



APEL Update

GridPP37, Ambleside

2016-08-31

Adrian Coveney (STFC)

Outline

- The APEL team
- Accounting system overview and accounting records
- REBUS
- Recent developments
- Indigo DataCloud
- APEL in Indigo
- How you can help us
- Contacts and links



The APEL Team

- Adrian Coveney
 - Team Leader
 - EGI and everything else
- John Gordon
 - WLCG and EGI
- Greg Corbett
 - Newest member
 - INDIGO DataCloud, EGI Federated Cloud
- (Stuart Pullinger)
 - Career Break



Accounting System Overview

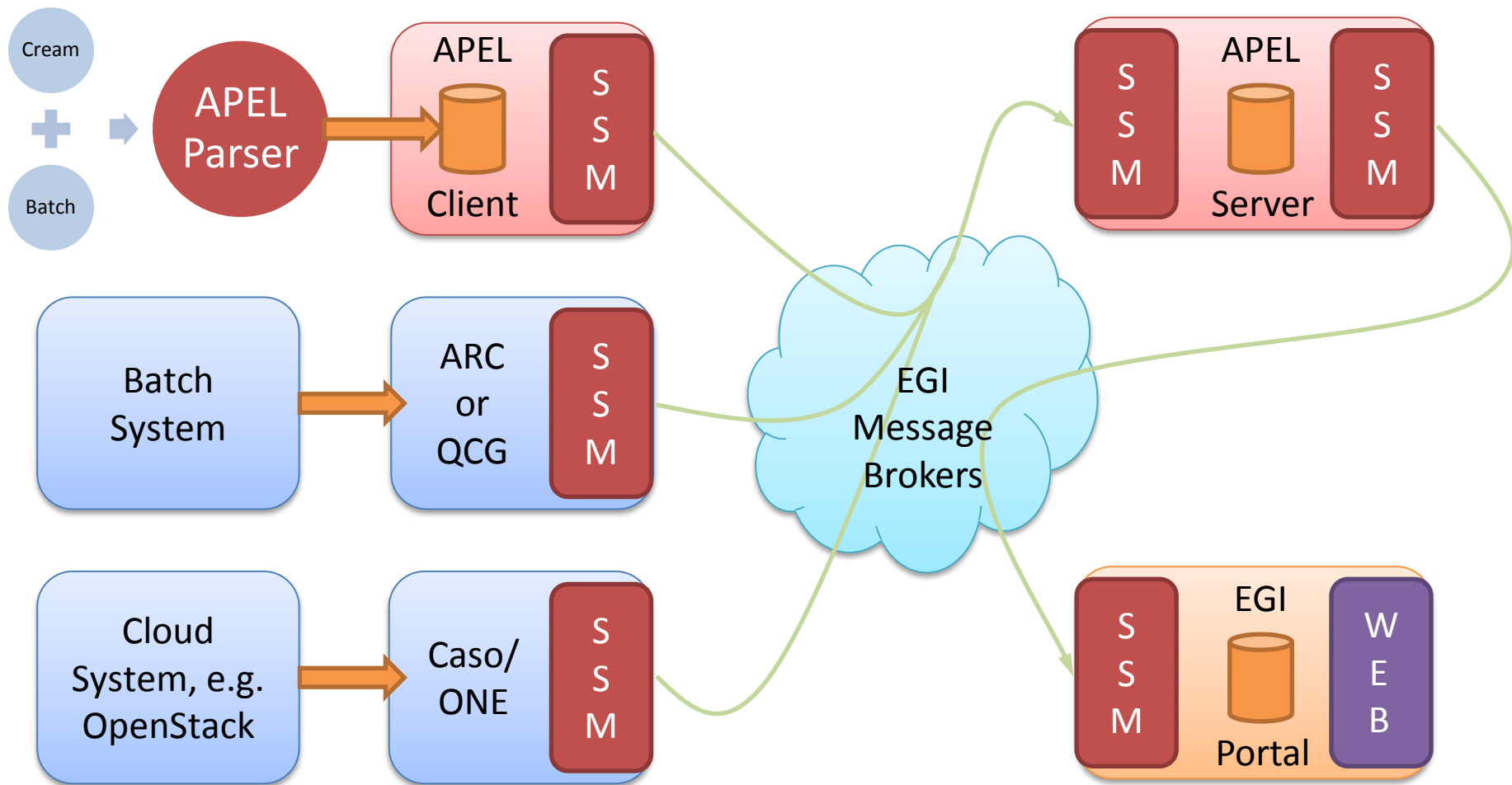
- APEL collects accounting data from sites participating in the EGI and WLCG infrastructures, and sites in other Grid organisations collaborating with EGI, including OSG, NorduGrid and INFN.
- Accounting data is collected from different sensors into a central repository, and then processed to generate statistical summaries that are available through the EGI/WLCG Accounting Portal.
- Statistics are available to view by Users, VO Managers, Site Administrators and anonymous users according to defined access rights.



Accounting records

- Job Records
 - Made from joining batch records with CE records.
- Summary Records
 - Aggregations of job records.
- Normalised Summary Records
 - As per summary records but with times normalised by benchmarks.
- Others
 - Cloud accounting/summary records.
 - Storage accounting records.
 - (Data set usage records.)





REBUS (I)

- WLCG **RE**source, **B**alance, and **U**sage
- Not controlled by APEL but it both takes data from, and provides metadata to, the Accounting Portal.
- A database at CERN which stores useful information for WLCG
- **Topology** – the definitive metadata on T1 sites, T2 federations in countries and their constituent sites. Used by Accounting Portal
- **Pledges** – WLCG resources pledged per T1 and T2 site. Input by WLCG country management (DB)
- **Capacities** – harvests data from BDII on installed capacity at sites. Physical and logical cpus, average HS06, total capacity.



REBUS (II)

- **Reports** – T1s can input their actual installed capacities per VO.
- **Trends** – plots of a few metrics as a function of time. Data taken from Accounting Portal.
- **Accounting** – Monthly WLCG reports showing usage vs pledges. Data loaded from Accounting Portal with possibility for T1s to correct/fudge their data. Reports then archived in WLCG wiki.



Recent Developments (I)

- **Storage Accounting**
 - Thank you if you've helped us test storage accounting. This has let us track down a bug at the central repository.
 - We will have fixed the problem soon and hope to roll out to more sites afterwards.
 - The fix will also simplify the addition of storage summary records at a later date.
- **Data Set Accounting**
 - Feedback received from EGI partners on first draft of metrics.
 - Looking to revise and test with a provider.



Recent Developments (II)

- WLCG Accounting Task Force
 - Generally happy with the new Accounting Portal.
 - Trying to verify data in portal from experiment data. Some success with ATLAS but also discrepancies.
 - Tool created at CERN to help automate the process of comparing data (ATLAS, ALICE, and CMS).
 - They would like unscaled wallclock but some batch systems scale the wallclock. (On long-term wishlist.)
 - CERN revamping how they report to APEL, so their data should soon be much more correct than previously.



Indigo DataCloud (I)

- **An H2020 project from April 2015 to September 2017**
- **Who: 26 European partners** in 11 European countries
- **What: develop an open source Cloud platform** for computing and data tailored to science.
- **For: multi-disciplinary scientific communities**
- **Where: deployable on hybrid (public or private) Cloud infrastructures**
- **Why: answer to the technological needs of scientists** seeking to easily exploit distributed Cloud/Grid compute and data resources.



Indigo DataCloud (II)

How the Service Works

- Top level - users submit “jobs” via TOSCA files
- Middle level determines what VMs, Docker Containers and Storage is needed
- Lowest level is the underlying infrastructure to instantiate these resources



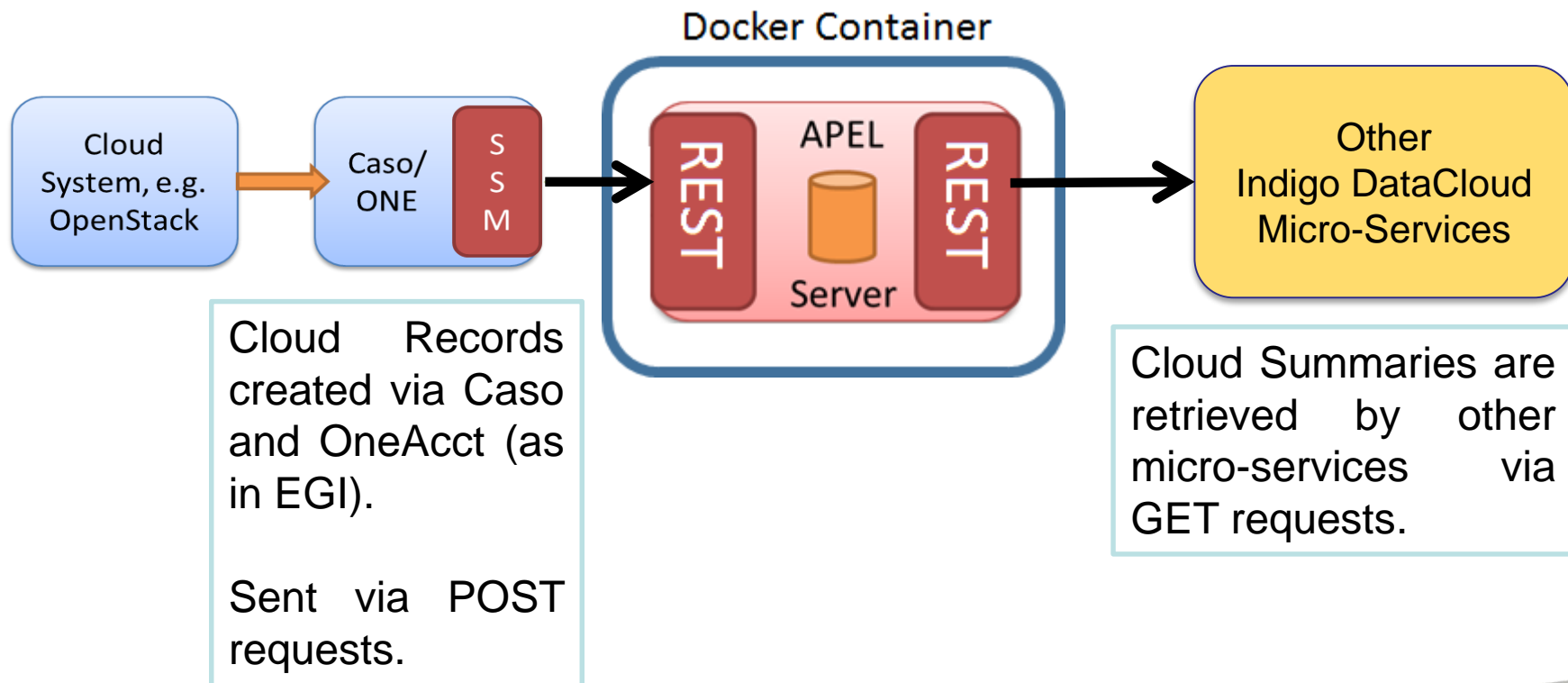
Indigo DataCloud Partners

APEL in Indigo (I)

- APEL will account for the underlying resource usage
- Developments undertaken
 - (Read + Write) REST interface to an underlying APEL Server
 - Summarises per day rather than by month
 - Packaged as a Docker container, deployable on Kubernetes
- Future
 - Support for cloud v0.4 records
- No changes to EGI Infrastructure



APEL in Indigo (II)



How you can help us

- Checking if batch systems change their output formats
 - Usually no notice or documentation.
 - When sites test a new release of a batch system, can they also check that the parsers still work and produce sensible results.
- Sending summaries instead of individual job records
 - More efficient in terms of network traffic.
 - Less load on the central database.
 - Also protects against individual job records going missing in transit as summaries are sent for the current and previous month, every day.



Contacts and Links

- GGUS Support Unit – APEL: <https://ggus.eu/>
- GitHub: <https://github.com/apel>
- Email: apel-admins@stfc.ac.uk

- New Accounting Portal: <https://accounting-next.egi.eu/>

- CERN comparison tool: http://wlcg-mon-dev.cern.ch/dashboard/request.py/siteview#currentView=ACCOUNTING_VALIDATION





Thank you.