HDFS and S3 plugins

Andrea Manzi
Martin Hellmich

13/12/2013
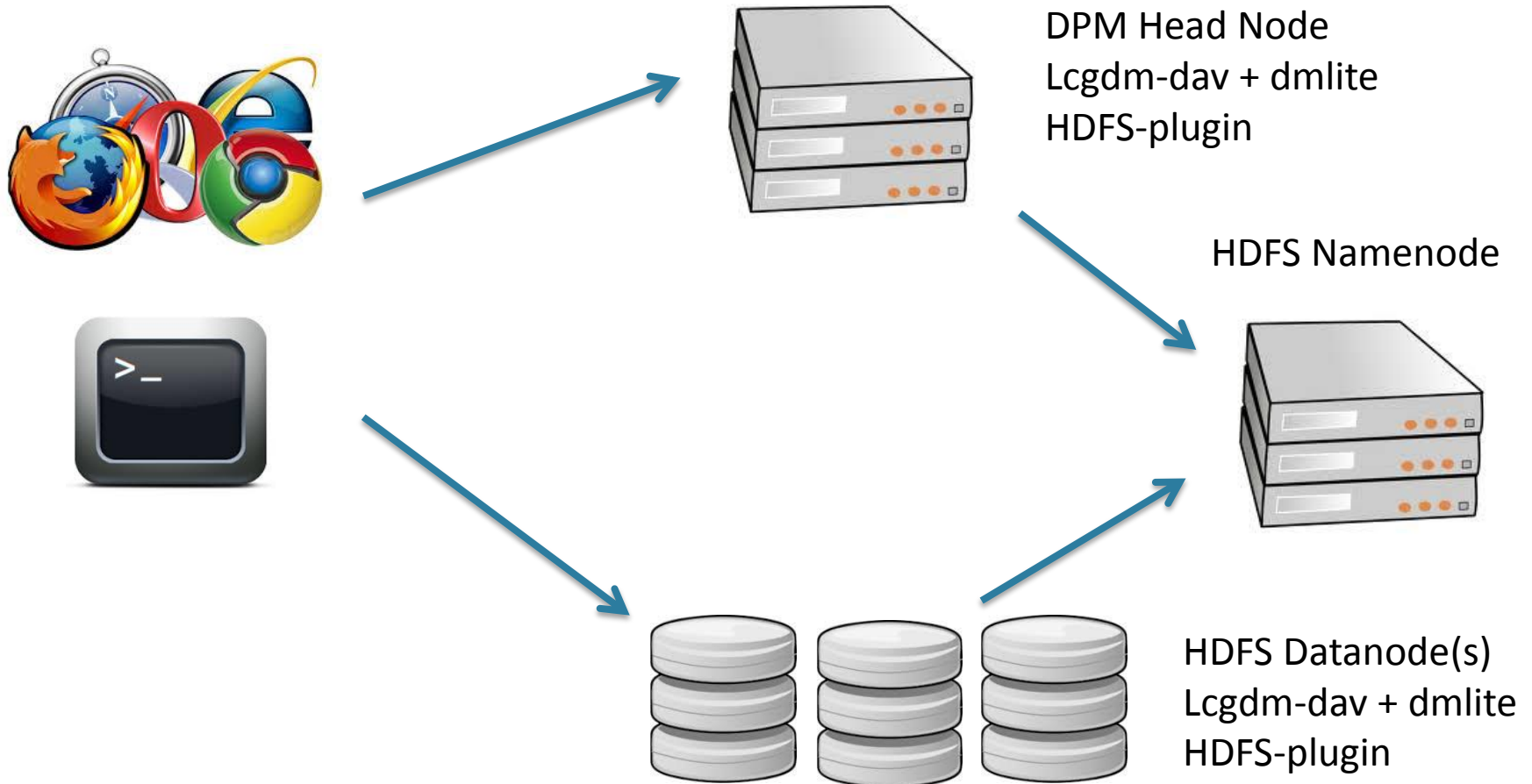
Plugins functionalities

NFS	HTTP/DAV	XROOT	GridFTP	RFIO

Namespace Management	Pool Management	Pool Driver	I/O
Legacy DPM	Legacy DPM	Legacy DPM	Legacy DPM
MySQL	MySQL	HDFS	HDFS
Oracle	Oracle	S3	
HDFS			
Memcache			
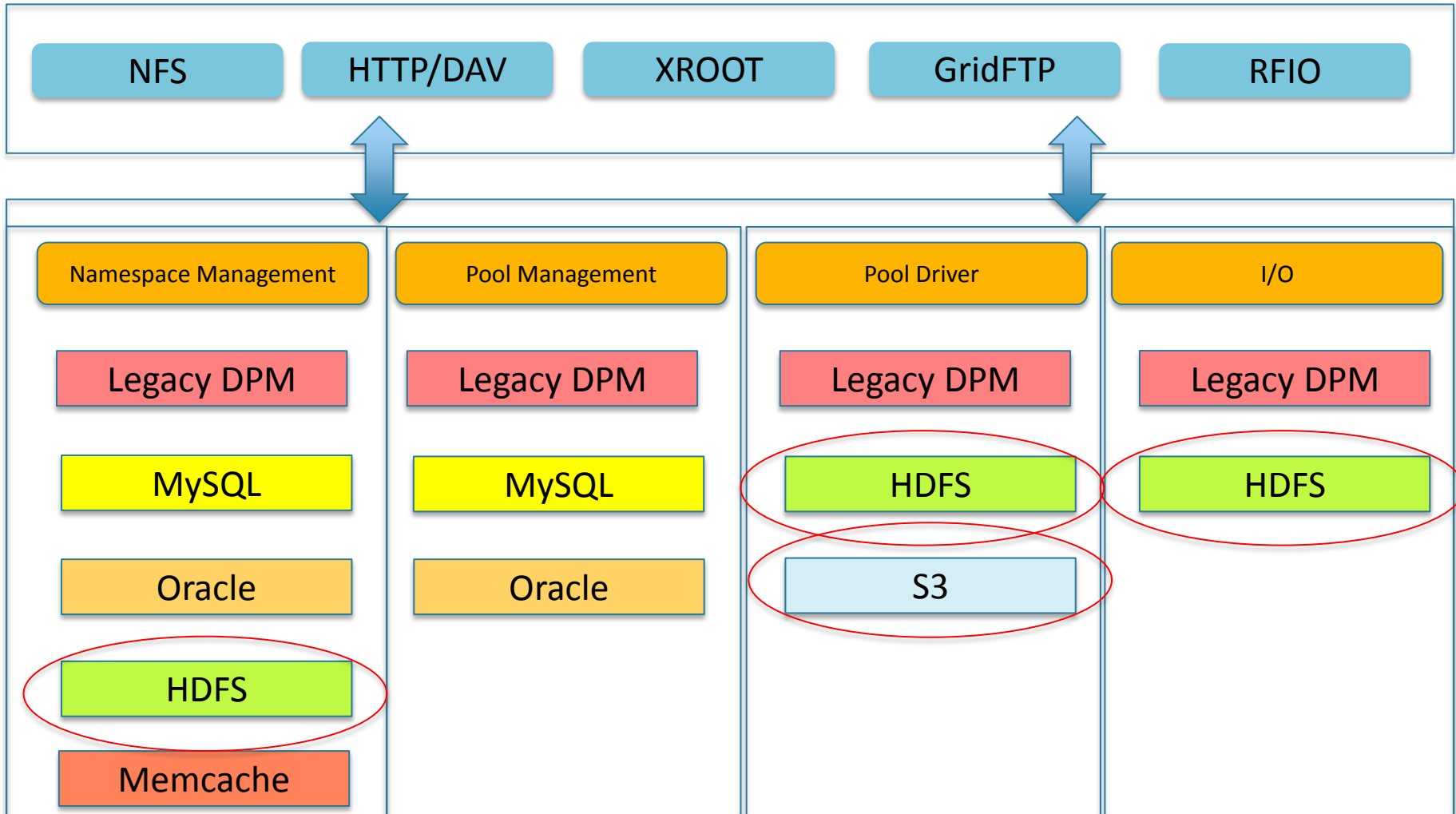
HDFS plugin

- dmlite plugin implementing I/O, pool driver and namespace functionalities through Apache Hadoop HDFS ensuring:
  - Automatic data replication
  - Fault tolerance to client's read
    - Dead of Datanode and Namenode
  - Scalability

Deployment with Lcgdm-dav

DPM Head Node
Lcgdm-dav + dmlite
HDFS-plugin

HDFS Namenode

HDFS Datanode(s)
Lcgdm-dav + dmlite
HDFS-plugin

Some details

- HDFS C APIs (libhdfs) do not implement functions to retrieve the available datanodes (LIVE nodes)
  - Patch implemented and submitted to Hadoop
  - hadoop-libhdfs rpm from our repo

- First version for Puppet installation is available.
  - To be adapted to recent dav/dmlite module changes

On-going issues

- Tested with new dmlite-based GridFTP plugin
  - Same deployment model as http/dav frontend or single node writing to HDFS
  - But…HDFS does not support **multiple write streams / random writes**:
    - OSG developed in-memory stream reordering in GridFTP in order to avoid this limitation (**gridftp-hdfs** DSI available also in Globus toolkit)
  - To test and understand integration

On-going issues

- SRM frontend does not speak dmlite

  - SRM calls through old dpm daemons do not handle properly new pools (as HDFS)
 - Patch to dpm daemon **to be implemented**

Future steps

- Distribution:
    - Need to understand how to distribute the plugin
    - HDFS client only in Fedora 20 and Rawhide
        - https://apps.fedoraproject.org/packages/libhdfs
- Support for security enabled HDFS clusters (Kerberos)

Performances
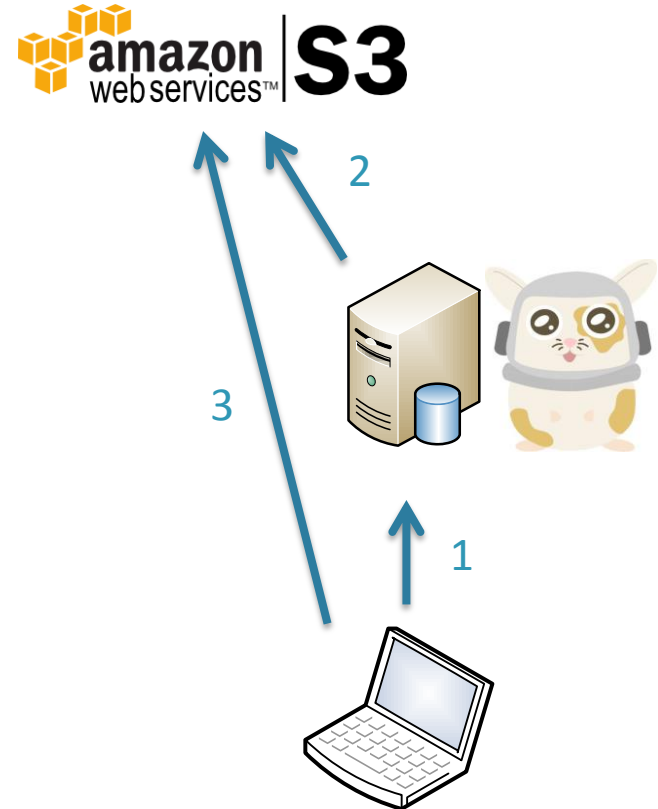
Tests through LCDM-DAV:

- HDFS Namespace
  - stat/s half performances compared to Mysql plugin namespace
  - To be optimized with Memcached in front
- ROOT analysis with massive Vector I/O and TTreeCache
  - Comparable performance with standard disk pools
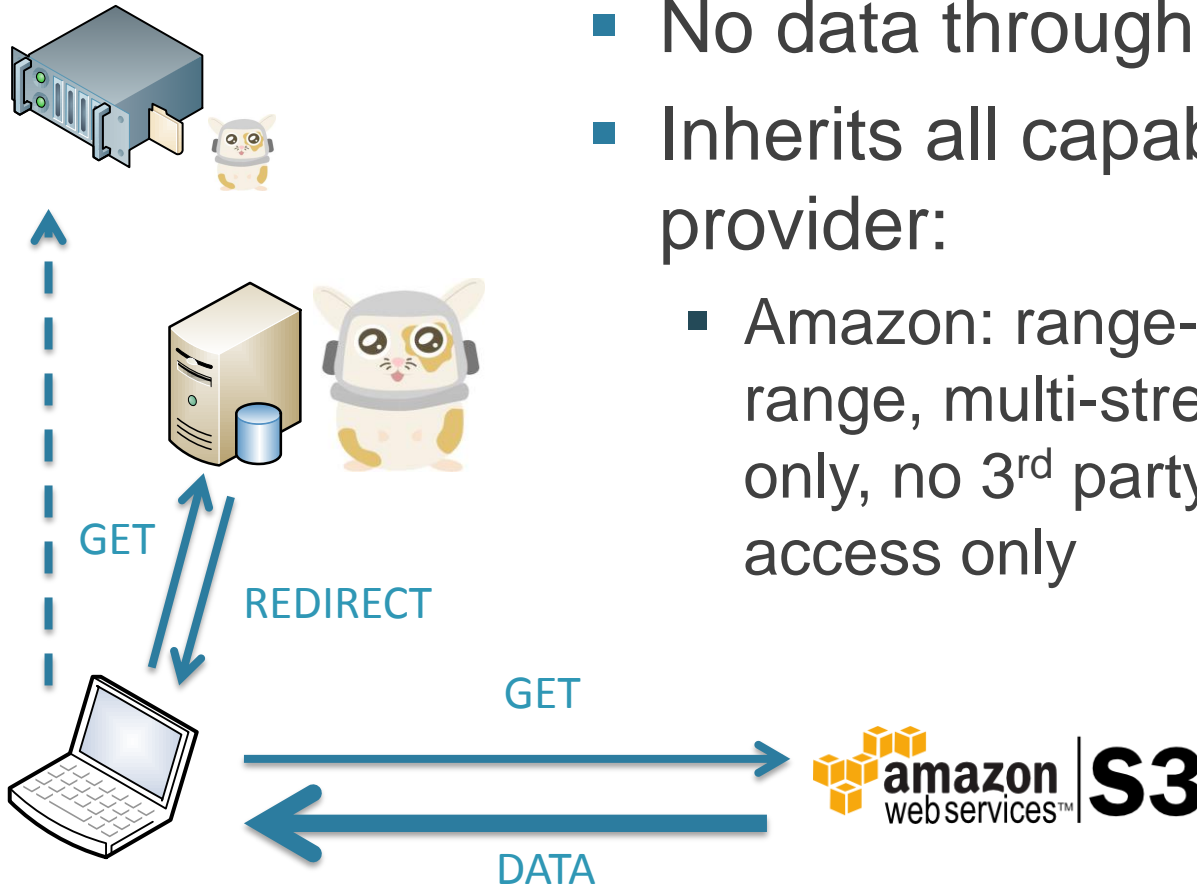
S3 plugin
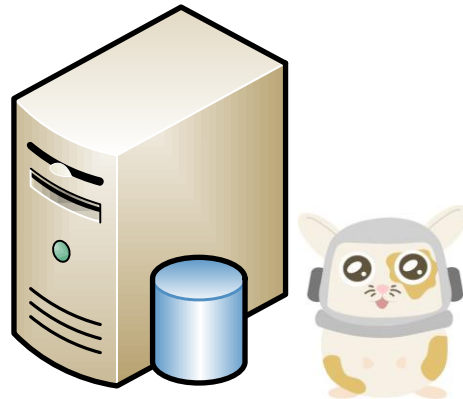
Key Facts

- Data directly to the cloud

- HTTP/HTTPS only

- DPM provides the namespace

Data in the Cloud

- No data through DPM
- Inherits all capabilities from S3 provider:
  - Amazon: range-header, no multi-range, multi-stream download only, no 3rd party copy, http access only

GET

REDIRECT

GET

DATA

How to install an S3 pool

- yum install dmlite-plugins-s3

- dmlite-shell

> pooladd poolaws s3
> poolmodify poolaws bucketsalt xFVlsrg
> poolmodify poolaws s3accesskeyid <ID>
> poolmodify poolaws s3secretaccesskey <SK>

- <create an s3 bucket on your storage>

More info

- HDFS plugin
  - https://svnweb.cern.ch/trac/lcgdm/wiki/Dpm/Dev/Dmlite/Plugins/HDFS

- S3 plugin
  - https://svnweb.cern.ch/trac/lcgdm/wiki/Dpm/Dev/Dmlite/Plugins/S3

Thanks!

Questions?

IT-SDC