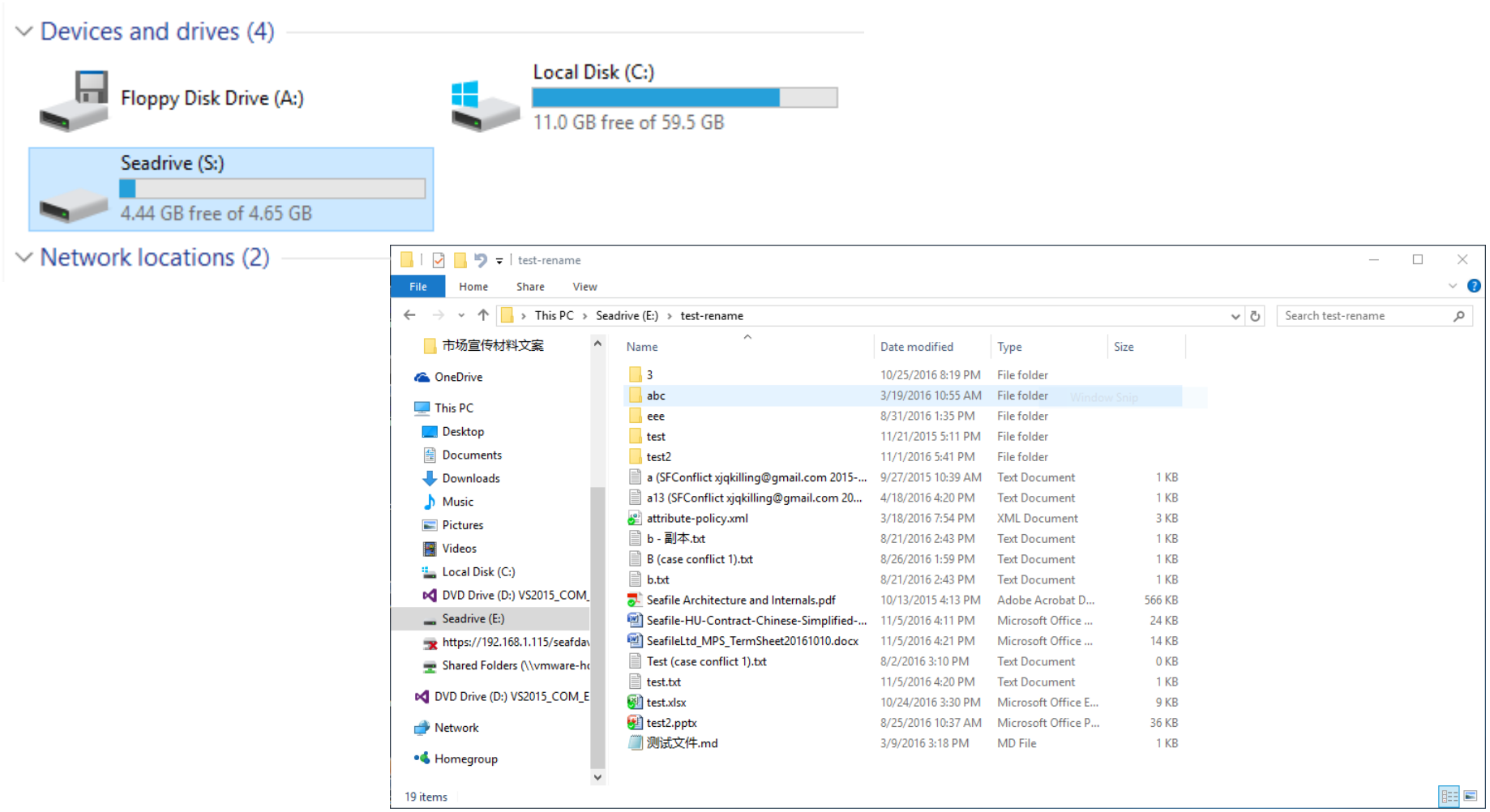


Photo copyright: Star Trek: The Next Generation

How about WebDAV?

- Map WebDAV as network drive
  - Pros
    - Can be used to map cloud storage as local drive
    - Most OS supports it out of box
  - Cons
    - Many redundant requests since the OS thinks it's local
    - Large file handling problems
    - Slow to upload many files
    - No offline access

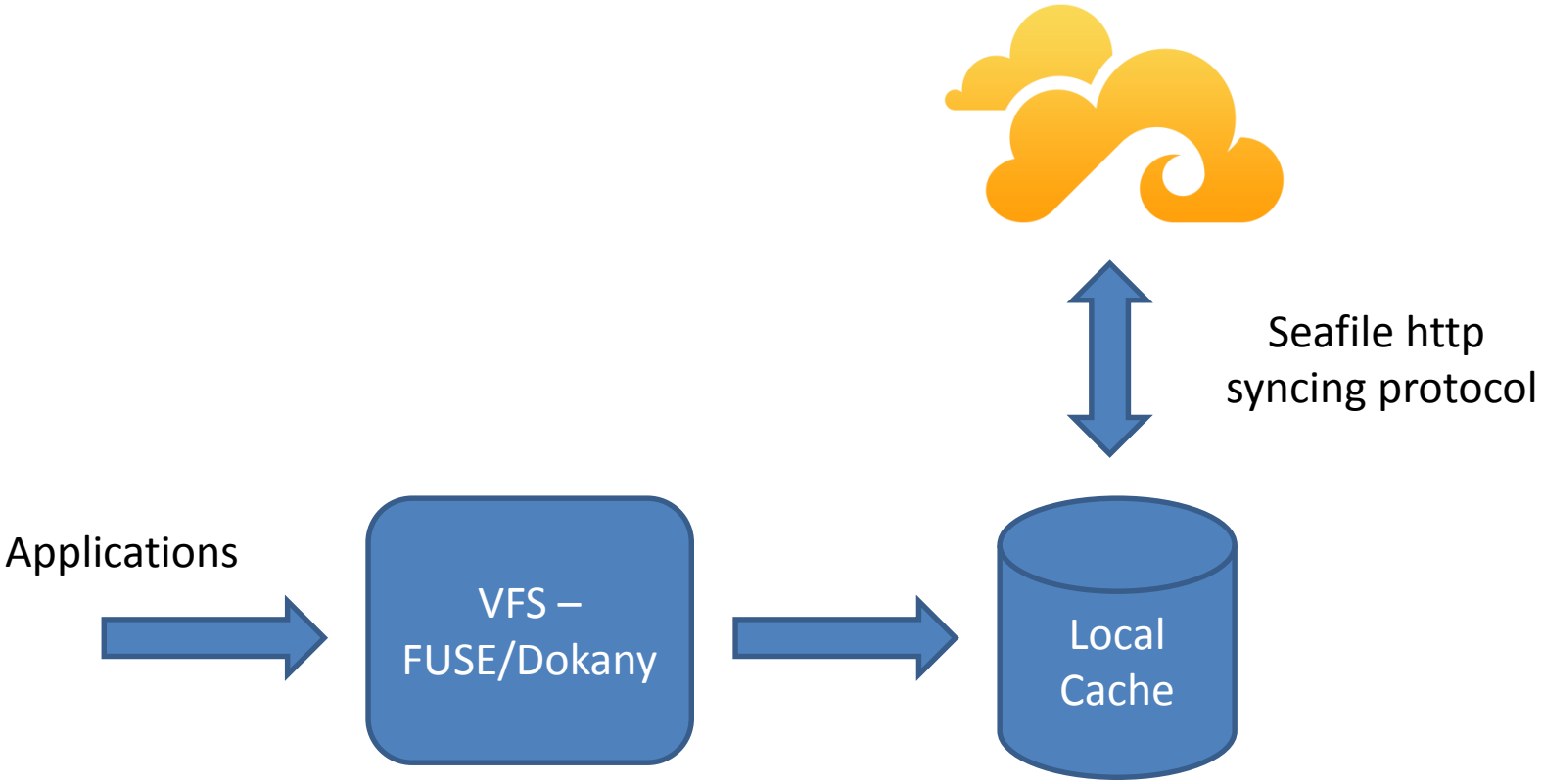
Seafile Drive Client



How it works

- Good from Windows Share
  - Map Seafile cloud storage as virtual drive. Browse files without syncing them.
  - Files are downloaded on demand. No unnecessary storage consumption.
  - Convenient and intuitive to use.
- Good from file syncing
  - File/folder list fully cached locally
  - Accessed files are cached locally for offline access
  - Manually cache folders/files for slow network or offline access (To be done). Replace “selective sync” function.
  - Can change files offline.
  - Easily share files

How it works


Seafile http
syncing protocol
Applications
VFS –
FUSE/Dokany
Local
Cache

Use Cases

- Satisfy 90% of file syncing use cases
- Replace Windows Share or Samba with Seafile.
- Access files on workstations with small local disk.
- Writing large scientific data directly to Seafile server.
- Better enforcement of data access regulations. Users access files from Seafile server exclusively. The locally cached files can be removed on logout.

Demo

Performance

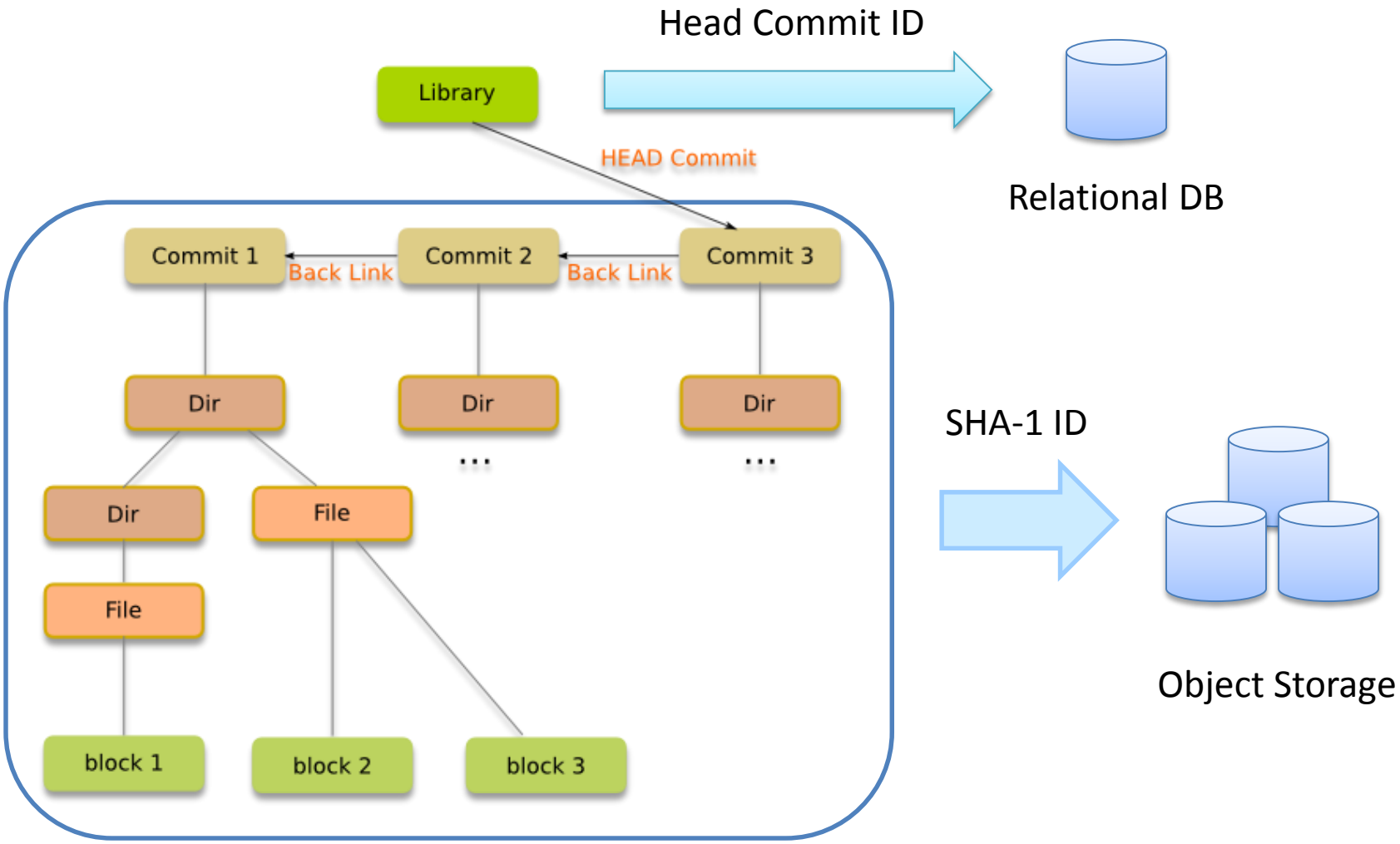
	SeaDrive	WebDAV	Windows Share
Copy out a 253MB file	6.6s	9.5s	5.4s
Read 10 positions in a 243MB file, read 1MB on each position	0.69s	9.8s	0.45s
Copy in a 253MB file	2.2s copy + 4s upload	11.6s	5.4s
Copy in a folder with 699 files, total 9MB	5s copy + 2s upload	8m9s	5s
Copy out a folder with 293 files, total 38MB	31s	36s	2s

Seafile Real-time Backup Server

Regular vs. Real-time Backup

- Regular backup
  - Backup the database and files from one server to another in fixed periods
  - Backup window: data written between backups are lost
  - DB and files may be inconsistent with each other
- Real-time backup
  - Smaller backup window: continuous, incremental backup from primary to backup server
  - DB and files are always consistent

Seafile File System Design



Data model similar to Git

How it works

- Sync libraries from primary server to backup server, with Seafile's syncing algorithm
  - Complete history of library is synced
  - Head commit info in the database is synced too
  - Auto detect updates on the primary server, nearly real-time
- All other information in the database is synced with MySQL replication
  - Sharing, ownership, etc.

Use Cases

- Continuous protection of Seafile data
- Multi-site replication
- High availability with geo-replicated hot backup

Thank You!