LHCb

LHCb

The “Beauty” experiment at CERN

Initiators of DIRAC, and up to now the real maintainers

DIRAC functionalities in use: WorkloadManagement

- We use pilots 2.0
  - Available since DIRAC v6r12
    - Follow Andrei's presentation tomorrow
  - And we extended it too
    - e.g.: we "install" (setup) LHCbDIRAC from CVMFS
      - with a fallback
  - Other minor extensions in LHCb up to now
    - Mostly for the pilots 2.0
- And all the rest

DIRAC functionalities in use: DataManagement

- We use everything in it right now, including the DIRAC File Catalog (DFC)
  - We migrated from LFC a couple months ago
    - That was not THAT easy
    - Careful planning
    - We can give you our recipe
    - Main issues found: authorization

DIRAC functionalities in use: Interfaces

- We extended most of the APIs
- the LHCbJob API contains all the logic for LHCb jobs, both production and user jobs

DIRAC functionalities in use: TS

- Extensively used and extended
  - All our production activities based on it
    - Sending jobs to the WMS
    - Jobs are created using the Job API
      - So, the Workflow package
  - All our data management activities based on it
    - Sending requests to the RMS
- Productions created using a LHCb extension
  - For more info: LHCb presentation at the 3rd DUW
- Scaling this system was critical
  - One day we will need to migrate to use parametric job submission

DIRAC functionalities in use: RMS

- Used extensively, both for the jobs failover and the data management (together with the Transformation System)
  - Not extended
    - We only added 1 operation handler

DIRAC functionalities in use: Accounting

- All the accounting types of DIRAC are used
  - Extended in LHCbDIRAC with few new types:
    - Both for storage usage and jobs

- WMSHistory is a “special” accounting
  - We split WMSHistory from the rest of the accounting
    - requires DIRAC v6r12
  - Generic “monitoring” solution
    - requirements: https://github.com/DIRACGrid/DIRAC/wiki/Monitoring-service
    - See Zoltan + Adri presentation tomorrow
      - RabbitMQ, ElasticSearch

DIRAC functionalities in use: Resources

- Almost all resources developed for LHCb
  - Few extensions, mainly for LHCb Bookkeeping catalog
  - Looking for TS catalog

DIRAC functionalities in use: RSS

- Used for managing SEs
  - We have several policies defined in the CS
- Extended in LHCbDIRAC
- We created it, and up to now we maintained it
  - We are open to suggestions, and open to your help
  - There are some tasks in GitHub and more might come

DIRAC functionalities in use: Workflow

- What is in DIRAC is a porting from LHCbDIRAC
  - Workflow package
  - Well extended in LHCbDIRAC
  - All our jobs, either production or user jobs, use this package
  - The Job API use this package
  - Apparently this package has been created in some extension, but not extended directly yet
    - Up to now, it has been maintained by LHCb, any user?

LHCbDIRAC systems

- Bookkeeping system
  - A file metadata and provenance system
  - Often, the real driver for LHCbDIRAC code
  - Backend is Oracle (DIRAC supported)
- Core (utilities)
  - Just utilities for running our jobs
- Production System
  - Provides management and submission functionalities
  - A web frontend is exposed to the LHCb community for requesting productions
    - And to launch them (Production managers)

LHCbDIRAC systems /2

- LHCbVMDIRAC
  - A collection of scripts for our Virtual Machines (mostly for contextualization)
  - We are running production jobs in these VMs launched via Vac, Vcycle, BOINC
  - We don’t use VMDIRAC
  - Pilot 2.0 is the real federator

Our Clients

- End users (physicists) can submit jobs:
  - using directly the LHCbDIRAC APIs
  - using Ganga to prepare and send jobs
    - this is is what most of them use
    - Ganga then submits jobs to DIRAC using the LHCbDIRAC APIs
- Productions requesters
  - every user, in theory
    - but productions go through an approval mechanism
- Production managers, GEOCs
- Production shifters
- Data managers
- Resources (sites) managers

Production setup

- 2 setups (production and certification)
  - We removed Development
- ~30 (V)machines for running components
- 5 MySQL servers
  - 4 are provided by CERN (DBoD)
  - 1 (accounting) is on 2 (old) machines that we manage ourselves
- 1 Oracle DB for the LHCb Bookkeeping

- About 40K jobs constantly running
- Peak of ~65K jobs running

Using several types of computing resources

Still dominated by “LCG” (CREAM), others growing fast

?