LIZARDFS

Is the MooseFS distributed filesystem in your future?

Back Bay Large Installation System Administration
BBLISA 11th of September 2013
MIT E-51, Room 145

Peter aNeutrino
(LizardFS.org)

How many users over the world ?

- The last release 1.6.27 had more than **4000 unique downloads** during the first month after it was released.
- We have tracked **7000 downloads** during the last 3 years from **sourceforge.net.**
- We are Open Source, so we do not know the exact number of users.
- TOP5: China, USA, Poland, France, Russia

LIZARDFS

What are the use cases ?

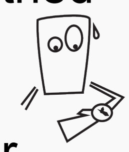
- backup,
- home directories,
- virtualization (with loopback block devices),
- media streaming,
- microscope images,
- computation.

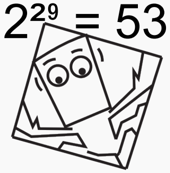
LIZARDFS

Why **don't** people use MooseFS?

- They don't know it exists.
- They read Jeff Darcy's blog:

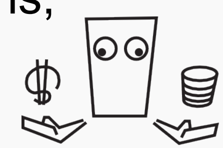
  http://hekafs.org/index.php/2012/11/trying-out-moosefs/
- They don't like rare releases.
- The community is very small.
- It was too slow for them (they tried lustre before).
- **They hate:**
  - SPOF on metadata server,
  - limit of files 2^{29} = 536 mln,



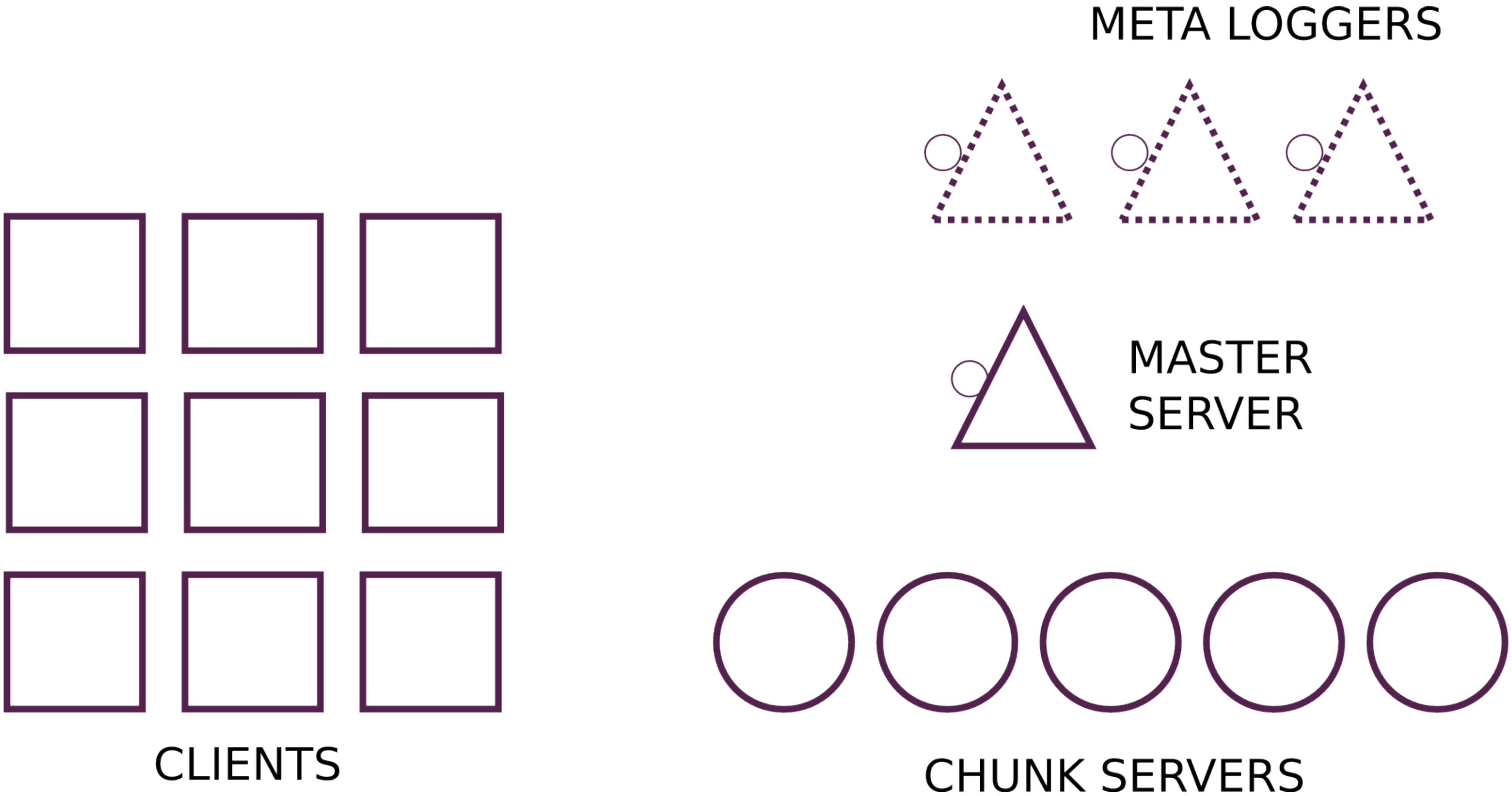


Why **do** people use MooseFS?

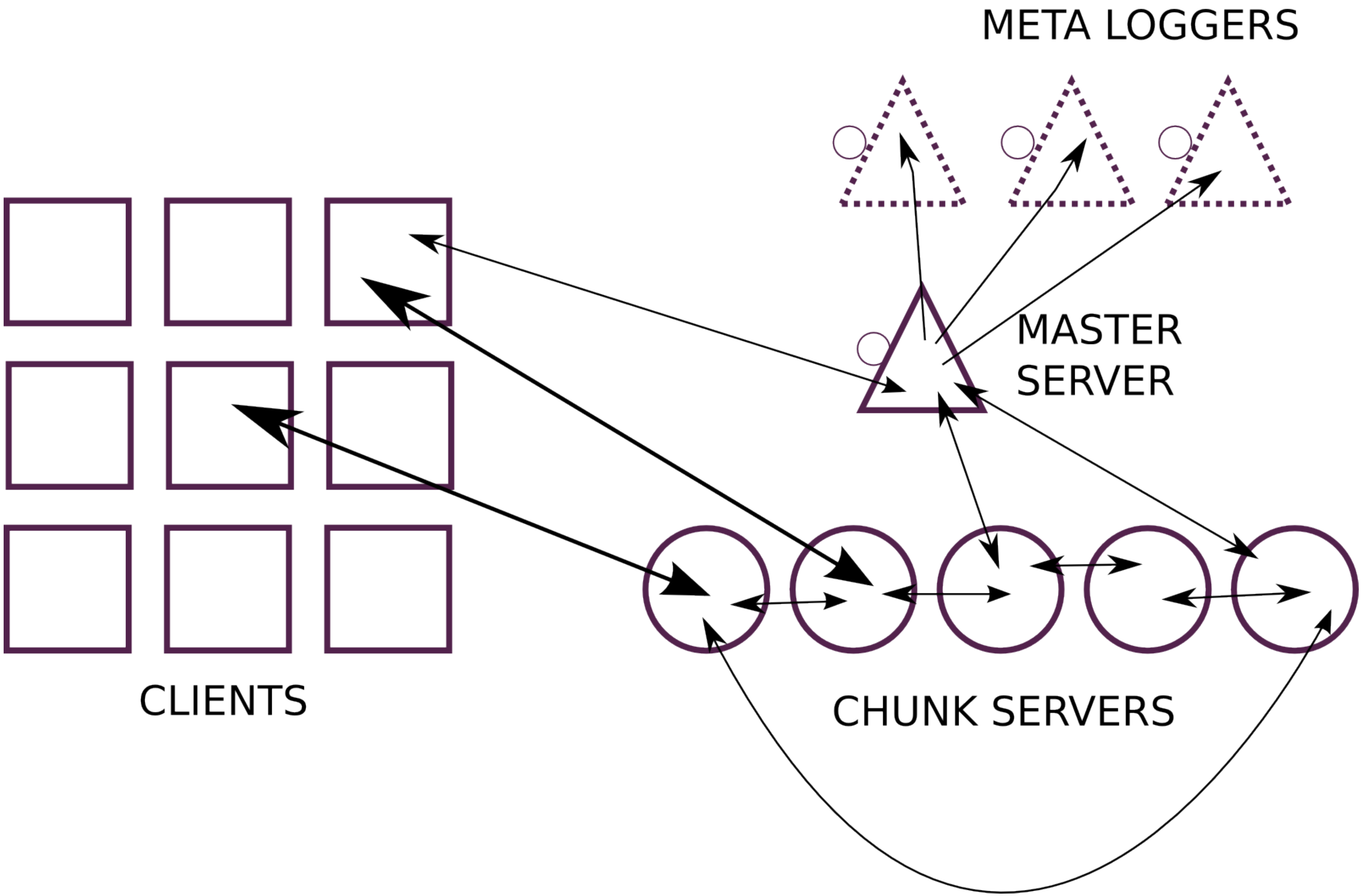
- They do know it exists.
- They saw only the chart on Jeff Darcy's blog...instead of read it.
- ...i am kidding, there are many 3rd parties blogs with positive results:
  - http://www.tinkergeek.com/?p=150
  - http://blog.opennebula.org/?p=1512
- **They love:**
  - fast & easy configuration,
  - snapshots,
  - fast replication,
  - build in trash,
  - speed of meta operations,
  - it is for free.



Components of the filesystem



Communication between components of the filesystem



Source code

```
ssh 192.168.122.5 -lroot

git clone https://github.com/lizardfs/lizardfs.git

cd lizardfs
```

Lines of code

```
cat $(find -name '*.h') $(find -name '*.cc') | wc -l
```

53459

Please note that this file system has:

- copy on write coherent snapshots,
- online replication of the data between nodes,
- build in transparent trash bin,
- CRC32 checksumming checked on both client and chunkserver sites,
- proactive checking of data integrity,
- auto-balancing of data,
- web based UI for monitoring,
- and was tested in production for 9 years.

Demo of installation (*.deb)

We install tools for building packages:

```
apt-get install dpkg-dev autotools-dev libfuse-dev pkg-config zlib1g-dev
```

Then we generate configuration:

```
time ./autogen.sh # 4 seconds
```

We build debs:

```
time dpkg-buildpackage # 40 seconds
```

...and we install them:

```
cd ..
dpkg -i mfs-common* mfs-master* mfs-cgi* mfs-cli* mfs-chunk*
```

We need to enable daemons to start:

```
sed -i -e 's/ENABLE=false/ENABLE=true/' /etc/default/mfs-*
```

Demo of installation (MASTER)

We copy default config file from template:

```
cp /etc/mfs/mfsexports.cfg.dist /etc/mfs/mfsexports.cfg
```

We start master daemon:

```
/etc/init.d/mfs-master start
```

At this point we could start using filesystem just **only metadata operations are available**. So we can `mkdir foodir` or `touch foo` or `ls`,

but we can't do: `echo foo > foo`

Lets run web based UI monitoring…

```
/etc/init.d/mfs-cgiserv start
```

… and go to: http://192.168.122.5:9425/mfs.cgi

LIZARDFS

Demo of installation (mountpoint)

We will play with **only metadata filesystem**, mounting it here:

```
mkdir /mnt/lizardfs
echo 192.168.122.5 mfsmaster >> /etc/hosts
mfsmount -H mfsmaster /mnt/lizardfs
```

Now we create some metadata:

```
cd /mnt/lizardfs
touch foo
mkdir foodir
ls
stat *
```

LIZARDFS

Demo of installation (chunkserver)

We need to have a directory (or mounted disk) where we will keep chunks:

```
mkdir -p /mnt/hd01 && chown mfs:mfs -R /mnt/hd01
echo /mnt/hd01 >> /etc/mfs/mfshdd.cfg
```

By default chunkserver daemon will try to connect to DNS name **mfsmaster**:

```
/etc/init.d/mfs-chunkserver start
```

Now we are able to create & read files with content:

```
cd /mnt/lizardfs
echo foo >> foo.txt
cat foo.txt
```

copy & paste >> bblisa_lizardfs.sh
run as root on your test server

LIZARDFS

time
bblisa_lizardfs.sh

real **0m59.240s**
user 0m32.840s
sys 0m2.684s

```
#! /bin/bash
MY_IP_ADDRESS=192.168.122.5 #must be NOT localhost NOR 127.*.*.*
if ip a | grep "inet $MY_IP_ADDRESS/" | egrep -v '127(.[0-255]){3}'
then echo ip ok ; else echo ERROR: wrong ip address. Please setup MY_IP_ADDRESS; exit 1; fi
git clone https://github.com/lizardfs/lizardfs.git
cd lizardfs
#We install tools for building packages:
apt-get install dpkg-dev autotools-dev libfuse-dev pkg-config zlib1g-dev
#Then we generate configuration:
./autogen.sh # 4 seconds
#We build debs:
dpkg-buildpackage # 40 seconds
#...and we install them:
cd ..
dpkg -i mfs-common* mfs-master* mfs-cgi* mfs-cli* mfs-chunk*
#We need to enable daemons to start:
sed -i -e 's/ENABLE=false/ENABLE=true/' /etc/default/mfs-*
#We copy default config file from template:
cp /etc/mfs/mfsexports.cfg.dist /etc/mfs/mfsexports.cfg
#We start master daemon:
/etc/init.d/mfs-master start
echo Waiting 10 seconds for master server first start...
sleep 10;
#At this point we could start using filesystem just only metadata operations are available.
#So we can mkdir foodir or touch foo or ls,
#but we can't do:  echo foo > foo
#Lets run monitoring…
/etc/init.d/mfs-cgiserv start
#…and go to: http://localhost:9425/mfs.cgi
#We will play with only metadata filesystem, mounting it here:
mkdir /mnt/lizardfs
echo $MY_IP_ADDRESS mfsmaster >> /etc/hosts
mfsmount -H mfsmaster /mnt/lizardfs
#Now we create some metadata:
cd /mnt/lizardfs
touch foo; mkdir foodir; ls; stat *
#We need to have a directory (or mounted disk) where we will keep chunks:
mkdir -p /mnt/hd01 && chown mfs:mfs -R /mnt/hd01
echo /mnt/hd01 >> /etc/mfs/mfshdd.cfg
#By default chunkserver daemon will try to connect to DNS name mfsmaster:
/etc/init.d/mfs-chunkserver start
#Now we are able to create & read files with content:
cd /mnt/lizardfs
echo foo >> foo.txt && cat foo.txt
```

LIZARDFS

Why LizardFS ?

Let’s go to http://lizardfs.org/

…and discuss it.

LIZARDFS

How make LizardFS the file system of your dreams ???

DISCUSSION…

LIZARDFS

THANK YOU SO MUCH :)

and please feel free to contact with me:

http://www.linkedin.com/in/aneutrino

aneutrino@lizardfs.org

Call me if you are in Warsaw

+48 602 302 132

We will go for a beer or/and talk next to the whiteboard!