k8s pre-GDB

A. Forti

GDB

11 December 2019

Motivation

- k8s not part of WLCG infrastructure officially yet
- Interest is growing
- Several activities are ongoing in different groups
- Similar efforts but no direction
- Not a community effort yet

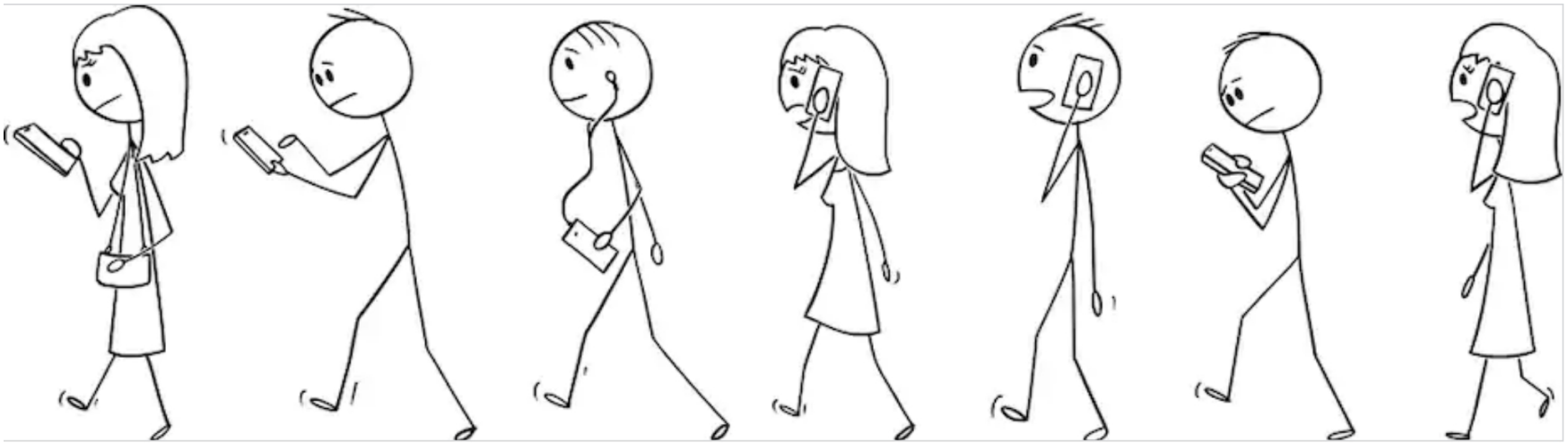
Themes from BoF @CHEP

1) Analysis facilities
   * as a batch system including federated clusters and schedulers
   * jupiter hub/binder hub/reana/kubeflow
   * submission from the experiments
   * easier access to alternative architectures
2) Service deployment
   * large scale service deployment: rucio, cms webservices, htcondor, other services
   * centralised remote deployment (slate)
3) k8s deployment and operations
   * on openstack
   * on baremetal
   * on a commercial cloud
4) Security
   * k8s security in general
   * k8s security for centralised remote deployment
   * AAI (x509, tokens, VO SSO...)
5) Storage integration
6) Image distribution (not strictly k8s but still)
   * registries/registries caches
   * cvmfs snapshotter
   * cvmfs plain cvmfs
7) CNCF landscape (if we want to interact with the community we need to know the landscape https://landscape.cncf.io)
8) k8s CRI and use of different runtimes

BoF google doc

A very long day

- 20+1 contributions
  - 4 time zones
  - Up to 60 participants
    - ~30 locally and 30 on vidyo
  - At 8 pm still 30 people
- Different projects
  - PRP
  - IRIS-HEP
  - WLCG sites
  - CERN-IT
  - ATLAS & CMS



Presentations

- Presentations can be grouped in few categories
  - Remote installation and maintenance of services: 6
    - 3 (IRIS-HEP), 2 CMS, 1 ATLAS
  - Local installation and maintenance of services: 6
    - 4 sites, 2 ATLAS/CMS
  - Using k8s as a batch system and multi cluster: 3
    - 1 T2, 1 CERN, 1 ATLAS
  - Image distribution using CVMFS: 2
    - 1 CERN-IT, 1 CVMFS
  - User perspective or current usage: 2
    - 1 ATLAS, 1 PRP
  - CNCF research group: 1s

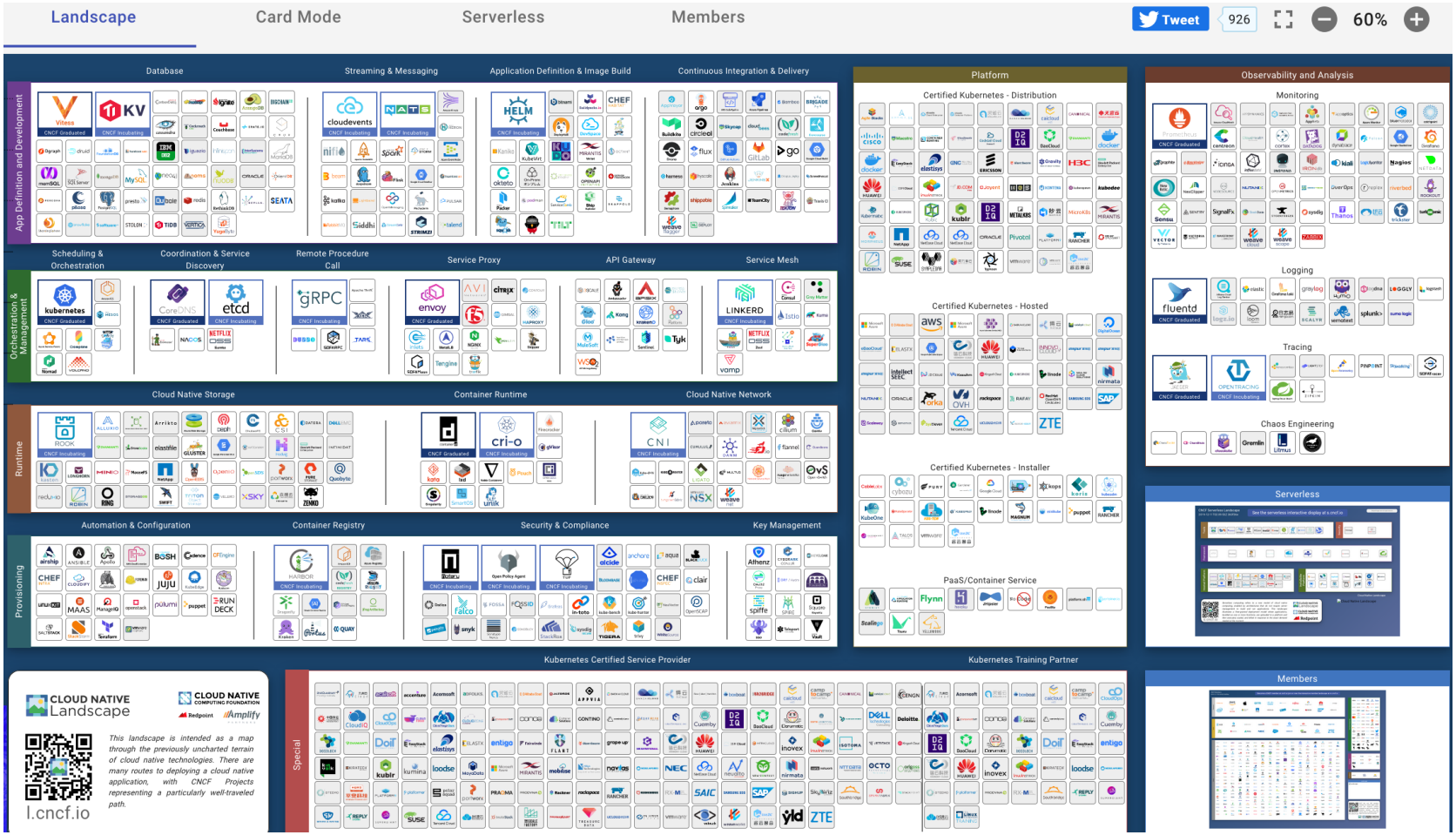
My notes

- Documentation/Training
- Can we do it?
  - Using k8s as a batch system
  - Traceability
- Different models of centralised deployment
- Distribution cvmfs but larger problem than k8s
- Common images and helm chart repo
  - Image content tracking
- Where's Europe?

- AAI → openID/tokens → usage skyrocketed
  - Can we add more tests to doma tpc ones?
- Common calls (ssl monthly call can EU site participate?)
- Completely different trust model
  - Slate/dodas/prp? participation to wlcg edge services wg
- Lot of replication of effort
- Cooperation between experiments and CERN-IT
- Need a WG

Docs, Training, Recommendations

- Filtering & recommendations
  - Landscape is huge often with several competing products changing fast



Docs, Training, Recommendations

- Filtering & recommendations
  - Landscape is huge often with several competing products changing fast
- Docs & Training
  - Learning curve is steep, documentation varies
  - Missing dummy examples on how to setup k8s toy cluster
  - Further examples on how to evolve
  - Changing configuration tools
    - Seems yesterday we moved from YAIM to puppet ;-)
    - Puppet → Helm
    - Puppet → kubespray (?)

Image distribution & CVMFS

- 2 containerd solutions to use CVMFS and avoid download everything from a registry
  - Particularly for users aim is to be able to use CVMFS for common layers and get only the user layer from the registry
  - Need to converge and cooperate on a common solution
- This is a long standing problem also for other types of container runtimes
  - Singularity also has different solutions being implemented either to use cvmfs or squids in front of a registry
- Benefits from common work.

Common infrastructure

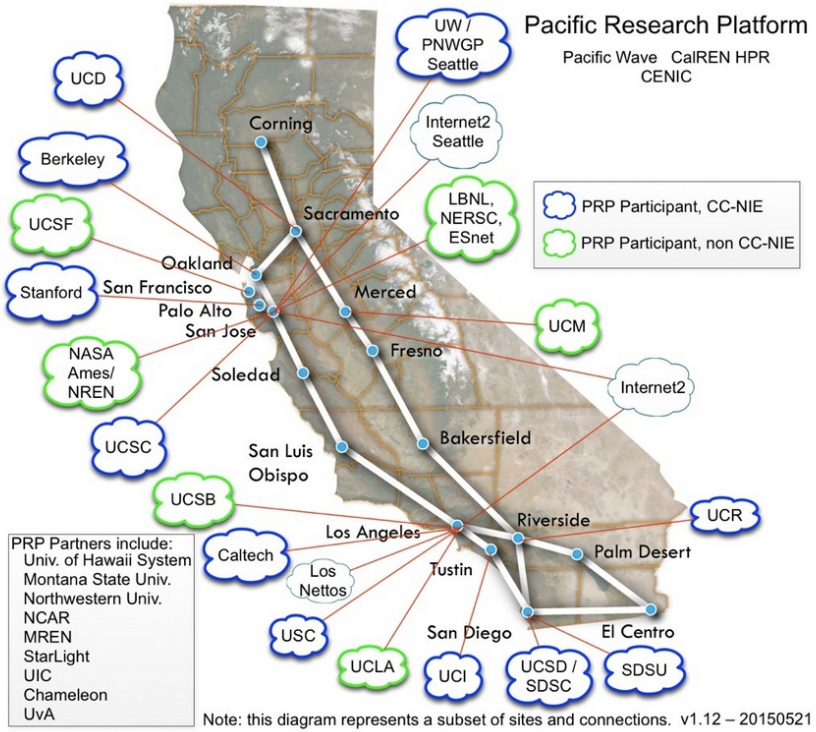
- Registries and configuration tools common repos
  - Nothing new we do it also for puppet at least for non confidential code
- Still images and Helm charts are slightly more than puppet modules and might need sanitising
- There are tools to do automated scanning and secrets can be isolated
- But the infrastructure would need to be agreed and built
  - Or we need to agree on a space on public repositories and how to maintain them and protect them
- Scanning images is not only a k8s problem

Not only k8s

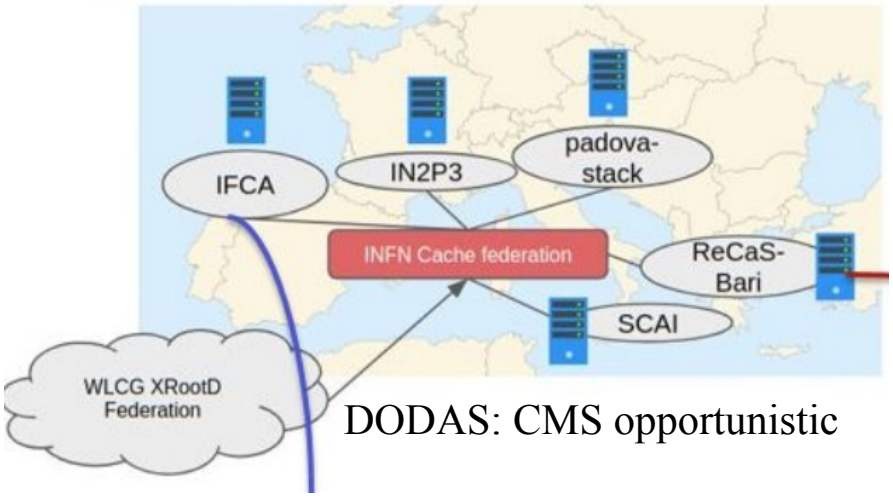
Centralised installation of services

- Depending on the model
  - Hardware owned by project managed remotely
  - Hardware owned locally needs access from external project
- Simplifies installation and maintainance of complicated services
  - Local people might need knowledge of k8s but little else
- Raises a lot of questions about security and trust model
  - There is already a WLCG WG about this started by SLATE
    - SLATE not the only one all projects that do install services at sites should participate
    - WG also dominated by US sites European sites need to participate too
    - WG charter

Centralised installation doesn't need to be global

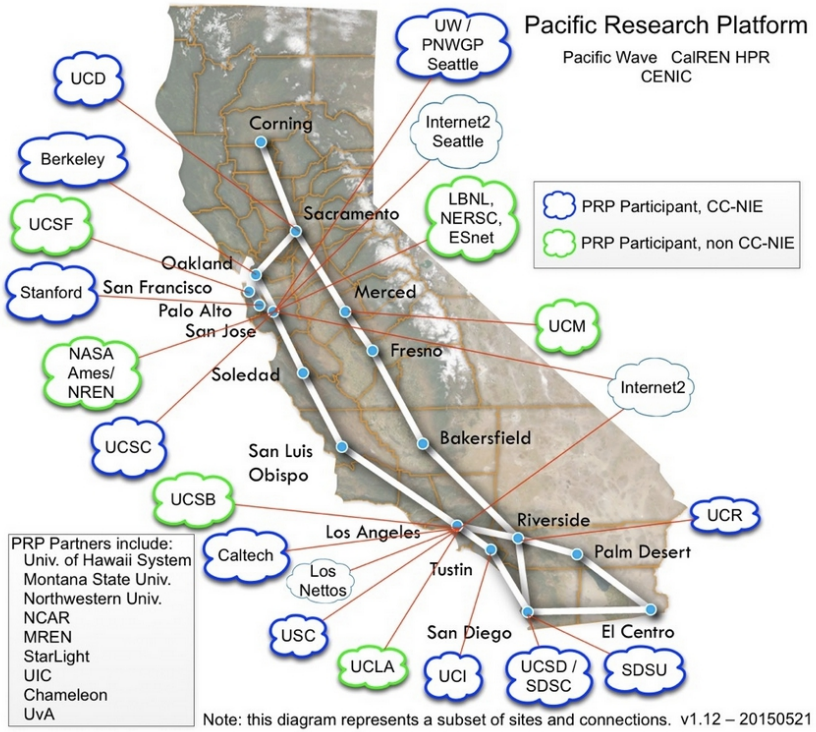


PRP (Pacific Research Platform)

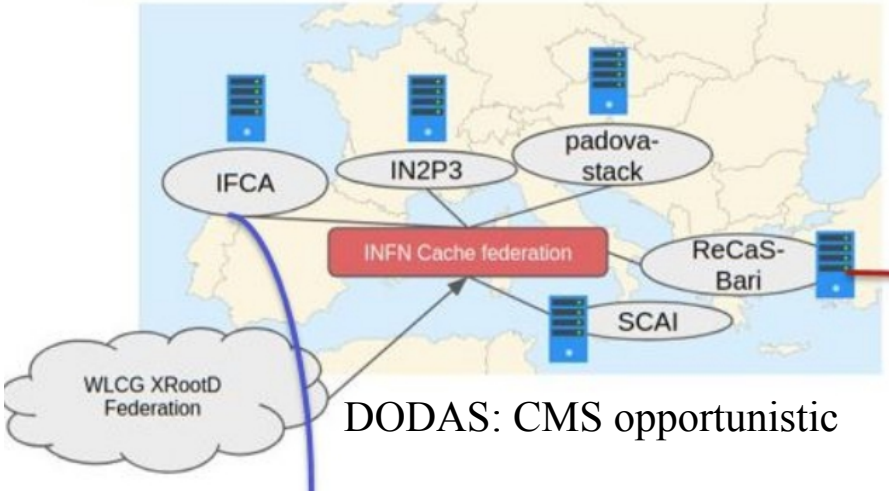


DODAS: CMS opportunistic

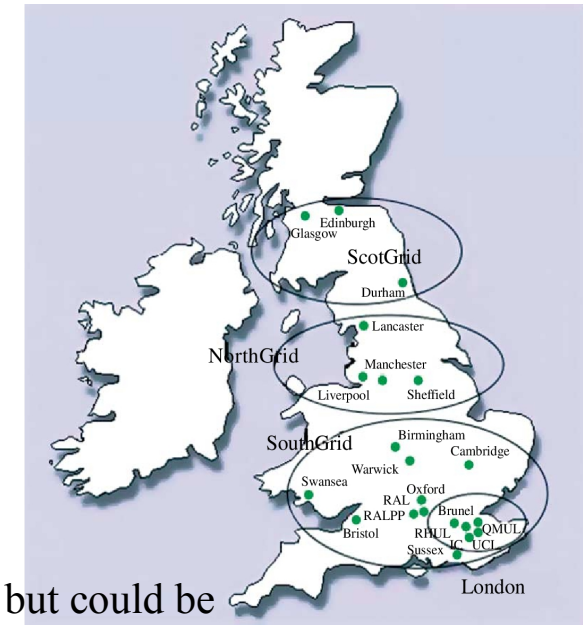
Centralised installation doesn't need to be global



PRP (Pacific Research Platform)



DODAS: CMS opportunistic



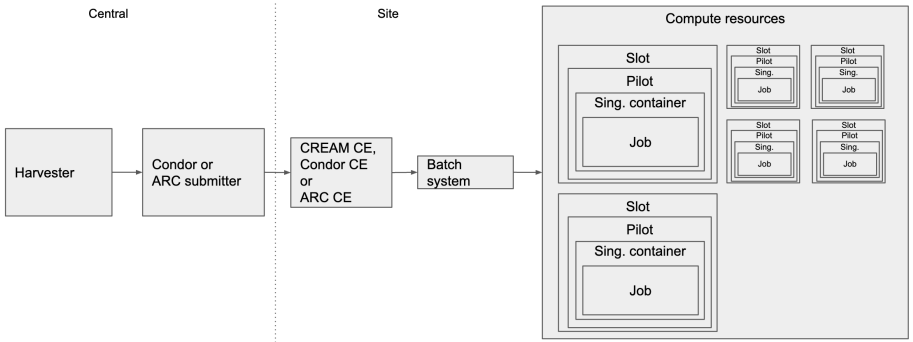
National projects

UK hypothetical but could be any other country

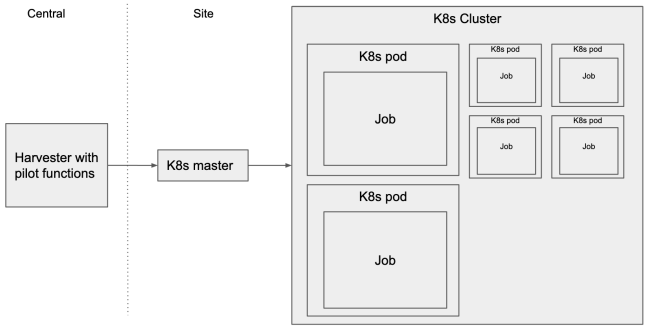
k8s as a batch system

- k8s can do resource management very well
- Still needs a lot of development to have some of the features that we take for granted
  - Multi-tenancy and fair shares
- Reasons to do this
  - Still potential to simplify a lot some of the experiment infrastructure by using some of the native functionality

Layers for ATLAS grid/batch setup



Ideal K8s setup



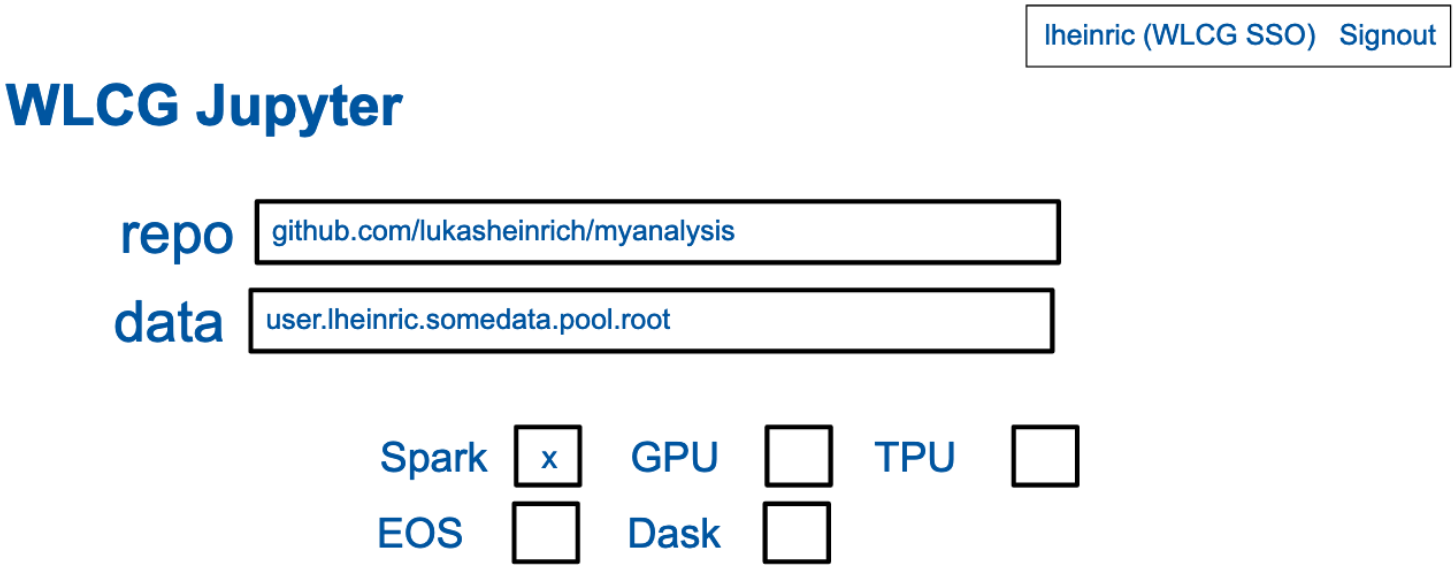
k8s as a batch system

- k8s can do resource management very well
- Still needs a lot of development to have some of the features that we take for granted
  - Multi-tenancy and fair shares
  - Traceability techniques might have to be relearned
- Reasons to do this
  - Potential to simplify a lot some of the experiment infrastructure by using some of the native functionality
  - Spill over cloud resources seamlessly without using custom made tools, but using native functionality
  - Integrating analysis infrastructure resources

Analysis Facilities

- Two types of services mostly required
  - Local batch system
  - Jupiter hubs
- Jupiter hubs handled by k8s can be also seen as an alternative to more classical batch system k8s still queues jobs even if the hub is interactive
- A more futuristic vision is to have federated jupiter hubs accessible using a federated identity
  - Components to do this are already there

Analysis Facilities



data requirement and service requirement

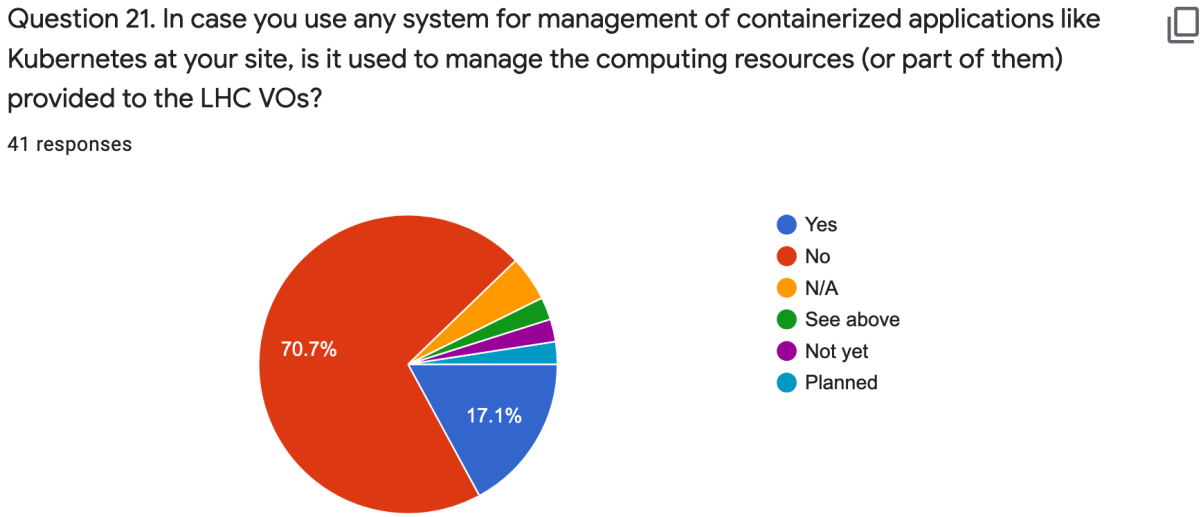
What the grid would look like if it was designed today?

AAI

- AAI repeatedly came up
- People have to do some gymnastics to integrate x509 the way we use it
- Several problems would be resolved by moving to openID Connect
- Work on using tokens already ongoing in DOMA TPC
  - This should be even more straightforward as a test case
  - Maybe we can add another testing activity?
- PRP cluster use “skyrocketed” with openID connect
  - Really easy for users to access the resources

Where's Europe?

- Feeling is that most of the work and the drive come from CERN and the US.
  - That is almost correct
- There is some effort also in Europe but driven by single institutes and not particularly visible in WLCG
  - No pressure from WLCG or experiments
  - Italy, France & Spain
  - Need to get sites interested too



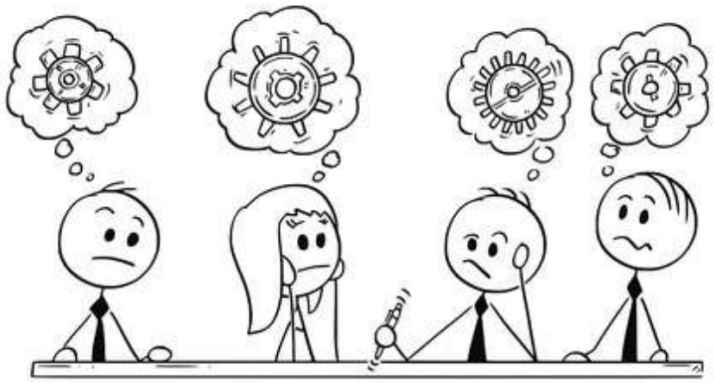
Interacting with the k8s community

- Attending kubecon and submitting presentations
  - But also participating in SIG and WG there not only internally
- For example for people interested in the development of k8s as a batch system for example can subscribe to
  - CNCF Research group
  - (check slides)

While we organise work internally... but also future work.

Outcome

- General consensus is that k8s is a useful development on multiple levels
- Work needs a more common direction
  - Point of reference
  - Documentation&training
  - Security review
  - Common work
  - Common Infrastructure
  - Interaction with k8s community
  - Avoid replication of effort
    - though some maybe useful to test different solutions



WG proposal

- No Ƀevolution
- Play & Evolve
- Open ended WG
  - Common Infrastructure
  - Experiments ↔ CERN-IT
  - Sites ↔ Experiment ↔ Site
  - Development ↔ k8s community
  - US ↔ Europe ↔ CERN
- DOMA model (?)
  - Sub groups mapped on areas of interest