After the hangover:
QuarkDB and the new namespace

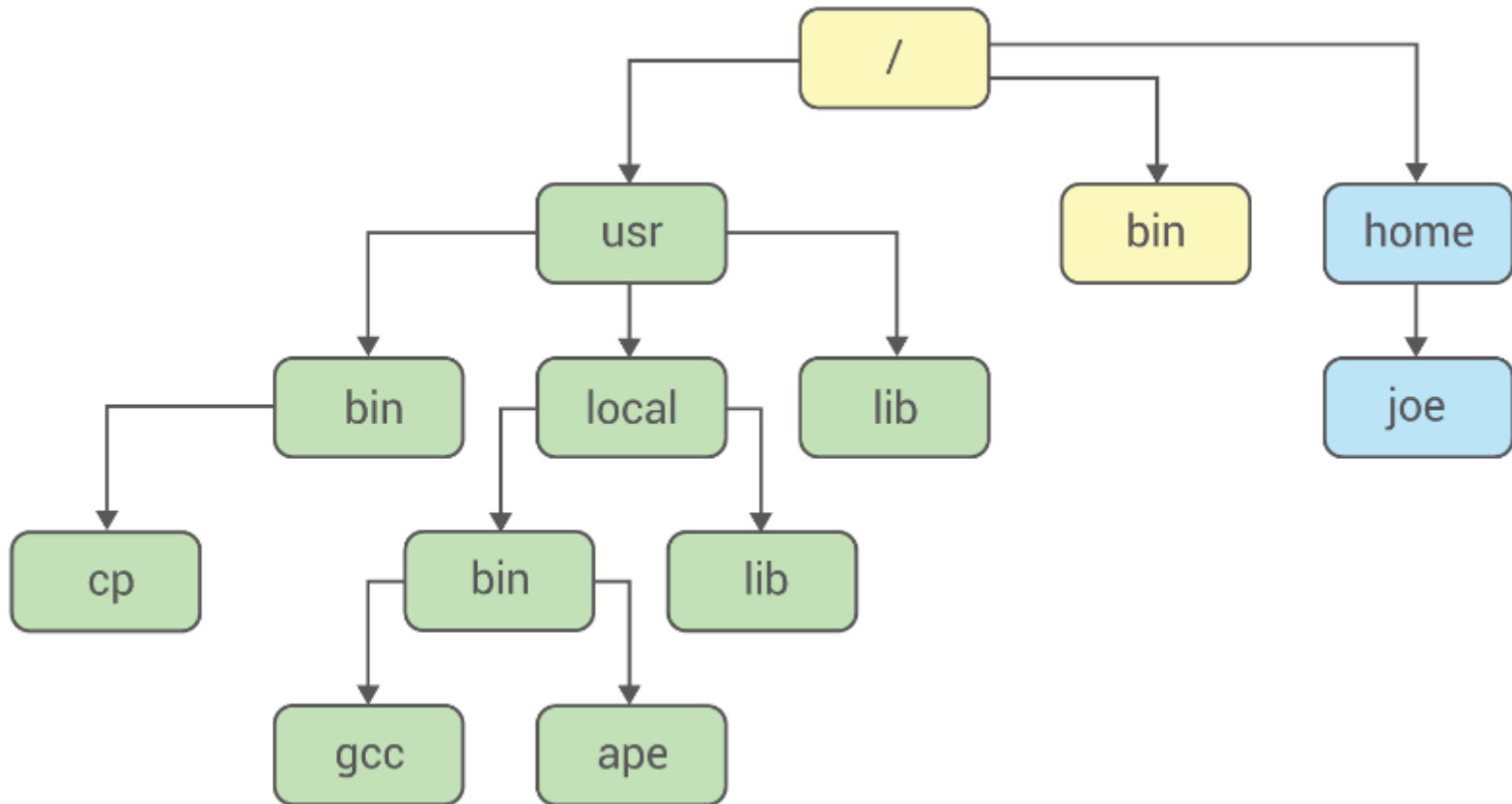Georgios Bitzes,
on behalf of the **EOS team**

The new namespace

- 2000+ **commits** spread across eos, QuarkDB, QClient repositories

- 483 individual **test scenarios**

- 2 load **testing tools**

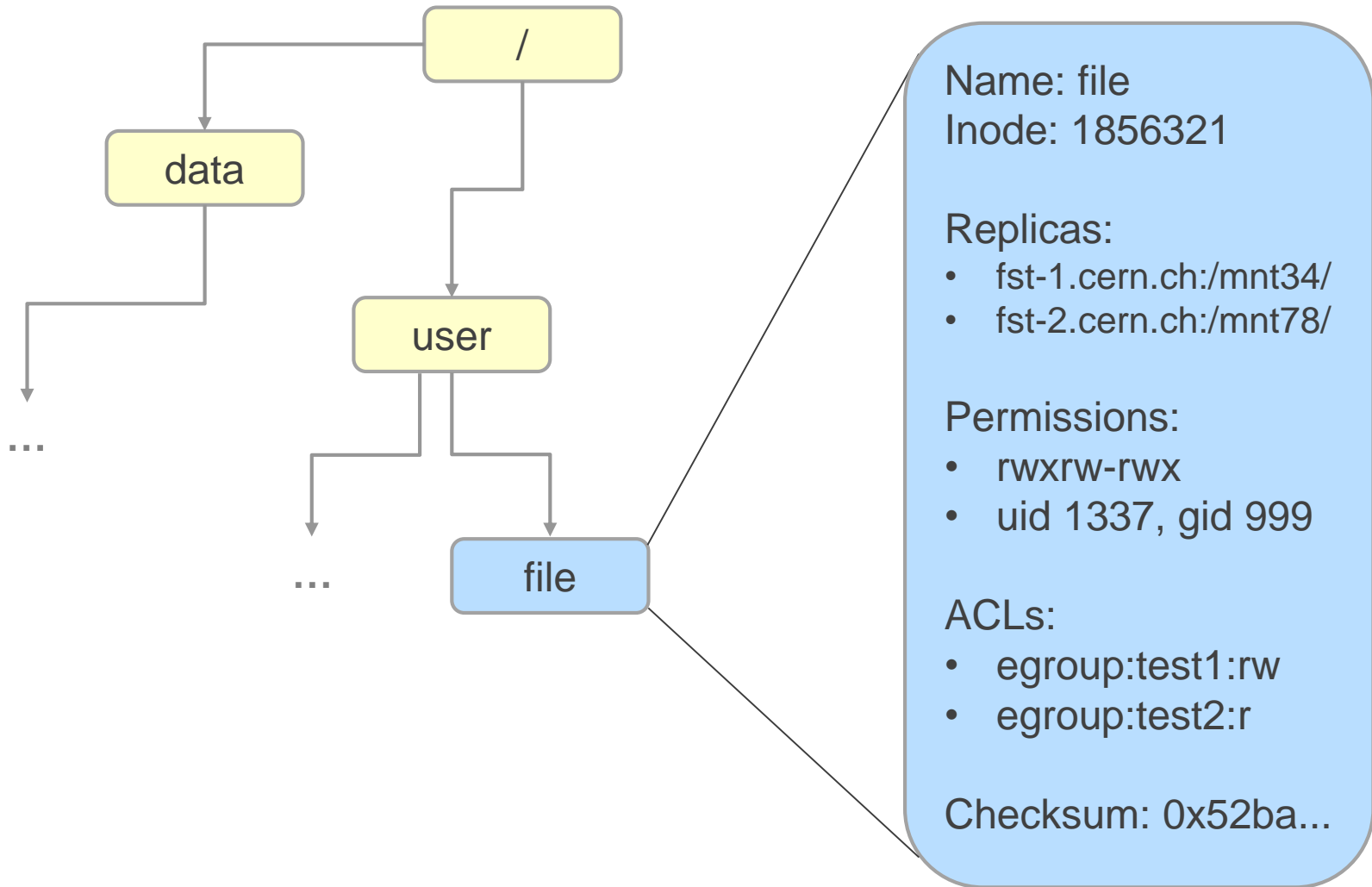- 15+ **prod EOS instances**

- 5+ billion **files**

One namespace

What is a namespace?

The eos namespace subsystem


```
                    / 
        ┌───────────┴───────────┐
      data                    user
        │               ┌───────┴───────┐
       ...             ...            file
```

Name: file
Inode: 1856321

Replicas:
- fst-1.cern.ch:/mnt34/
- fst-2.cern.ch:/mnt78/

Permissions:
- rwxrw-rwx
- uid 1337, gid 999

ACLs:
- egroup:test1:rw
- egroup:test2:r

Checksum: 0x52ba...

Summary

- Old namespace
  - Kept entire contents in-memory – often required **100-1000GB** of RAM on MGM
  - Took ages to reboot: **15 – 60 minutes** typical

- New namespace
  - ✓ In-memory **namespace cache**, configurable memory consumption
  - ✓ **Instant MGM reboot**, 1-10 seconds
  - ✓ **Just as fast** as old namespace

Project status: Completed

- Heavy development on the server-side completed since a year ago
- <u>All CERN instances migrated</u> (minus one)

Research into namespace alternatives, evaluation of options	Development on QuarkDB, and MGM namespace code	Heavy testing, production deployment in August	Instance migrations now routine; in-mem NS phaseout
2014-2016	**2016-2018**	**2018**	**2019**

Why migrate your instance to QuarkDB? (1)

- The carrot:

  - MGM now starts up in 10 seconds or less, **regardless of namespace size**

  - RAM requirements become reasonable, **configurable namespace cache** size

  - Powerful new tool **eos-ns-inspect**, complete visibility into QDB contents *without an MGM running*

Why migrate your instance to QuarkDB? (2)

- The stick:

  - In-memory namespace now receives minimal usage and testing – **bugs could slip through**

  - Several new features **not supported** with in-memory namespace (reworked master-slave; new fsck; eos-ns-inspect)

  - We'll soon **entirely deprecate and remove** the in-memory NS

Stability?

- QuarkDB uptime often reaches 6-9 months, usually interrupted for upgrades
  - Most bugs so far **infrequent** and **low-impact**

- Just 2-3 (substantial) bugs found and fixed in MGM namespace code during the past year; namespace code **Done™** and under **maintenance mode**

eos-ns-inspect
Full visibility into QuarkDB contents


```
$ redis-cli lhget eos-file-md 464244431
"\xa7\x03\x9b\xee\x8d\x00\x00\x00\b\xcf\x9d\xaf\xdd\x01\x10\xa1\xf6u
\x18\x98\x96\x05\xcb\x15(\xd6\x92\x010\x92\x82@8\xa4\x03B\bchickensR\x10
\xf2(^\x00\x00\x00\x00\xb9\x80\x99\t\x00\x00\x00\x00Z\x10\xf2
(^\x00\x00\x00\x00\t\t\xa2\n\x00\x00\x00\x00b\x14\x06m\xa0\xfd\x00\x00\x
00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00j\x04\xd0\x02\xa0\
x02z%\n\rsys.eos.btime\x12\x141579688178.161054905\x00\x00\x00"
```



```
$ eos-ns-inspect print --fid 464244431
ID: 464244431
Name: chickens
Container ID: 1932065
uid: 84760, gid: 2763
Size: 18774
Modify: Wed Jan 22 10:16:18 2020 Timestamp: 1579688178.178391305
Checksum type: adler, checksum bytes: 066da0fd
Expected number of replicas / stripes: 2
Locations: [336, 288]
Full path: /eos/user/g/gbitzes/chickens
...
```

Future plans

- Improve internal, inherited C++ namespace API: Make **certain operations atomic**, update namespace data transactionally
  - Refactor **a lot** of MGM code

- Investigate improvements to performance and **MGM responsiveness**.
  - example: Andreas' talk "A Mutex Curse" 👻 👻

- Infinite, elastic **scalability**? In the far future perhaps.

QuarkDB as MQ? Current status

- Necessary commands **implemented**:
  - PUBLISH
  - SUBSCRIBE
  - PSUBSCRIBE
  - VHSET, VHGET

- EOS code refactoring **in progress**, most of the work done already.

- **Prototype** for testing to be available soon-ish™ (a few months?)

Security considerations

- Please ensure your QuarkDB instances are **secure**!

- Instances left wide-open *will* be attacked, probably within days or hours.

- Use **password authentication** – will become required soon. Firewall your ports, if possible.

  https://quarkdb.web.cern.ch/quarkdb/docs/master/AUTHENTICATION.html

C'est fini

- Warm thanks to everyone who made this project successful!

- Documentation:

  https://quarkdb.web.cern.ch/quarkdb/docs/master/
  http://eos-docs.web.cern.ch/eos-docs/quickstart/ns_quarkdb.html

Questions, comments?