EOS Cross Tier Federation

Geoffray Michel Adde, Andreas-Joachim Peters – CERN/IT

Open Source Storage

Introduction

EOS is a storage software developed at CERN to store over 100 PB of experimental LHC and non-LHC and user data. The core of the implementation is the XRootD framework providing a feature-rich remote access protocol.
The storage system is running natively on commodity hardware with disks in **JBOD** configuration. The software has been evolved in the recent years to federate geographically distributed storage resources into a **virtualised storage cloud**. IO plugins allow additionally to attach external **S3** storage resources and **object disks** - *OpenKinetic* /SEAGATE standard.
The storage system provides three server components:

MGM: meta data and management server **FST:** file storage server. **MQ:** message queue server

Evolution

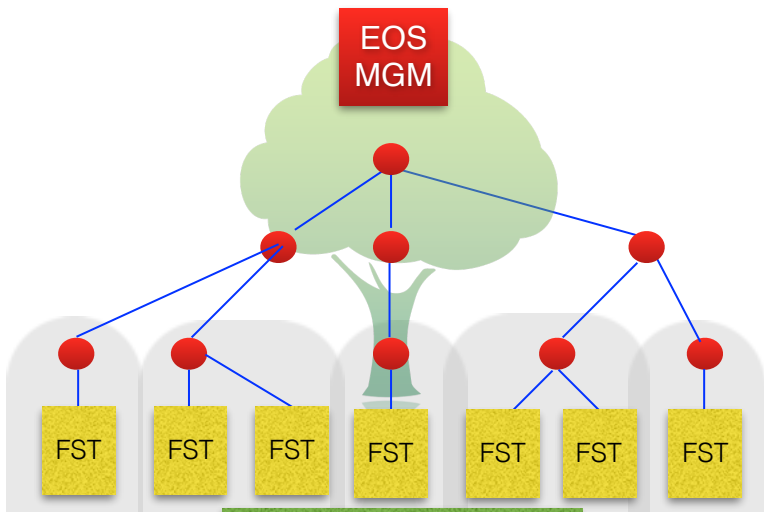
NO FEDERATION

EOS Amber **V0.2**
Flat scheduling with round-robin groups

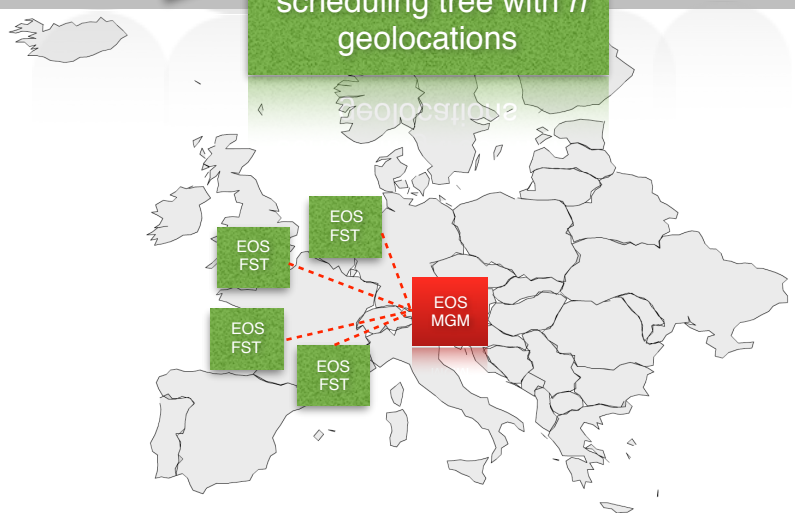
DUAL SITE FEDERATION

EOS Beryl **V0.3**
Flat scheduling with round-robin groups and two geographic locations

MULTI SITE FEDERATION



EOS Citrine **V4.X**
Flexible hierarchical scheduling tree with *n* geolocations



National T2 federation with external T1 namespace

Cross Tier Deployment Model

EOS allows to connect distant storage installations (FSTs) to a shared namespace instance (MGM). The installation procedure for a site administrator consists of five simple steps:

1. receive and install a shared secret **/etc/eos.keytab** on each storage node
2. install **eos-server** rpms
3. configure geographic location settings
4. register storage mount points using **eos-fst-register**
5. start the service **service eos start**

Cross Tier Placement Model

Geographic aware scheduling in EOS supports three generic placement strategies:

- maximum distance
- hybrid
- co-located

Introduction of a **virtual group concept** representing a virtual placement tree **enables cross tier placement** policies (e.g. one replica@ T1 - one replica@ T2)

IO plug-ins and Proxy Scheduling

Storage Backends

EOS supports FST detached storage backends e.g.

- **OpenKinetic** ethernet disks
- **S3** storage like AmazonAWS or RadosGW
- **XRootD** remote storage

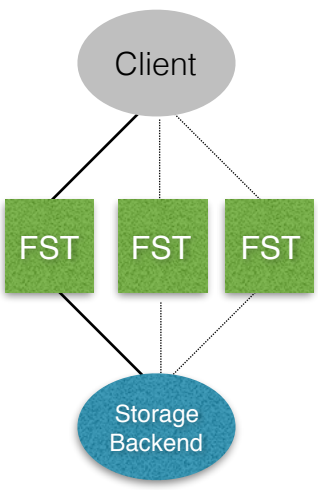
To provide scalable access to these backends EOS provides multi-path scheduling via configurable proxy gateways. No external load balancers like HA-Proxy or DNS load balancing are required!

File Stickiness

A backend like *OpenKinetic* profits from caching functionality inside the proxy gateways. This means that a given file is accessed via the same gateway machine to improve the **cache hit rate** - as long as the route is available.

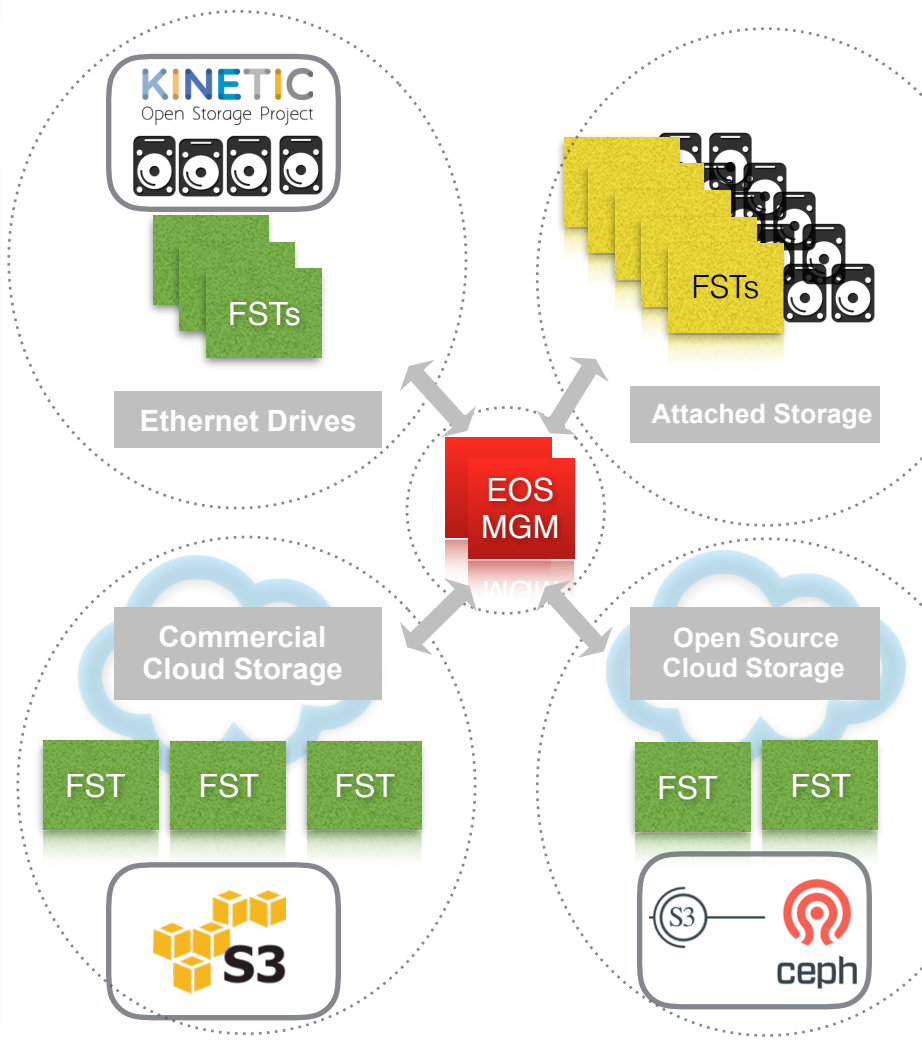
Proxy Scheduling

Proxies are tagged in the same way storage server are tagged - clients are scheduled in the same way to proxies: this allows e.g. to select the **closest/ optimal proxy routes** for site-internal and site-external clients.



Multi-path proxy setup
the black line illustrates a sticky access route for a given file

Deployment Scenarios



Multiple backend federations

Contact: Andreas.Joachim.Peters@cern.ch