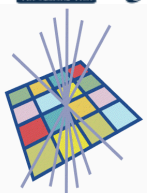
Better Scheduling Decisions

Passing additional job parameters to Batch Systems

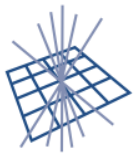
Douglas McNab (Glasgow) and Dennis Van Dok (Nikhef)

With thanks to Massimo Sgaravatto (INFN)

University of Glasgow | Department of Physics & Astronomy



ScotGrid
Scottish Grid Service



GridPP
UK Computing for Particle Physics

NIKHEF

Overview

- Why
- The Advantages
- Solutions
- How it works
- Other Uses
- Open Questions
- Conclusions

Why

- Currently these is a disconnect between what can be expressed in JDL and what a batch system submission can understand. Some may say important features are missing.
- The 'Requirements' section of JDL is only used by the WMS to match-make your requirements against what a site offers. They do not make it into the submission itself.
  - i.e. MaxCPUTime and MaxWallTime are lost once a site/queue is selected
- This not only applies to CPU/Wall times but for other possible use cases such as consumable resources /email / node properties that you wish to request for MPI etc.
  - i.e. these values could be site dependant or adopted grid wide

- By changing this users will have ability to specify more information about their jobs to allow the scheduler to make better informed decisions.

The Advantages

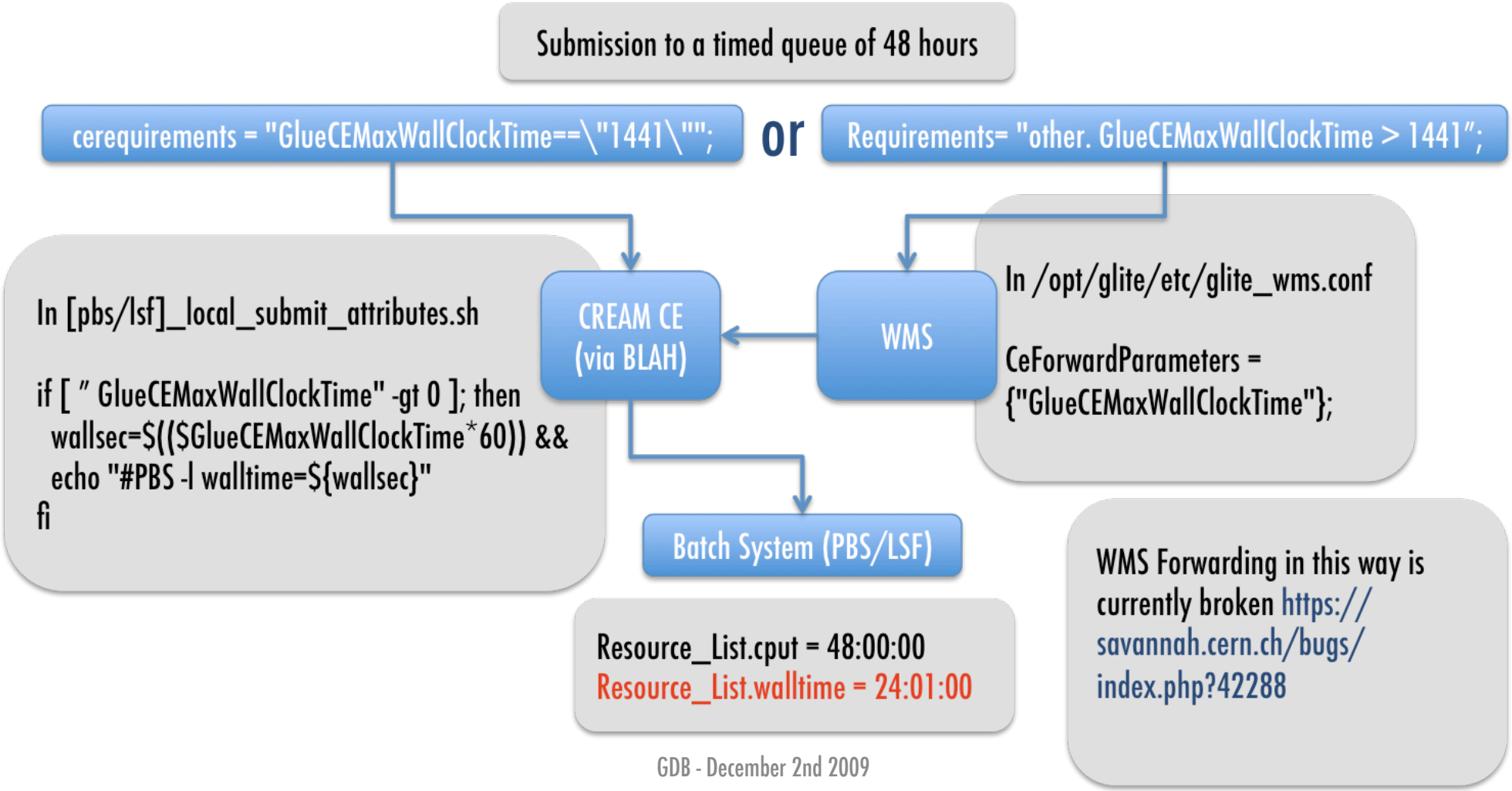
- Better Scheduling
  - Currently the queue times are used as the job CPU/Wall time
- Accurate Requests of resources
  - Be than memory, wall times or specific site resources.
- Backfill
  - With increased accuracy comes the opportunity to backfill increasing amounts of shorter jobs into free slots
    - An example might be an MPI job waiting on slots on a node to become free. Rather than leave them empty the scheduler could backfill with short jobs.

Solutions

- Hand crafted patches to the CE job managers
  - Disadvantages: maintainability
- Use CREAM's '*cerequirements*' functionality via BLAH
  - Advantages:
    - maintainability,
    - arbitrary values allowed
    - works with WMS submission (when CeForwardParameters set)
  - Disadvantages:
    - arbitrary values remain site specific
    - grid wide adoption required to provide realistic chance of use
    - WMS requires sites to support a common set of CeForwardParameters
      - currently broken due to https://savannah.cern.ch/bugs/index.php?42288

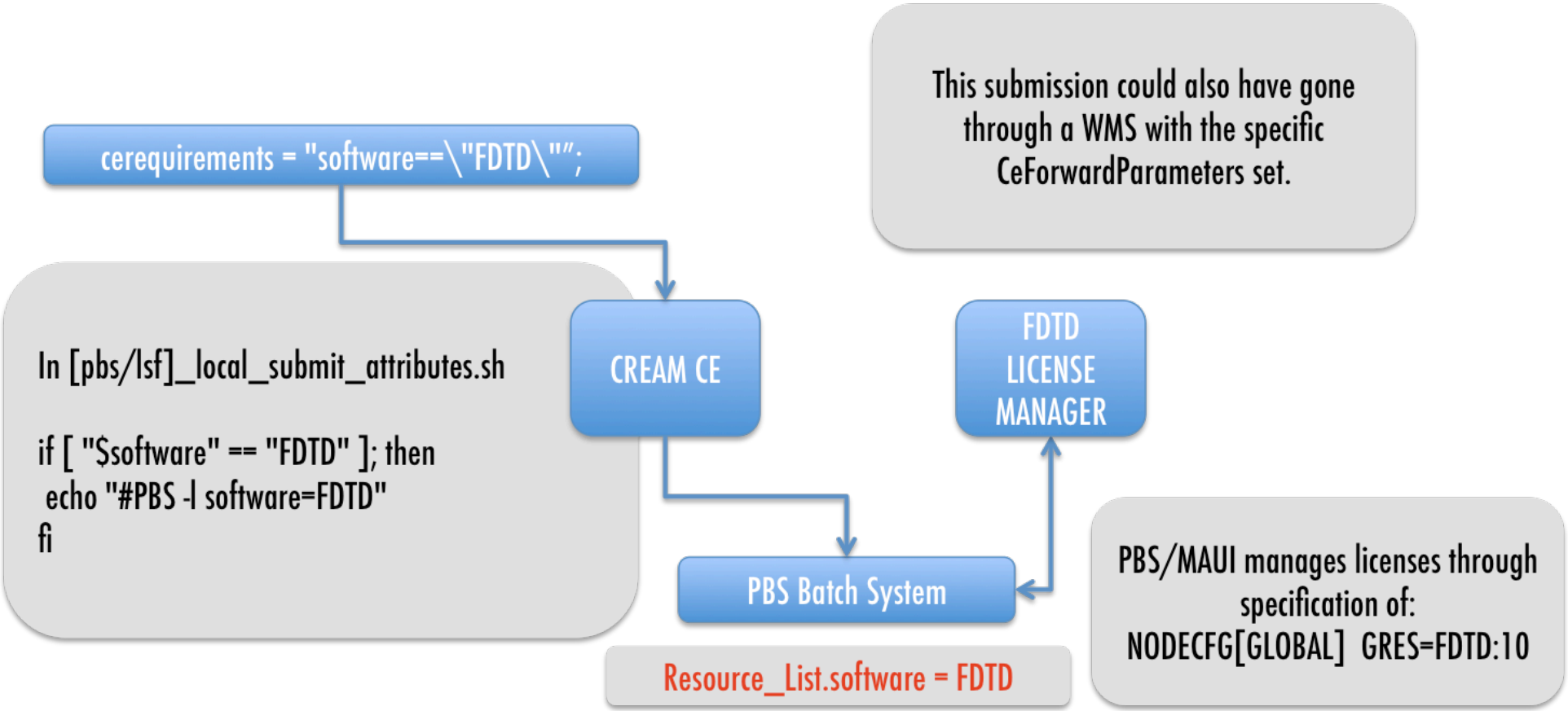
How it Works

- An example specifying a MaxWallClockTime: example from Nikhef



Other Uses

- Consumable Resources such as licenses: example from Glasgow



Open Questions

- For grid wide adoption a standardised set of attributes should be supported by all CREAM CE's.
- What set are supported?
  - Values should be qsub IEEE standard values?
    - i.e. walltime in seconds or 00:00:00
- MatchMaking requirements are a good way to make sure requested resources actually exist.
  - Since you could end up with jobs that just queue forever on the batch system never meeting any requirements
    - Requesting too much memory or resources that don't exist.

Conclusions

- Passing additional job parameters through CREAM is a promising development for users and site admins.
  - Better scheduling at sites makes more efficient use of resources.
- Standardisation required for grid wide adoption.
  - Care should be taken to follow existing standards i.e qsub
- Still useful for local sites to support custom functionality such as license management and specific resources.

Useful Links

- http://grid.pd.infn.it/cream/field.php?n=Main.ForwardOfRequirementsToTheBatchSystem
- https://savannah.cern.ch/bugs/index.php?42288
- https://savannah.cern.ch/task/?9461