Moose FS

Ralph Krimmel

4. März 2014

Verteilte Dateisysteme

Viele bekannte Namen

- ► Ceph
- ► Lustre
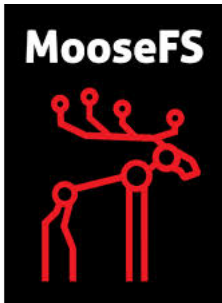- ► Glusterfs
- ► ...

Plattformen

Clients (aka mfsmount): Alle mit funktionierender FUSE Implementierung

- ► Linux (ab Version 2.6.14)
- ► FreeBSD
- ► OpenSolaris
- ► MacOS X

Nur Server, Metalogger und Chunkserver

- ► Solaris
- ► Windows mit Cygwin
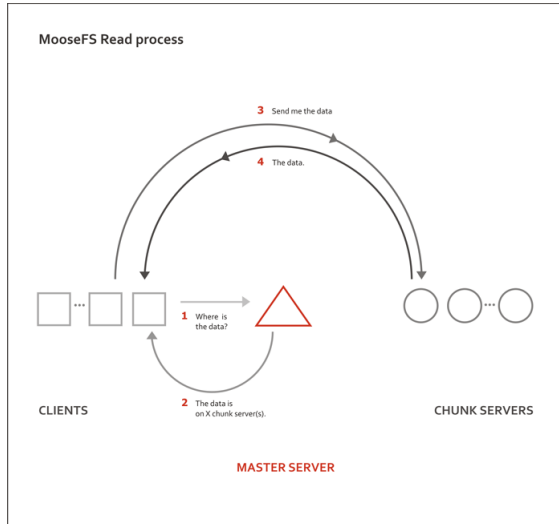
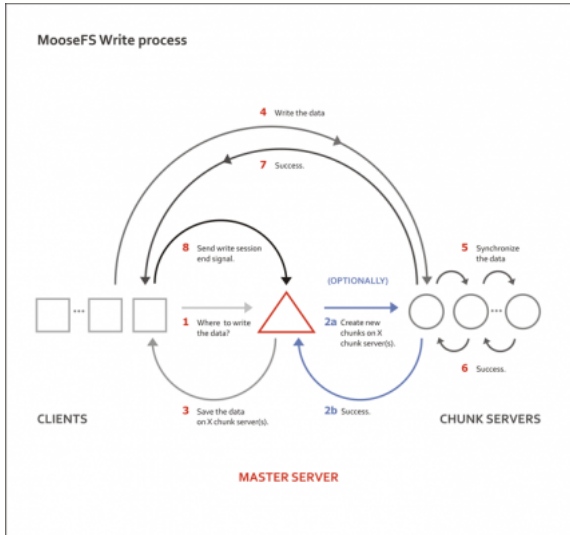# Eher unbekannt: Moosefs

Standard Features

- ▶ Hierarchisch (Baumstruktur)
- ▶ POSIX Datei Attribute:
    - ▶ POSIX Rechte
    - ▶ Last Access Time
    - ▶ Modification Time
- ▶ Spezialdateien
    - ▶ Pipes
    - ▶ Sockets
    - ▶ Blockdevices
- ▶ Symbolische und Hardlinks

Einleitung
OOOO

MFS Technik
●O

Unstructured data
OOOOOO

Questions?
O

Lesen

Schreiben

Definition

▶ Unstructured data refers to any data that has no identifiable structure.

Einleitung
oooo

MFS Technik
oo

Unstructured data
o●oooo

Questions?
o

But there is always a structure!

Einleitung
0000

MFS Technik
00

Unstructured data
000●000

Questions?
0

Yes, but..

- ▶ Each document itself may have a structure (e.g. HTML)
- ▶ But: Structure can be not helpful for the processing task
  In HTML: Tags just for layout, not semantics

Examples

Word processing documents

PDF files

E-mail messages

Blogs and Web pages

Video/Audio

...

Challenges

Most of today's data is unstructured (80%)

Challenges

- ▶ Access to textual data (different file formats)
- ▶ Language of the data
- ▶ Volume (See Big Data presentation)
- ▶ Security of the data
- ▶ Searchability (the content has to be understood)

Current approaches

- ▶ Manual tagging
- ▶ Dataminig/Text analytics
- ▶ Document management systems (e.g. Products from SAS, Provalis Research, Inxight, SPSS)

Einleitung
0000

MFS Technik
00

Unstructured data
000000

Questions?
●

Questions?

Questions?