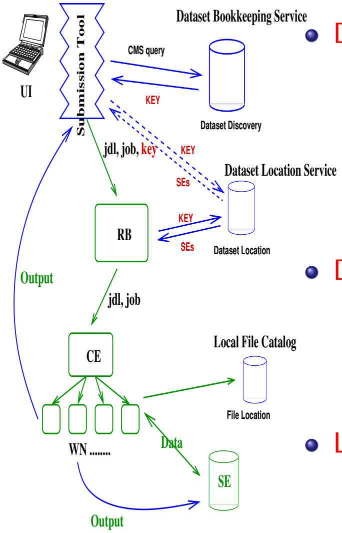
Workload Management Baseline Services for CMS

Stefano Lacaprara

Department of Physics
INFN and University of Padova

PEB Baseline Services Group, 15 april 2005

CMS Schema for WM



- Dataset Bookkeeping Service (CMS)
  - higher level, interface to physicist
  - provide query mechanism
  - output is set of *Data chunk(s)*
  - Data Chunk is an unbreakable unit (Atom). granularity defined by DM–WM (today is Dataset . . .)
- Data Location Service
  - Given key identifying DataChunk ⇒ list of SE(s) ⇒ RB get CE(s)
  - Can be done at UI or RB level
  - Use only *abstract Data*, not files!
- Local File Catalog
  - Available at local sites
  - GUID to PFN mapping

General

- CMS does not want to develop its own WMS
- We do want to use Resource Broker and LCG WMS
- Actual RB functionalities are not far from the expected need (see after)
- Performances are not yet

RB

- Performances
- Must be able to submit $\mathcal{O}(1000)$ jobs in timescale of ′(10) *s* seconds
- Must allow load balancing (if needed) and fault tolerance
- A set of RB should be available from UI and if one is down, the “next” should be selected w/o human intervention
- Data Location Service interface: RB must be able to talk directly with Data Location Service (DLS) *e.g.* via DLI (or DLI–like) interface, to query location of given data block
- Must be able to deal with DAG

Job Cluster

- Must be able to deal with Job Cluster
  - Job with same requirements, same executable, . . .
  - Data Input can change from job to job
  - Configuration can also change per job, possibly just a small part
- Bulk operation: submit, query, kill, etc . . .
- Allow also single job operation
- Job splitting: probably not baseline, but the possibility should stay open for future (hopefully near)
- User defines as input a set of input data (data blocks), and RB splits the job (at the granularity defined by data blocks) according to available resources
- For scalability reason, it would be nice if the resolution of a job cluster to a set of single jobs is dealt with at CE leven and not at RB level.

Policy and Priorities

- Today possible only at CE level
- Must be possible at VO level
- CMS must be able to define priorities according to role/group/user
- *Higgs analysis group is close to discovery, so all resources should be given to them from today to next conference date*
- As now, CE owner is allowed to define priority of use of his own resources among local user and CMS
- On top of that, CMS should be able to define priorities for all sites, or for selected sites

Other

- Interactive jobs: interesting, can be useful for debugging, but we could live without it
- Reproducibility: WMS should be able to resubmit a job to the very same resources on which have been submitted before
- For sure at CE level, possibly at WN (even if interaction with LBS can be very tricky or not possible . . .)
- Job provenance: WMS should provide all relevant information about job submission (where, when, etc . . .)
- CMS will take care of provenance of application itself
- Global (VO) monitoring: we must be able to knows how many jobs have been submitted by who to where. Not a generic user (which see just his jobs) but CMS manager (with proper role) should have a global view (including history)
- Input and output sandbox should not, under any circumstances, fill up RB disk space and kill the RB