

Statistics on disk usage CC-IN2P3

Pre-GDB / Data Management

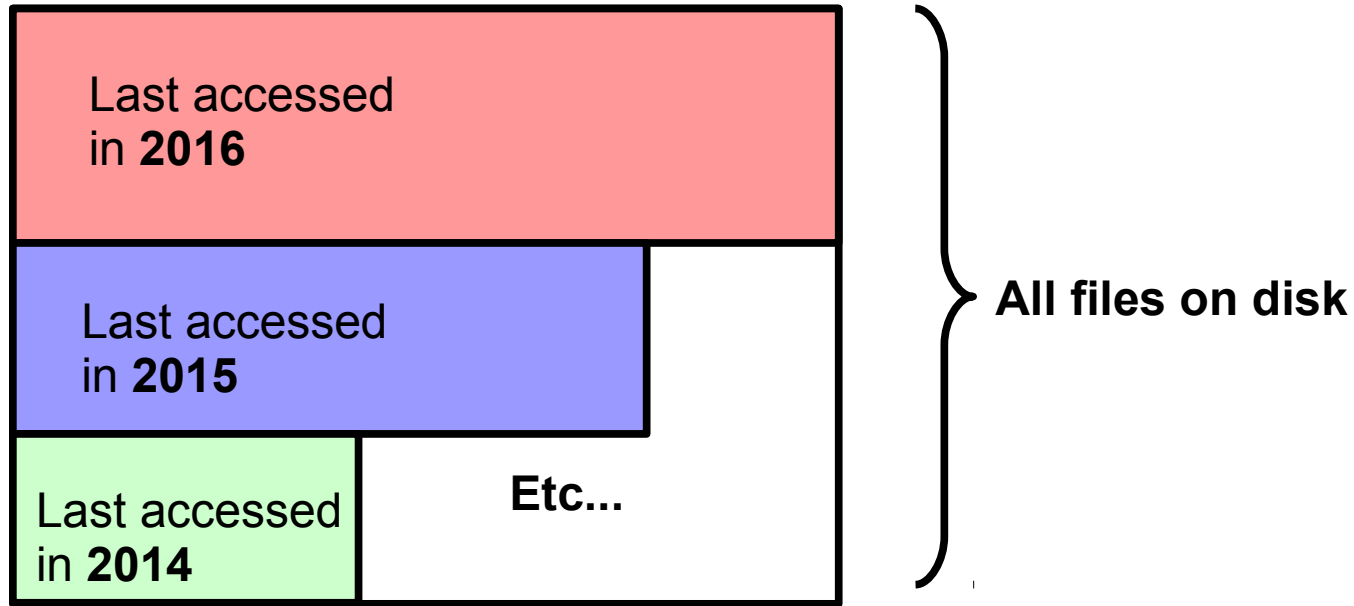
Fabien Kleinbourg, Renaud Vernet

Procedure

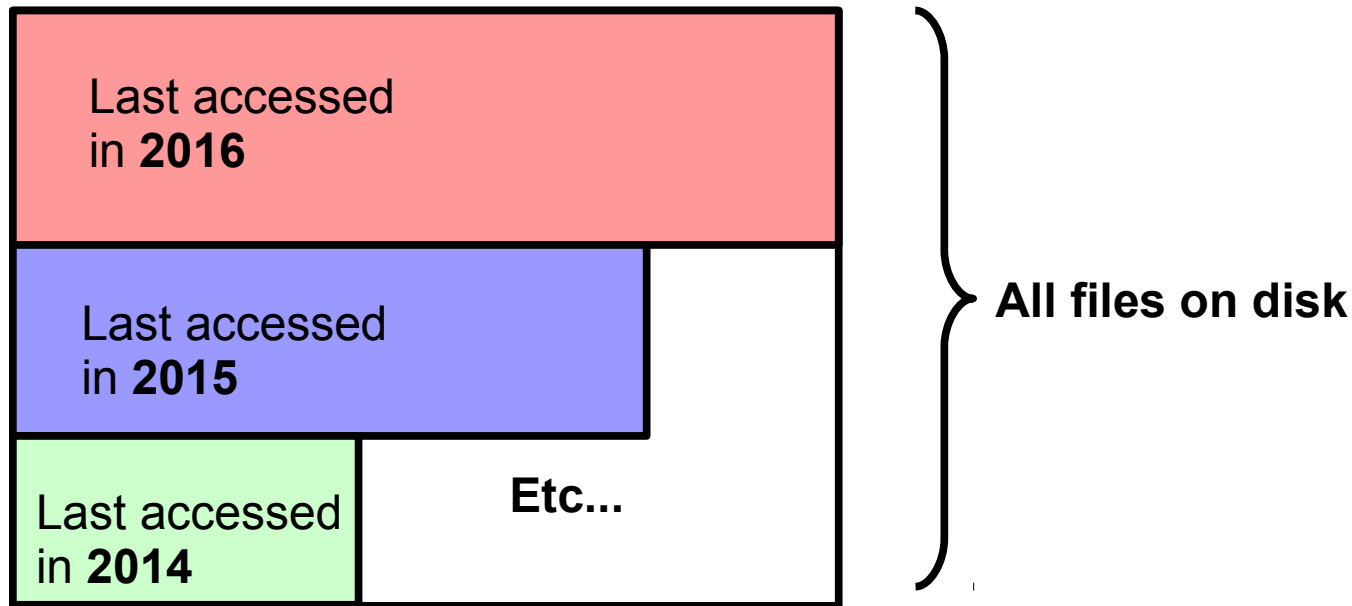
- LHC data @ ccin2p3
 - ATLAS, CMS, LHCb : dCache
 - ALICE: xrootd
- For (file on disk) :
 - `stat` file
 - record last 'modify' and 'access' times

```
OK kleinbou@ccdcacli005 /data/pool-cms-analysis-li005a/pool $ stat data/0000000180B3739F42B7A3AEDC9D8E769FF5
File: 'data/0000000180B3739F42B7A3AEDC9D8E769FF5'
Size: 2229756060      Blocks: 4355000      IO Block: 4096      regular file
Device: fd02h/64770d  Inode: 17186875534   Links: 1
Access: (0644/-rw-r--r--)  Uid: (  0/   root)   Gid: (  0/   root)
Access: 2015-12-09 10:40:04.159793778 +0100
Modify: 2015-12-07 14:33:29.869000000 +0100
Change: 2015-12-07 14:33:29.868464574 +0100
Birth: -
```

Grouping by access period



Grouping by access period

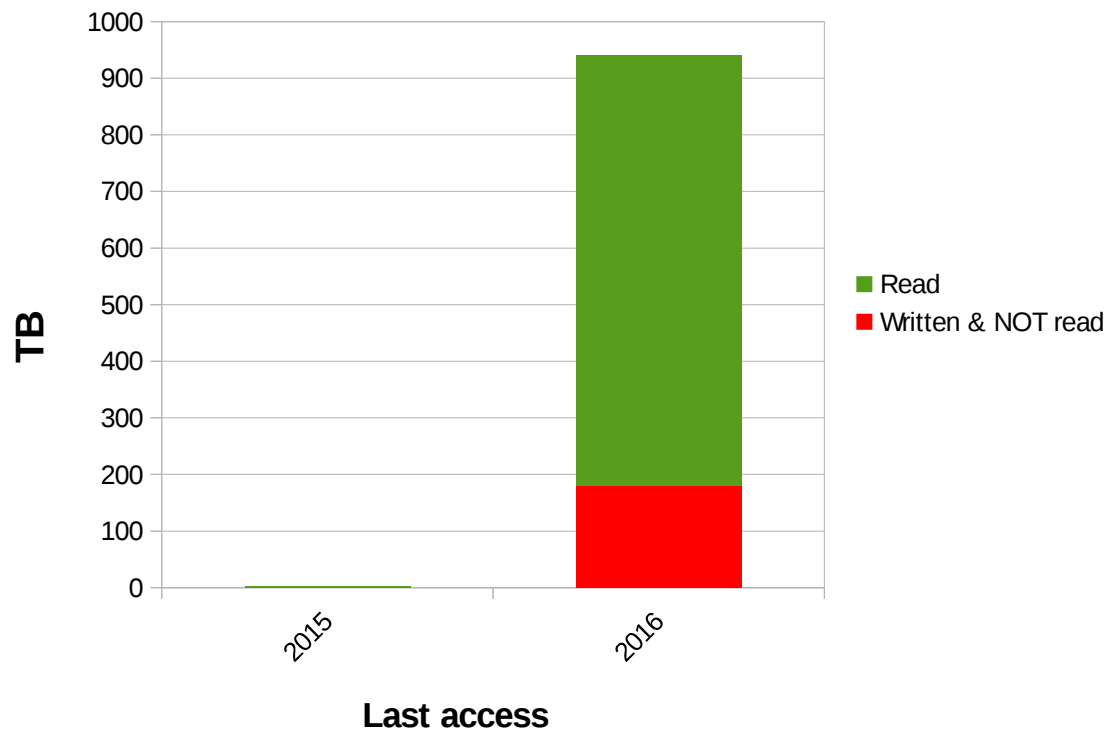


- File creation time not so important
 - File read in 2016 may have been written in 2013
 - If a file is read, it is 'alive'
 - → focus on files created but never read
- Statistics from snapshot
 - Get the picture at a given 'time' (last week)

Goal:
Get the amount of
inactive files (size)

Data usage : CMS on dCache

pool cms-analysis

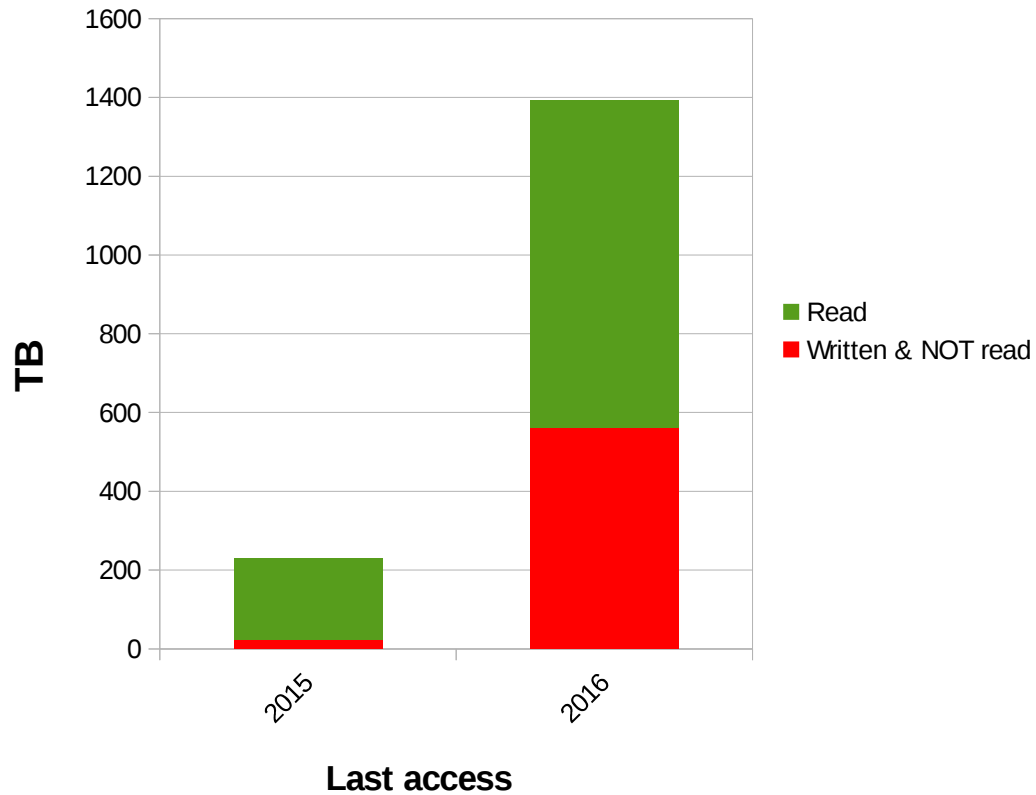


→ ~all data present has been accessed this year

→ ~ 200 TB written this year but not read (yet)

Data usage : ATLAS on dCache

pool atlas-aod-read

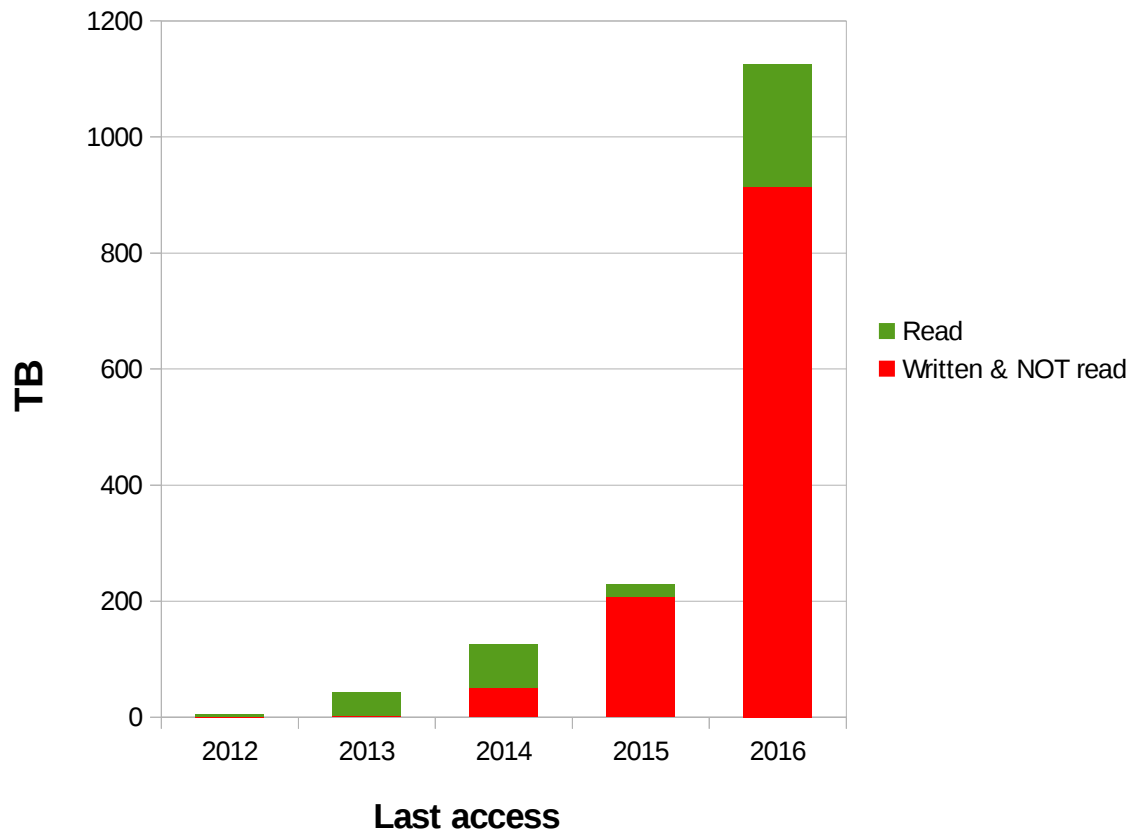


→ 200 TB not read since 2015

→ 600 TB written this year but not read (yet)

Data usage : LHCb on dCache

pool lhcb-dst



- most data accessed in 2016
- 900 TB of the 2016 data not read (yet)
- 400 TB of data not accessed after 2015
- 250 TB written never read after 2015

Summary

- Example of detailed statistics on data usage
- Insight on 'active' and 'inactive' data
 - Eg : When to move data to tape
- Can this be helpful to you ?