GridPP DIRAC: Status and Plans

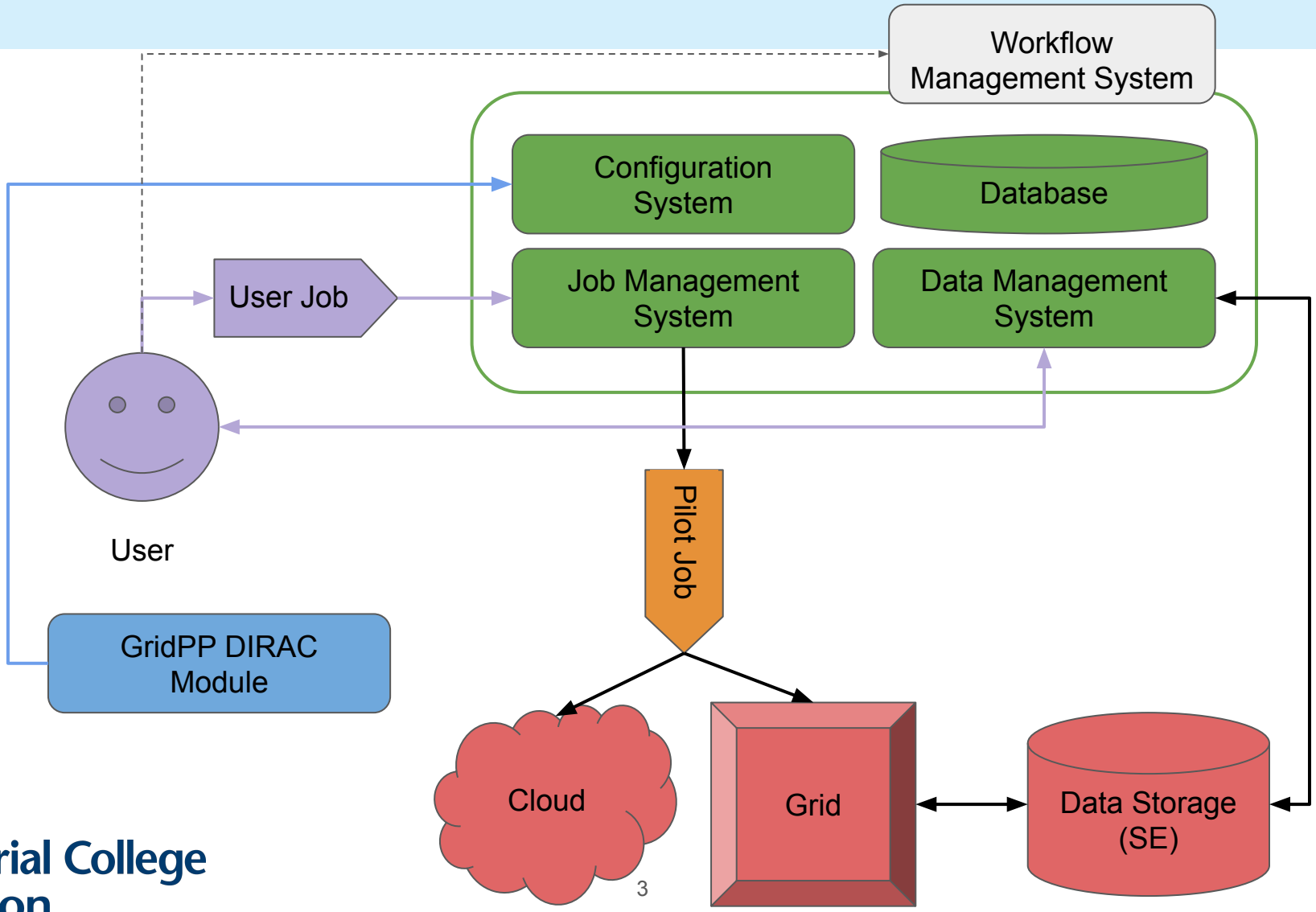
Daniela Bauer & Simon Fayer

Imperial College London

Overview

- The GridPP DIRAC service and the UK's role in the DIRAC consortium
- GridPP DIRAC: Recent projects
- GridPP DIRAC: Planned upgrades and developments
  - Major version upgrade: v7 -> v8
  - Token support in DIRAC
  - Resource discovery in a DIRAC context

DIRAC Overview

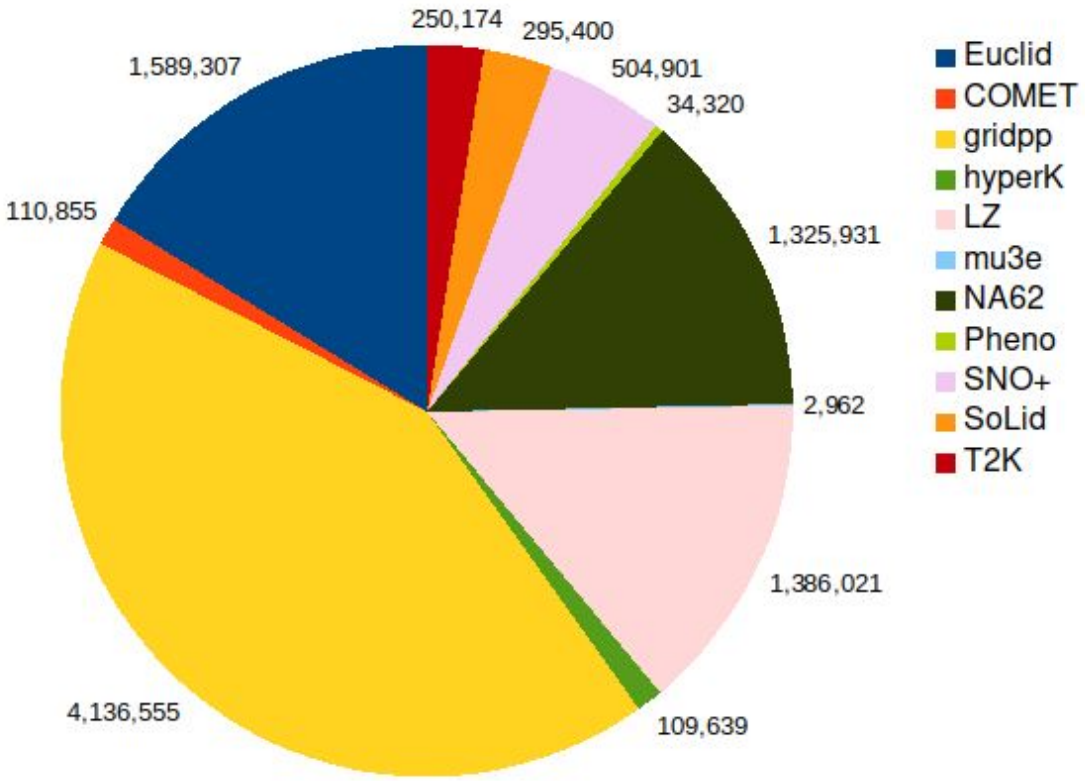


Imperial College London

Introduction

- The GridPP DIRAC instance is currently used by ten experiments to manage their workloads
  - A subset of these also uses it for data management.
- The gridpp VO is used for onboarding
- All supported experiments have UK involvement

Executed jobs by experiment: March 2022 to March 2023



Imperial College London

UK involvement in DIRAC

- DIRAC was originally developed by LHCb, but is now overseen by the DIRAC consortium of which the UK is a member via Imperial College.
- We use this to ensure that our user communities' requirements are taken into consideration.
- We regularly contribute features that are useful to our communities.
  - These are often then also used by other DIRAC instances.
- We also do (some) quality assurance for the project as a whole.

CERN

Institute of High Energy Physics
Chinese Academy of Sciences

KEK

cnrs

Imperial College London

Recent UK DIRAC projects (GridPP/IRIS/SwiftHEP)

- IRIS activities (Euclid, LZ) are driving interest in cloud usage, making this a focus for UK DIRAC developments.
  - The most recent development replaced the bespoke DIRAC cloud interfaces with a single reusable module using Apache libcloud:
    - This will lessen the maintenance burden and increase reliability
    - There will be a CHEP poster with all the details :-)
- DIRAC is a pilot job system, hence pilot logs form are a crucial tool for diagnosing failures:
  - Current pilot logging system insufficient: Limited log retention, site and technology dependent
  - New system actively sends logs at regular intervals to a central collector
  - Code nearing completion: CHEP poster :-)

Code can be found at https://github.com/DIRACGrid/DIRAC

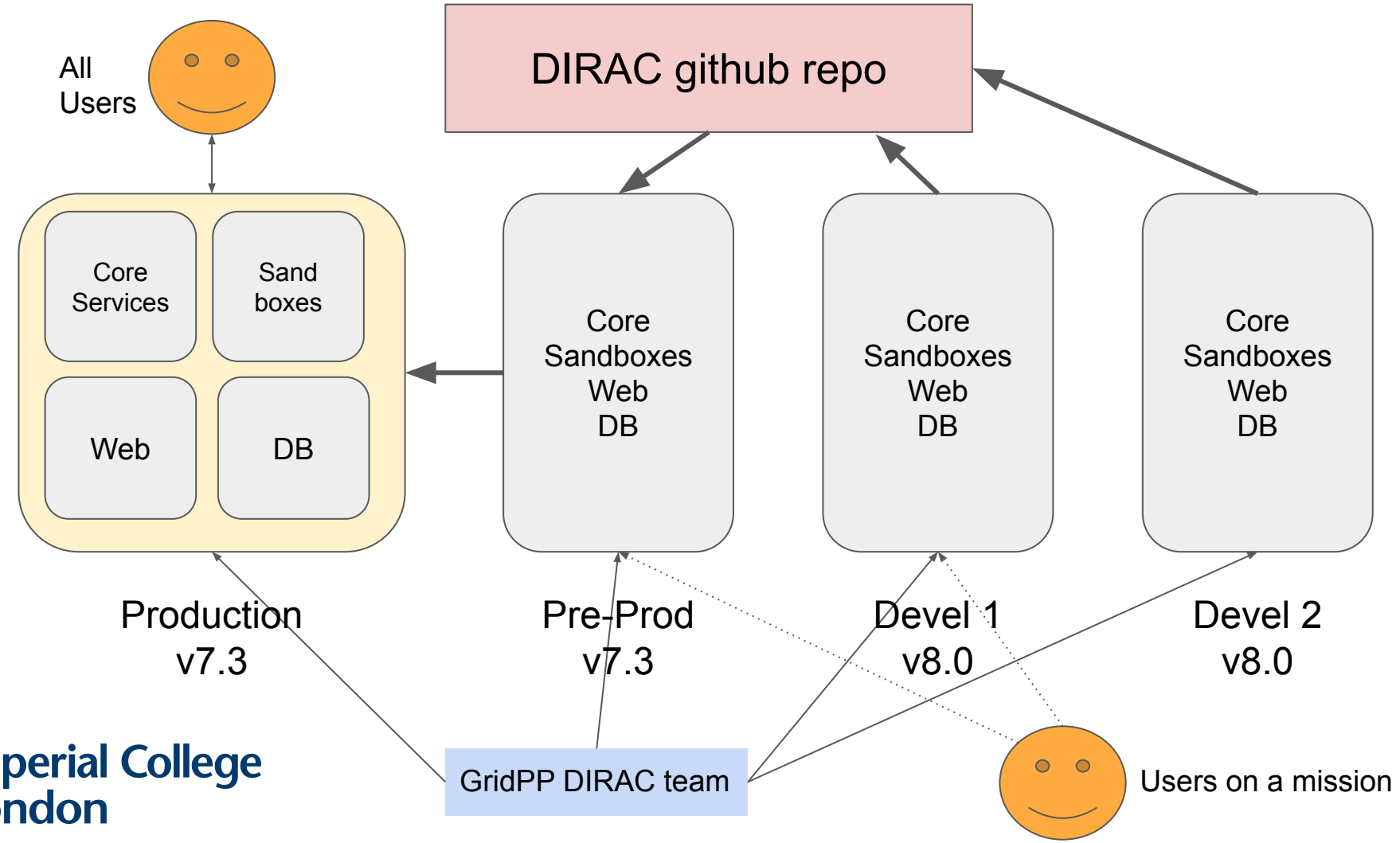
Recent UK DIRAC projects: User facing

- Simplify DIRAC command line interface:
  - Targets are small to medium communities that usually do not have dedicated user support.
  - EGI has a similar user base, but their product (‘comdirac’) had been mostly unmaintained due lack of available developer effort.
  - After a complete code revamp by the UK Team the simplified commands are now part of the DIRAC core and have already found some happy users. No CHEP poster, though.
- Feature deployment on user request:
  - Deploying a new feature in production carries a fairly large overhead:
    - New features often need initial “make GridPP compatible” modifications to avoid interfering with our current setup.
    - Requires developing monitoring tools.
    - Requires new Pre-Prod QA.
  - We can afford to be more lenient on the development servers, where the nuclear option is available:
    - This enables users to try out features before deciding whether they are useful to them.
  - Deployed two SWIFT-HEP requests for testing: Workflow management system for automated file processing and HTTP-based service interface.

Imperial College London

Need something ? Please ask.

GridPP DIRAC Setup: Production and Development



Imperial College London

The next major release upgrade: v7 to v8

New in v8:

- First release that will have token support.
- More https service support.
- New pilot logging in v8 release track only.
- Simplified user interface.
- Python 3 only.

Current target date for upgrade: July 2023

- Also deploy webdav everywhere.

Imperial College London

Tokens

DIRAC is following a model similar to CMS & Atlas for introducing token support:

- Initially only pilot job submissions are authenticated by tokens, but payload still uses X.509 proxy
- Code development behind WLCG token roadmap, but latest pre-prod releases pass initial token tests (using the wlcg VO):
  - uses Indigo-IAM
  - also looking to support EGI Check-in
  - focuses on a non-CERN infrastructure to avoid the “works for LHCb but nowhere else” pitfall
- Imperial College is one of the sites used for testing
- Aiming to keep the transition as transparent as possible to our users, in line with WLCG roadmap(*), i.e. users will be switched over last

Imperial College London

(*) see e.g. Token Transition Status at the WLCG workshop in Lancaster November 2022 or Tom’s GridPP49 talk.

Resource discovery and automated configuration in DIRAC

- DIRAC relies on the bdii for resource discovery and configuration
  - So far there is no replacement that contains the same information for all VOs:
    - CEs: type, supported VOs, operating system generation of WNs, RAM/jobslot
    - Storage: supported VOs, protocols (srm, XRootD, WebDAV) & ports, storage paths
  - With 10 VOs which do not have the computing resources of the WLCG experiments, going down the WLCG route of having the site admins email/open tickets for any change in configuration would be “not ideal™”
    - Plus, have you seen my typing ? Hand-editing config files is error prone.
  - Imperial runs their own topbdii, so CERN retiring theirs is not an issue

- DIRAC also has an inbuilt module to query the GOCDB

Resource discovery and automated configuration in DIRAC

- GridPP DIRAC currently uses an extension to configure the resources in DIRAC:
  - Fully automated for CEs, with a few (< 4) hacks for ‘special’ sites/queues etc
    - CEs are auto-deleted if we don’t see them for 2 weeks.
  - Fully automated for (most) SEs, but SEs are never deleted.
    - Special cases (you know who you are) added by hand.
  - This has worked well for the last 8 years.
- IRIS Digital Asset: Rewrite the GridPP module to use:
  - bdii
  - GOCDB
  - storage.json
- If you are wondering why this topic has two slides, as opposed to the one slide about tokens:
  - This reflects the (current) workload that is put on the GridPP DIRAC admins.

DIRAC - Future

For an overview of developments for the whole project, please see the upcoming talk at CHEP 2023:

F. Stagni *et al*: “DIRAC: current, upcoming and planned capabilities and technologies”

Conclusion

- The GridPP DIRAC instance is the workhorse for a number of non-WLCG VOs to manage their workloads.
- Project is supported throughout GridPP 7.
- We’ll keep working to ensure the GridPP DIRAC instance meets our users’ needs for the foreseeable future.