Experiment plans for batch system usage

Federico Carminati

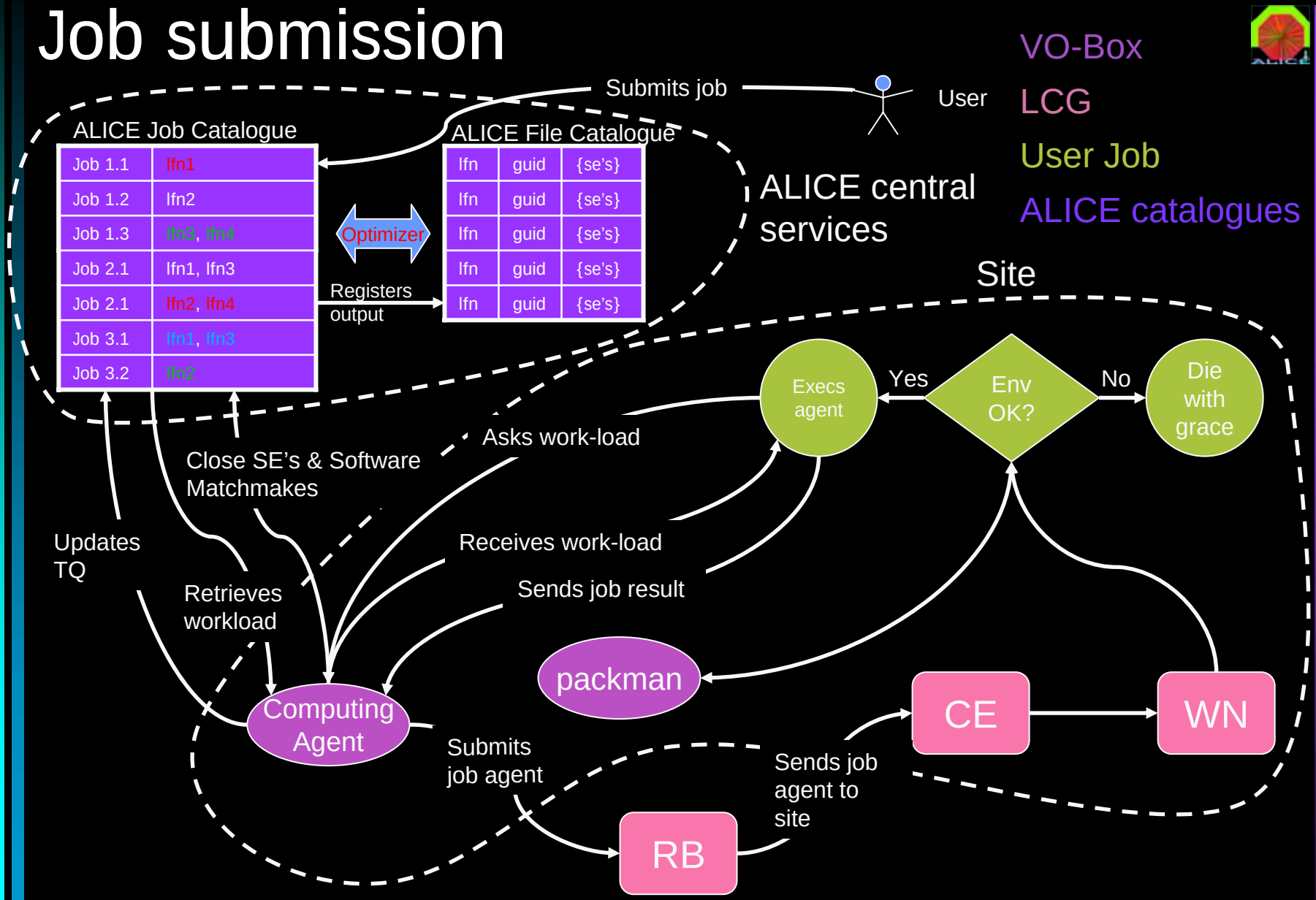
HEPiX

Rome, April 4, 2006

ALICE computing model

- For pp similar to the other experiments
  - Quasi-online data distribution and first reconstruction, calib and alignment at T0; prompt analysis @CAF
  - Further reconstructions at T1's
- For AA different model
  - Calibration, alignment, pilot reconstructions, prompt analysis@CAF and partial data export during data taking
  - Data distribution and first reconstruction at T0 in the four months after AA run (shutdown)
  - Further reconstructions at T1's
- T0: First pass reconstruction, storage of RAW, calibration data and first-pass ESD's
- T1: Subsequent reconstructions and scheduled analysis, storage of a collective copy of RAW and one copy of reconstructed and simulated data to be safely kept, disk replicas of ESD's and AOD's
- T2: Simulation and end-user analysis, disk replicas of ESD's and AOD's

Job submission



Job submission

- Job agents
  - Sent only when needed
  - Avoid waste of resources and “useless” updates of the ALICE Job Catalogue
  - Eliminate “black hole” effect
- Job location determined by the data location
- WN outbound connectivity required
  - We are working on removing this constraint
- System used for large production
  - 22,500 jobs, 540 KSi2K hours, 20TB
  - 2.5% inefficiency thanks to job agents

Batch systems use in ALICE

- Past use
  - Through AliEn – use of all flavours of batch schedulers (LSF, PBS, BQS, SGE, Condor) at many computing centres worldwide
  - Few separate queues for different jobs types
  - Job priorities handled in the central TQ
- Present status
  - Practically no direct access to batch queues: shielded by the GRID (LCG, OSG, ARC) CE
  - Middleware is increasingly ‘taking away’ the functions of the batch systems (job prioritization based on job length, queuing)
  - Fewer users submit jobs locally: ultimately all offline computing tasks in ALICE will be performed on the GRID (production, calibration, analysis), users will submit all jobs to the GRID interface

ALICE requirements

- We see the interaction with the batch systems (specific submission commands, error handling and reporting, log and output files, etc...) as part of the GRID service
  - Therefore we have no special preferences to the type of batch systems deployed at the sites
- Connected with this we still do not have a properly secured sandbox
- For that we would probably need Job Agent to
  - Start virtual machine
  - ...or start another process under different user id using glexec/sudo mechanism
- However this is not a show-stopper for us

ALICE requirements

- From practical point of view, presently we require
  - One single long ALICE-specific queue
  - Would like a uniform publishing of queue length in kSI2k•h (ultimately also a GRID function) across sites
  - Ability to guarantee the share of computing resources for ALICE
  - Ability to specify the amount of memory needed by a job
  - A minimum memory requirement of 2 Gb per core
  - Scratch space of several GB
  - A shared home directory for software installatinos etc..


"That's
all
folks!"