

ButlerWG Meeting 2017-08-16

Attending

- [Tim Jenness](#)
- [Unknown User \(xiuqin\)](#)
- [Michelle Gower](#)
- [Russell Owen](#)
- [Dominique Boutigny](#)
- [Simon Krughoff](#)
- [Jim Bosch](#)
- [Brian Van Klaveren](#)
- [Unknown User \(pschella\)](#)
- [John Swinbank](#)
- [Kian-Tat Lim](#)
- [Unknown User \(npease\)](#)
- [Unknown User \(kooper\)](#)

Agenda

- Overview of data access/databases/file-storage-systems in operations.
- Specify/split up and group/consolidate use cases.

Overview

- [Michelle Gower](#) presented an overview of the systems used in operations. These are primarily:
 - Operation production (archive, raw files, metadata in databases, etc.)
 - Operation L1
 - Release (qserv, read-only that external users query, provenance, metadata, tables, etc.).
- [Michelle Gower](#), [Jim Bosch](#), [Brian Van Klaveren](#) contributed additional data access systems:
 - Developer systems (subsets of `Operation production` used for developing and debugging, etc.)
 - Personal databases (accessible from Jupiterlab, live in qserv (probably the same instance as `Release` to enable cross-joins))
 - [Brian Van Klaveren](#) mentioned that creating and ingesting data into these user databases is not done by the user itself)
 - External local pre-cache
 - External remote on-the-fly
 - External large local (e.g. DESC@NERSC, HSC@Japan)
- [Jim Bosch](#) then suggested to go through the use cases and decide for each of them for which of the above systems it was relevant. A case from [Simon Krughoff](#) was selected as an example: "A science user sees a bad coadd and wants to find out what went wrong by inspecting the full chain of things that went into this coadd." This was further specified into:
 - 1.a Find inputs into coadd (warps, calexp, etc)
 - 1.b Get files into my sandbox (exists on disk in DB)
 - 1.c Get intermediate files that were used to generate this product ([Brian Van Klaveren](#), [Michelle Gower](#) does this involve regenerating data on-the-fly?)
 - 1.d Inspect files by eye.

Additional discussion

- Brief discussion about need to read/write e.g. panda's datatable into user database ([Jim Bosch](#) yes, [Brian Van Klaveren](#) no).
- Discussion between [Brian Van Klaveren](#), [Tim Jenness](#), [Jim Bosch](#) and [Michelle Gower](#) about being able to reconstruct the qserv database from files on disk. Plan is to be able to do that (so qserv is not ultimate reference?).
- [Kian-Tat Lim](#) and [Jim Bosch](#) talked the about availability of qserv instances for development.
- [Brian Van Klaveren](#), [Unknown User \(xiuqin\)](#) and [Jim Bosch](#) discussed the spectrum between regular users (expected to use IVOA interfaces to access data vs advanced users that will use the butler).

Conclusion and task for next meeting

- Go through each use case and:
 - split it up into specific steps,
 - then mark each step based on relevance to each data access system (using template matrix).