

Unified Quality-of-Service and Data-Lifecycle Definitions for Data Storage and Access

... or how to managing expectations



HELMHOLTZ
ASSOCIATION

Paul Millar



RDA Plenary 6 BoF (2015-09-25)

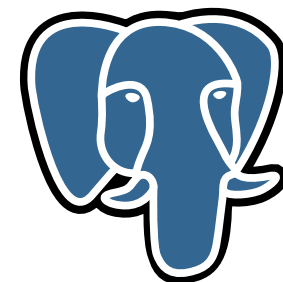
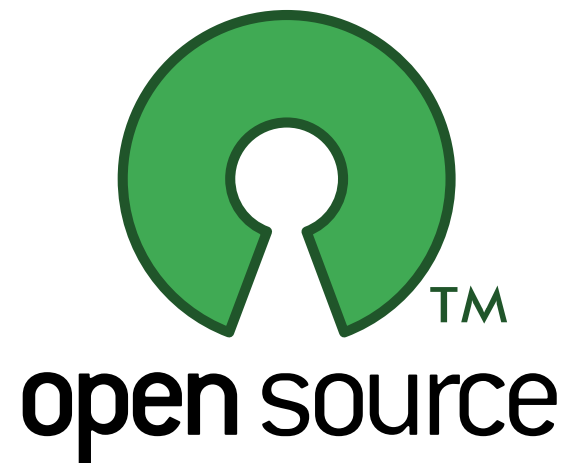
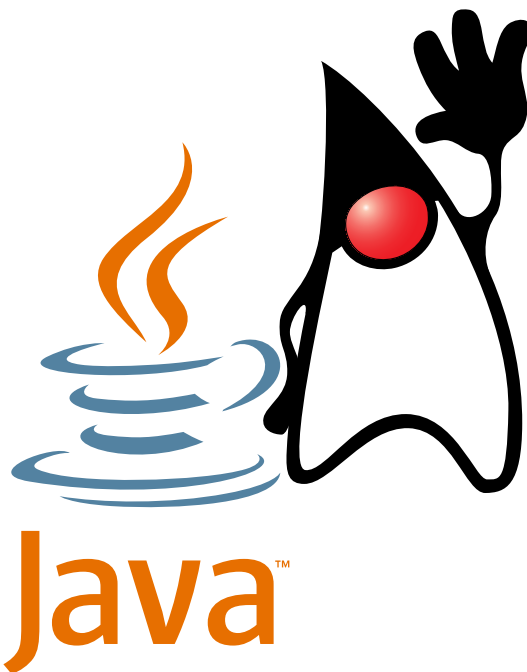
INDIGO DataCloud

Why are we here?

In INDIGO-DataCloud ...

- We've identified a problem,
(well, two actually),
 - We want to fix this problem,
(we hope you do too!)
 - We want your help in fixing it
(we hope you do too!)
-

Storage software: Free, Open-Source



<https://github.com/dCache/dCache>

<mailto:dev-subscribe@dcache.org>

Software running throughout the world



dCache and INDIGO-DataCloud



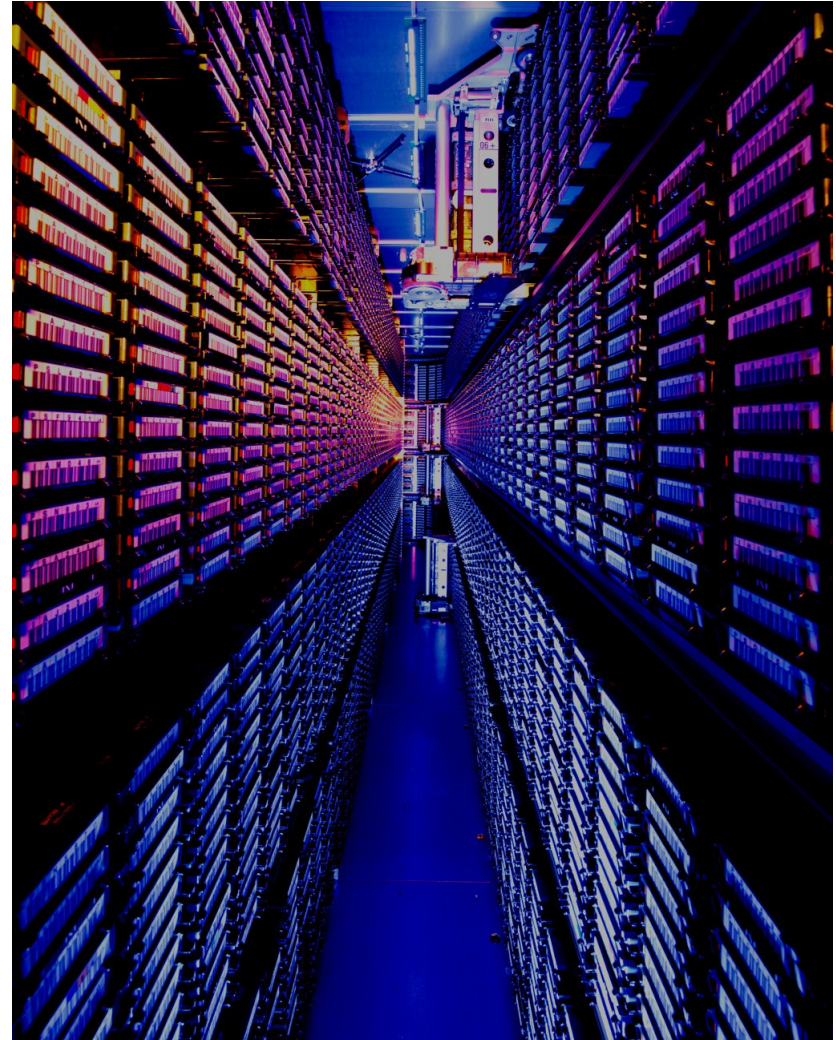
The problem...

Quality of Service
and
Data Life-Cycle

Quality of Service



Store data on disk or tape?



Now we have more media options



Replicating data

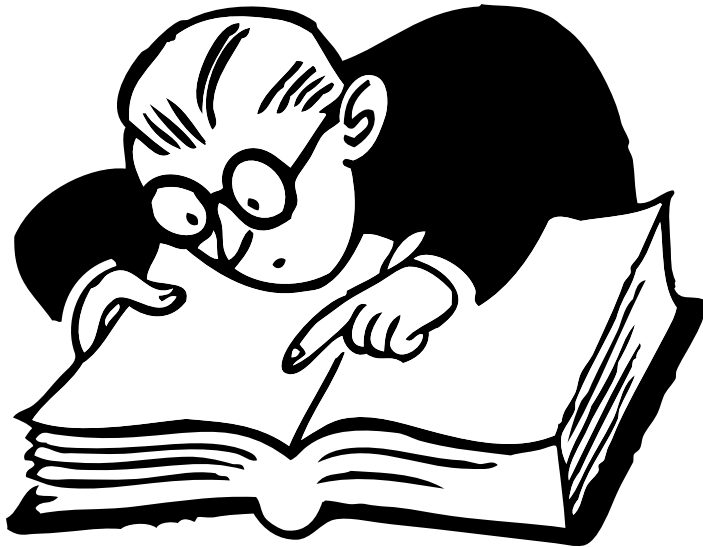


How many copies? Where are they located?

Motivation: budgets



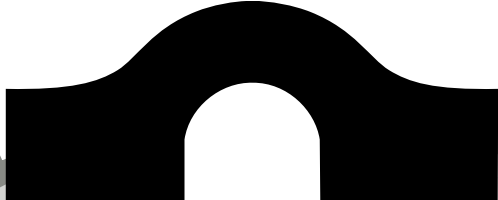
How to make this a possibility



What are my options?

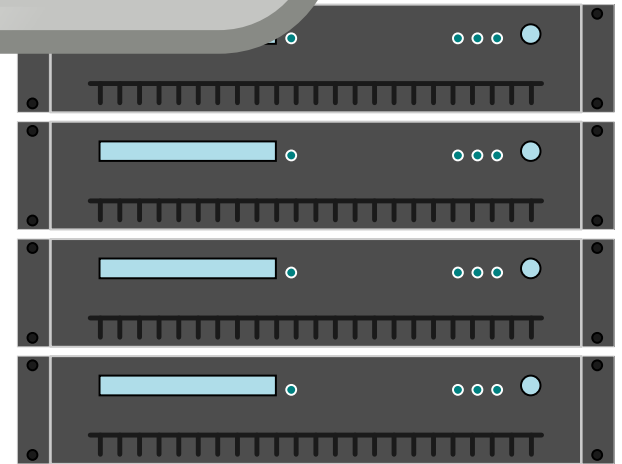
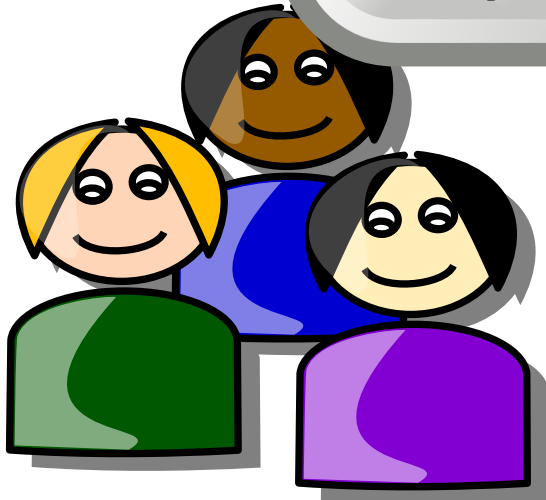
How do I choose?

Bridging the gap

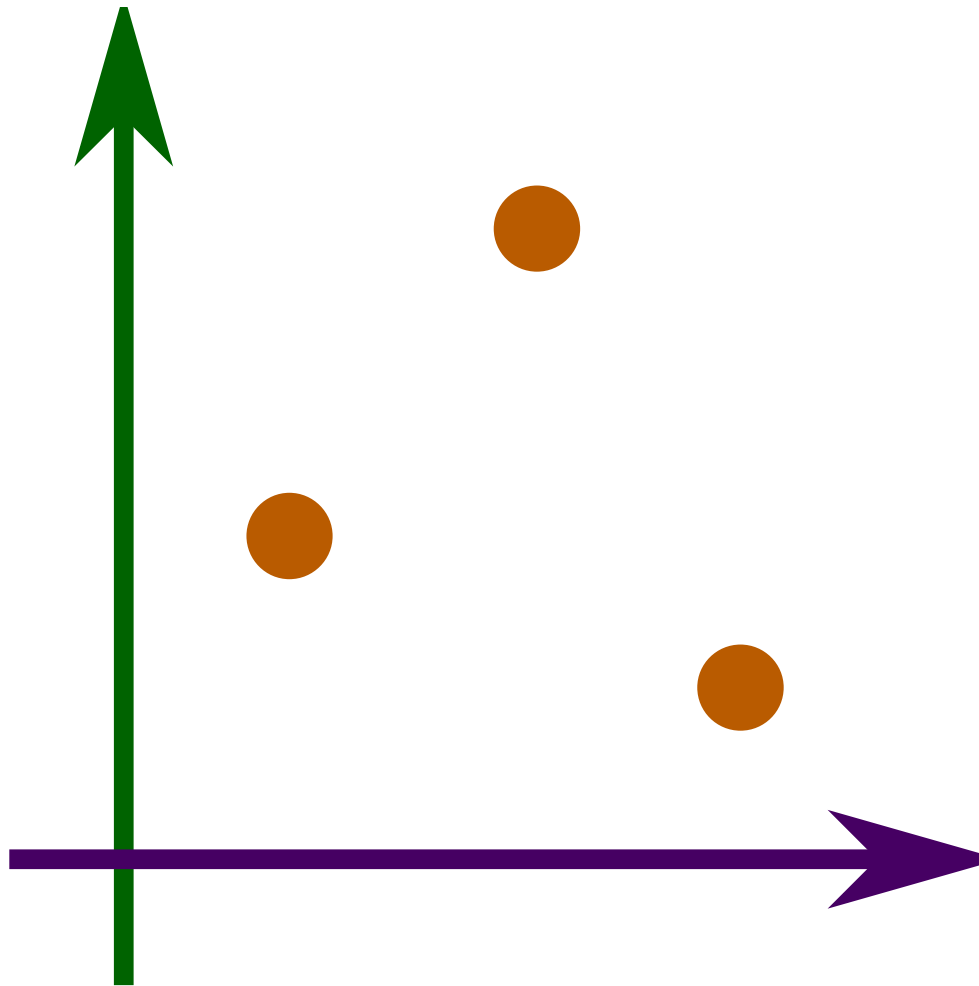


Concepts that users easily understand

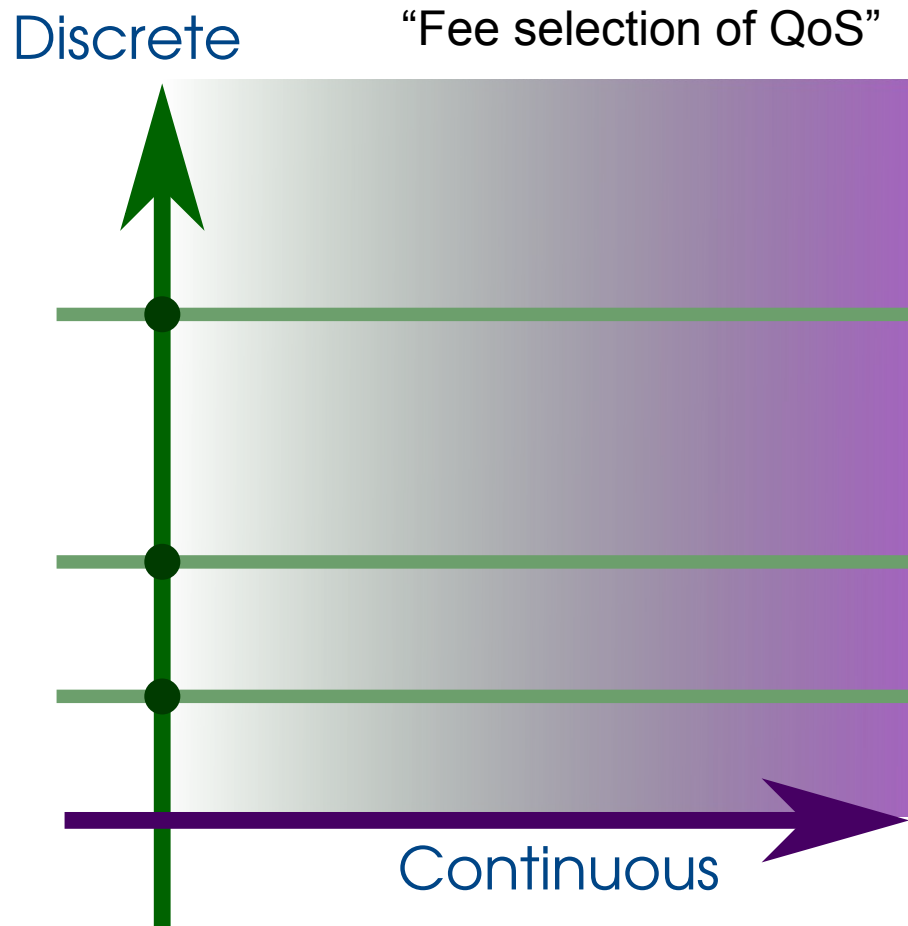
Concepts that storage systems easily understand



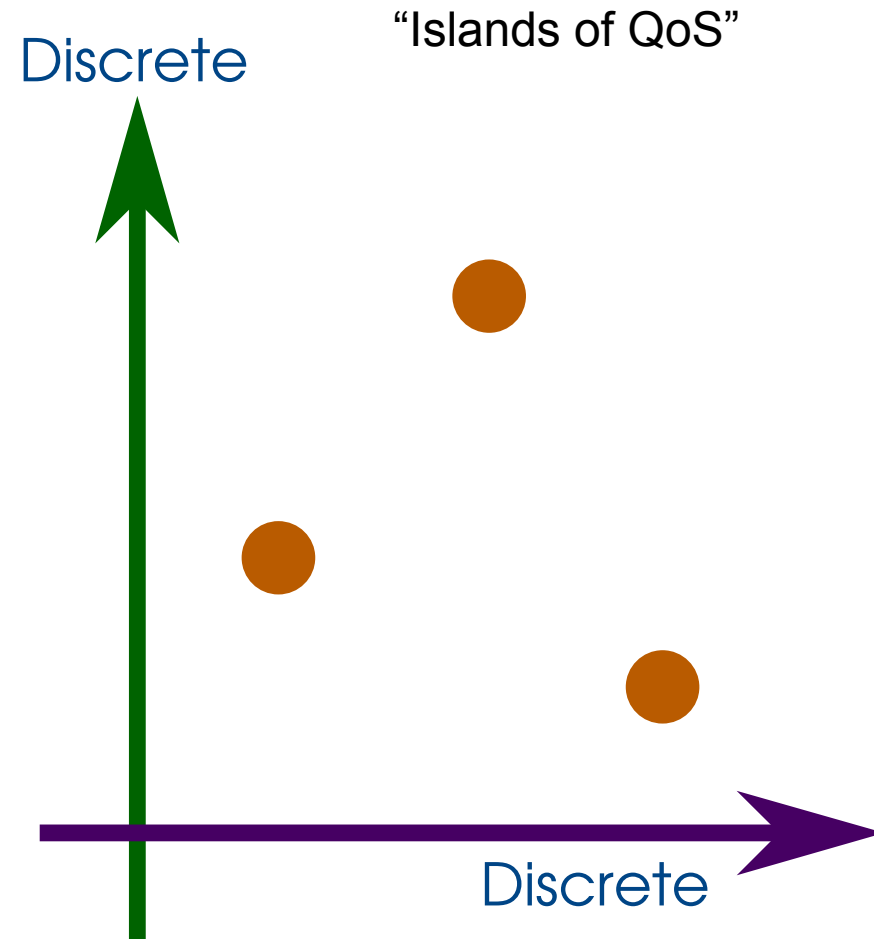
Attributes and islands



Combining QoS attributes

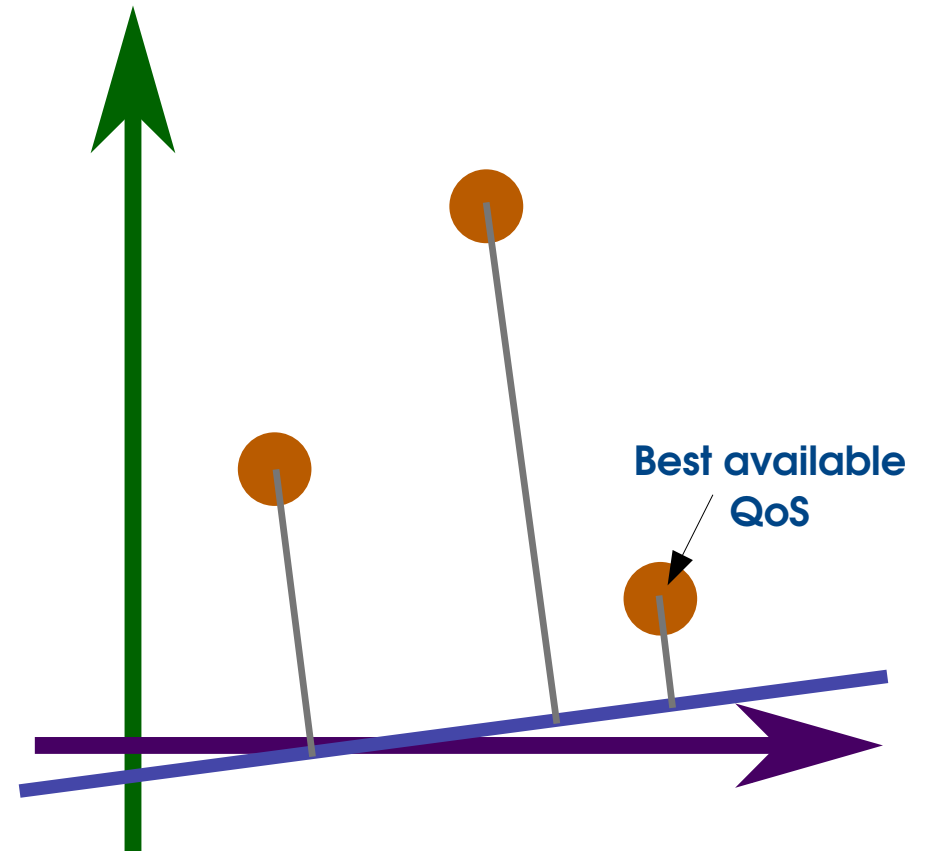
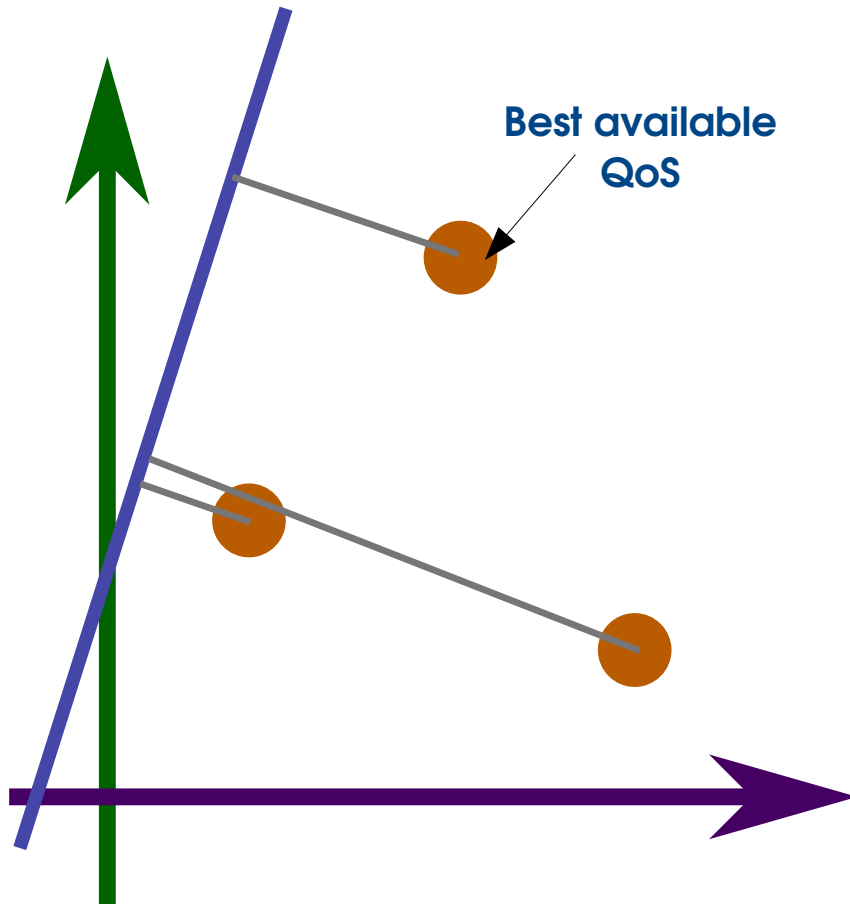


Independent



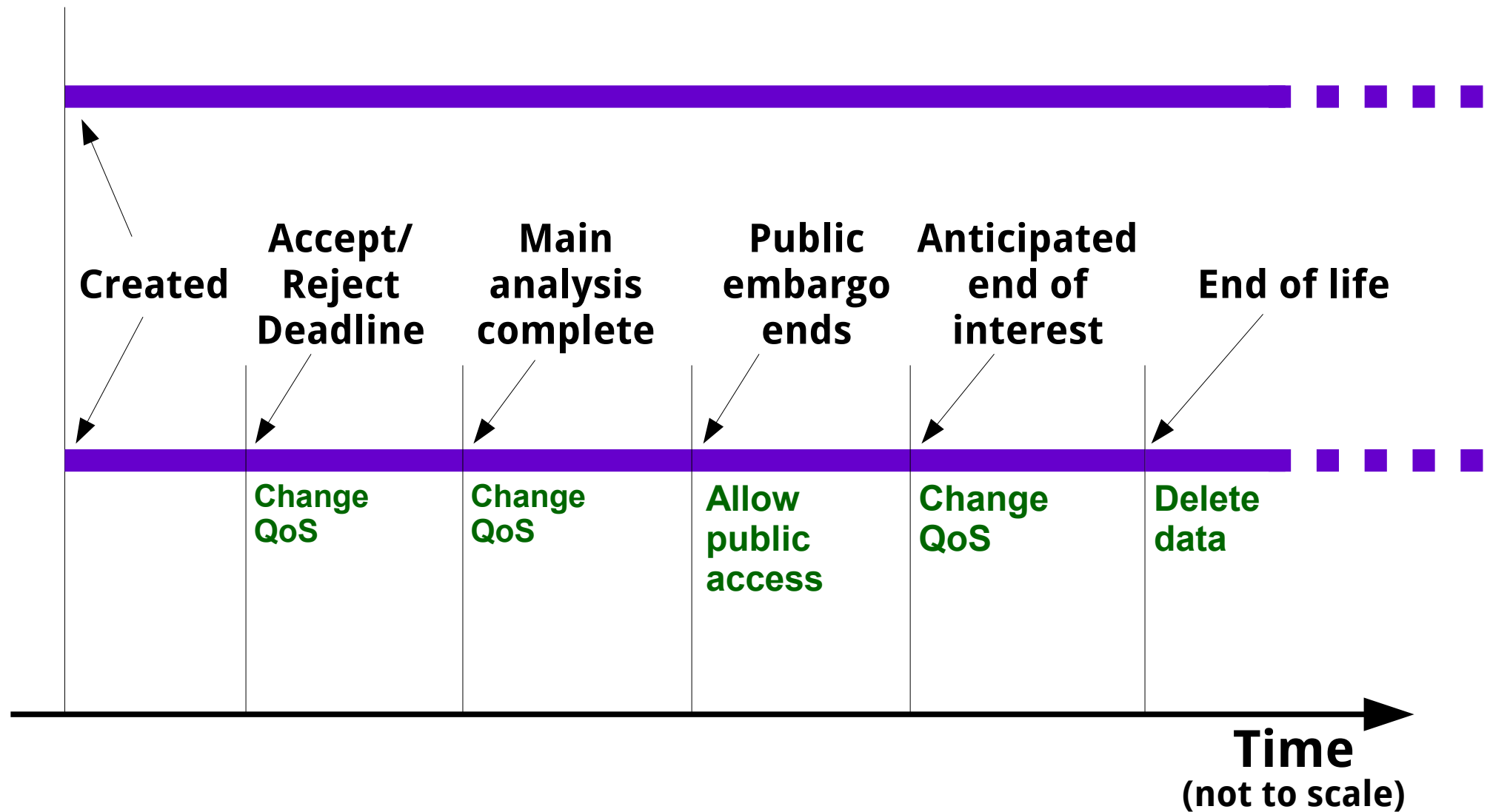
Dependent

Figure-of-merit: allowing decisions



Data Life-Cycle

DLC use-cases: the story of a file



Format for DLC rules

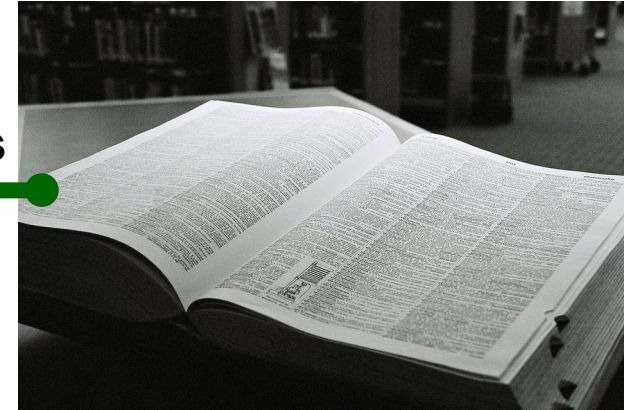
<trigger> <action>

(e.g., <after 6 months> <add public-access ACE>)

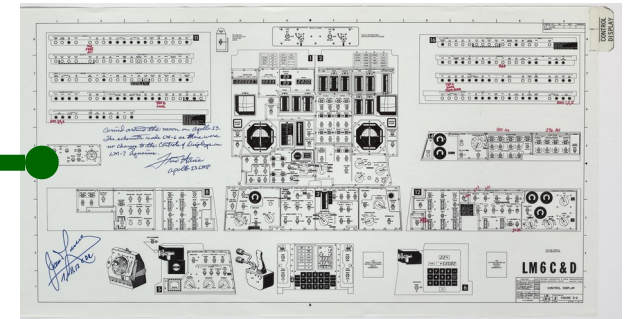
The plan:



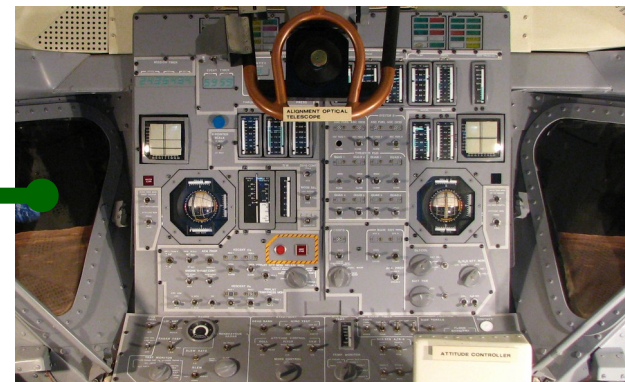
Definition of terms



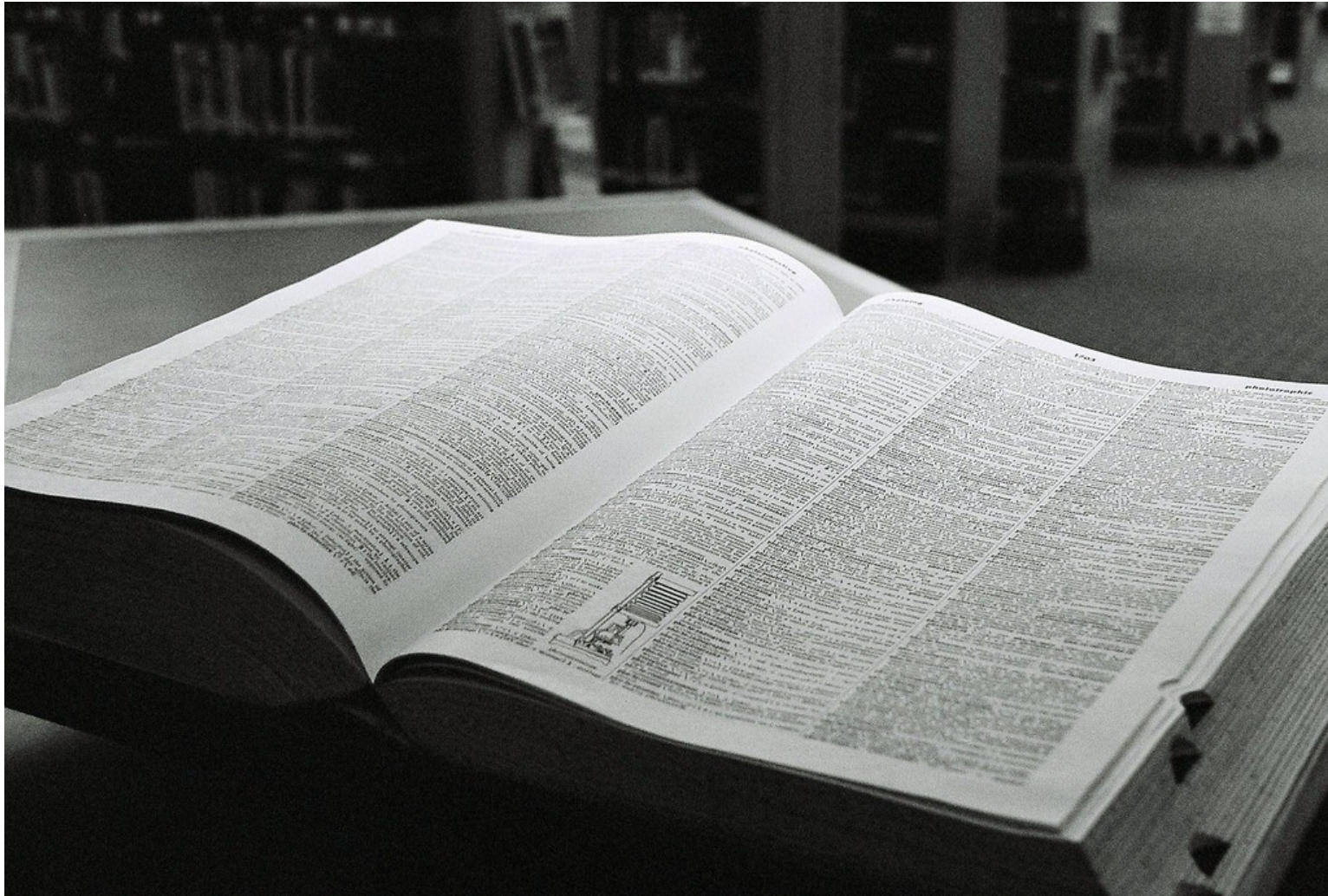
Protocol definition



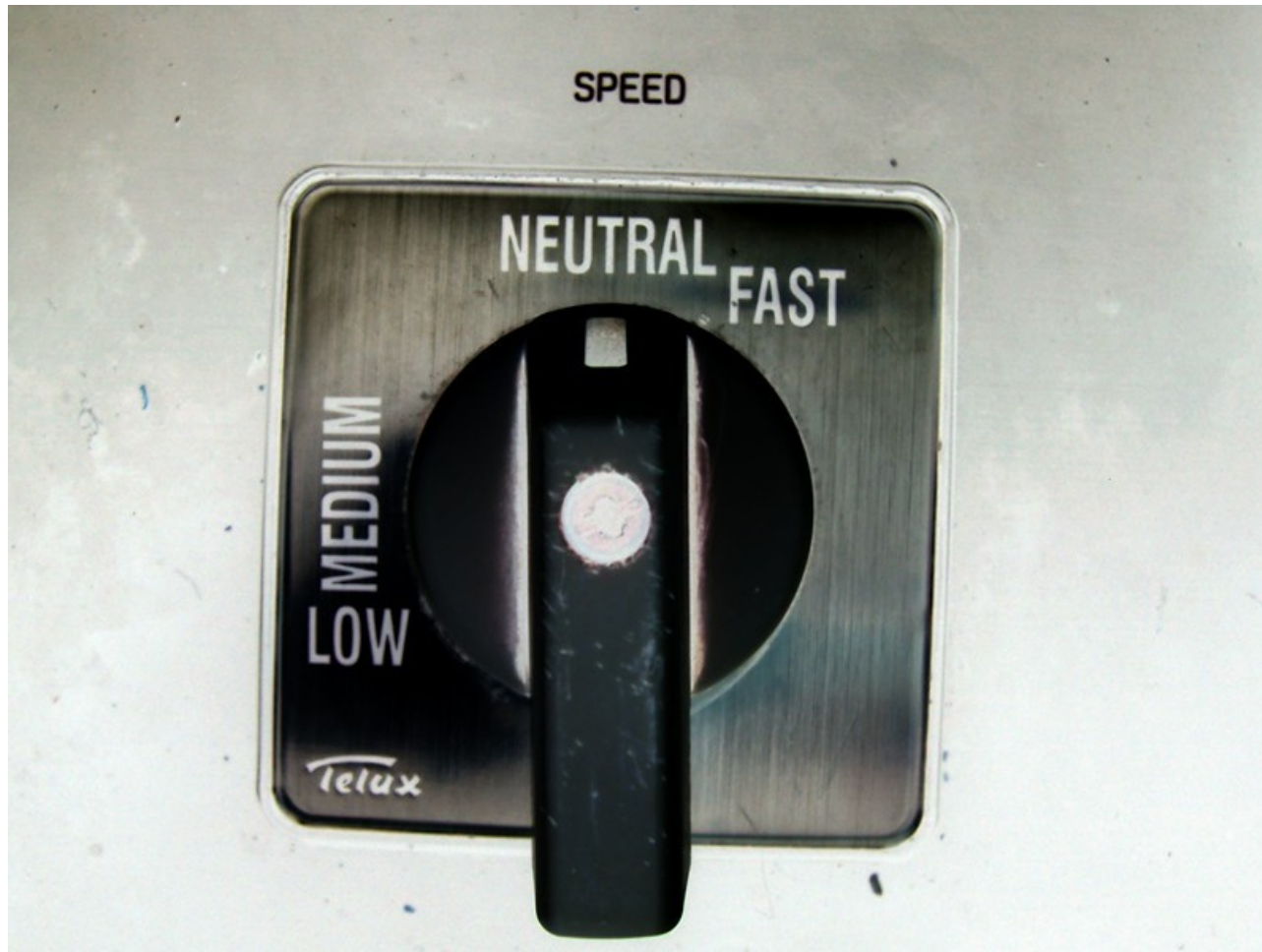
Implementations



Proposal: RDA WG “dictionary of terms”



“Speed” access-latency or bandwidth?



How fast is “High”?



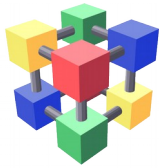
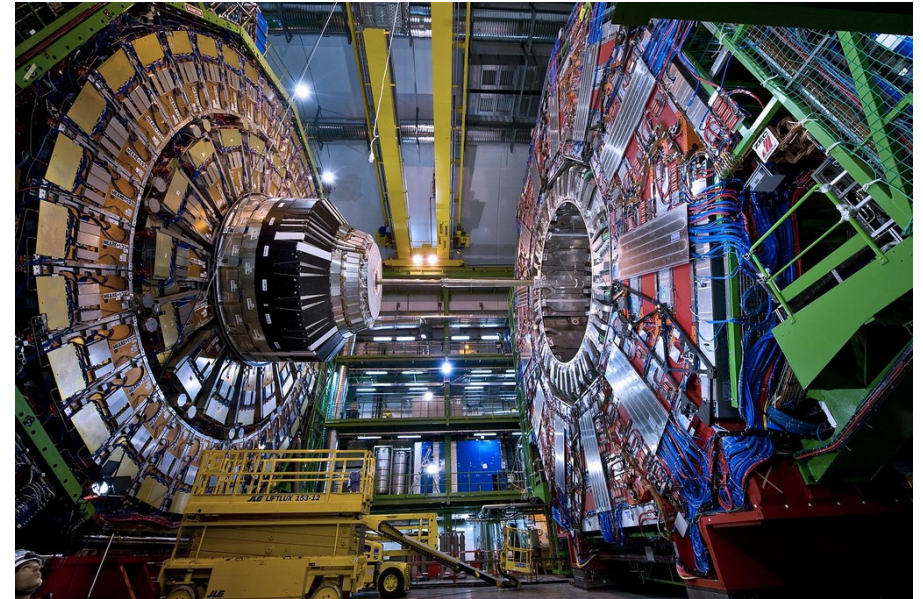
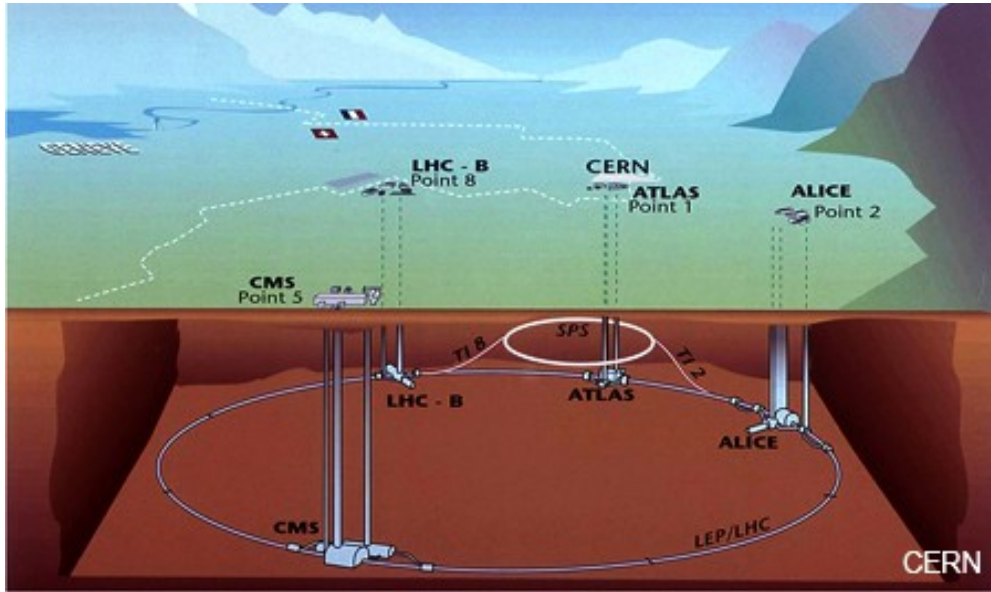
Would you be able to work this?



Thanks for listening



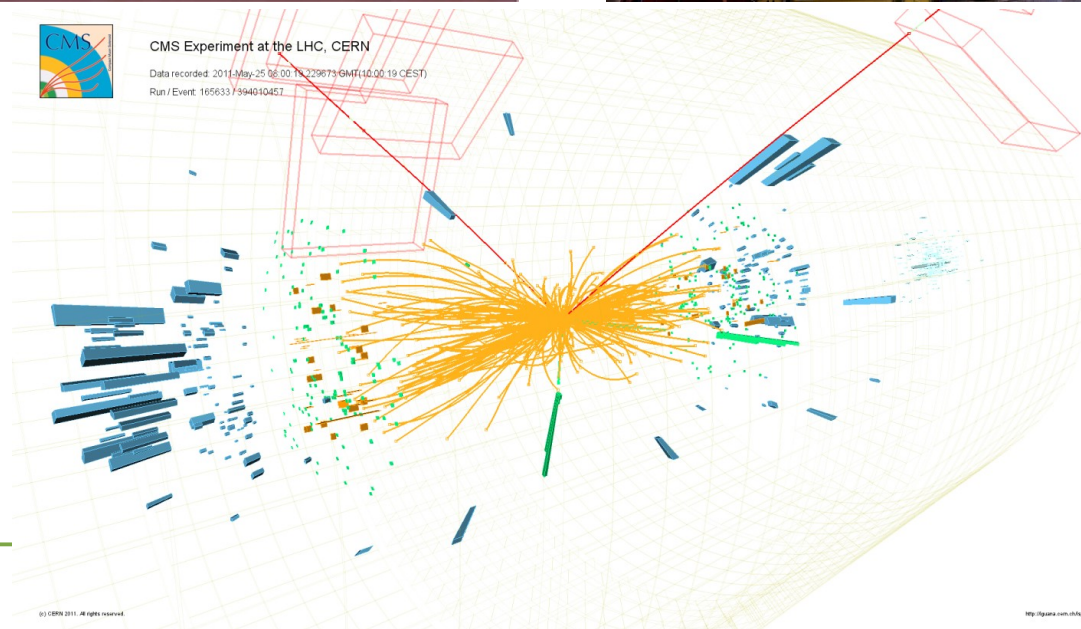
Used to search for the Higgs boson



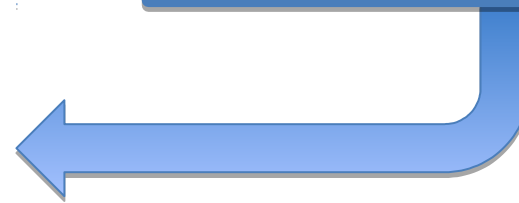
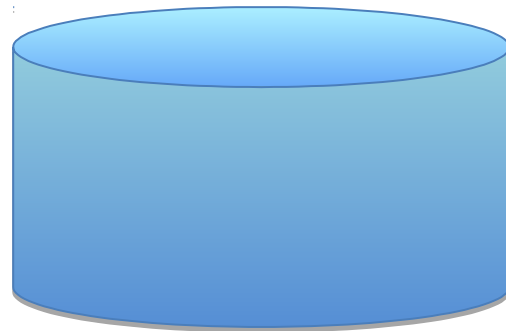
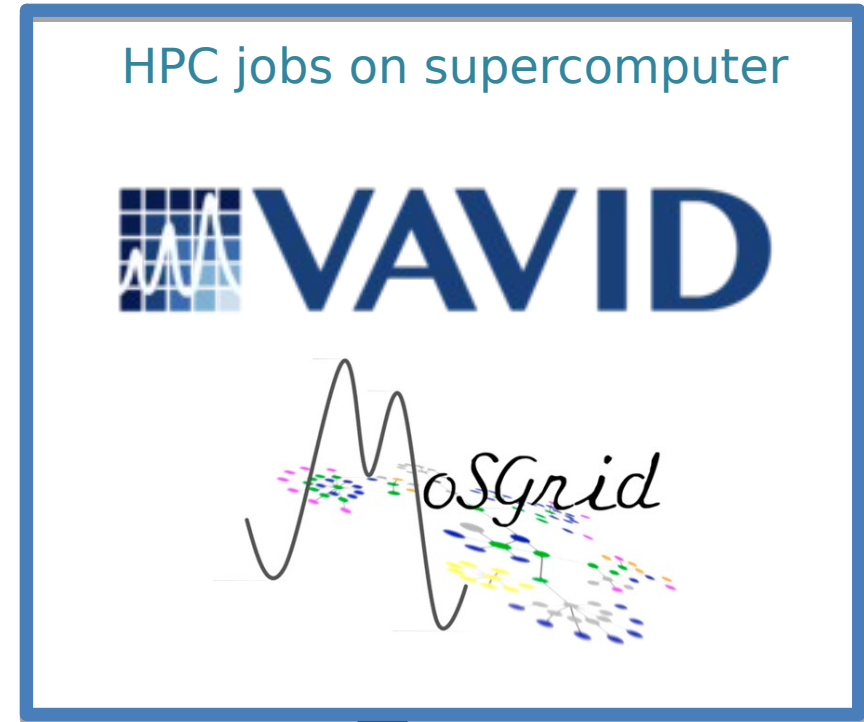
WLCG
Worldwide LHC Computing Grid



CMS Experiment at the LHC, CERN
Data recorded: 2011-May-25 06:00:15.229673 GMT(10:00:19 CEST)
Run / Event: 165633 / 994010457



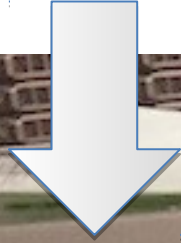
Feed data for HPC applications



HPC jobs get access to dCache storage.

Research: pushing frontiers

Power supply



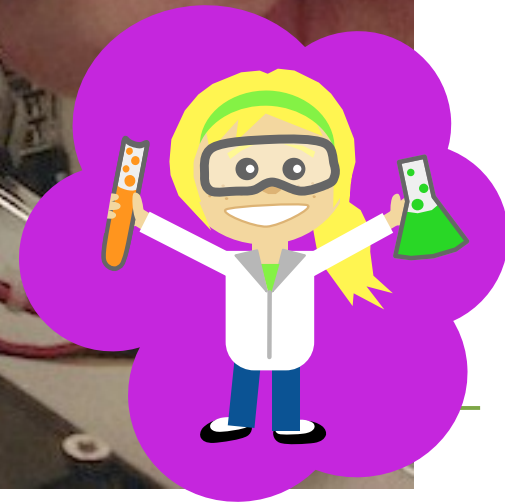
Clip



Yves



**HGST
Disk**



Software that scales up to tens of PiB



 Fermilab



... and down to a single Raspberry Pi

