

Cloud Computing Wrap-Up

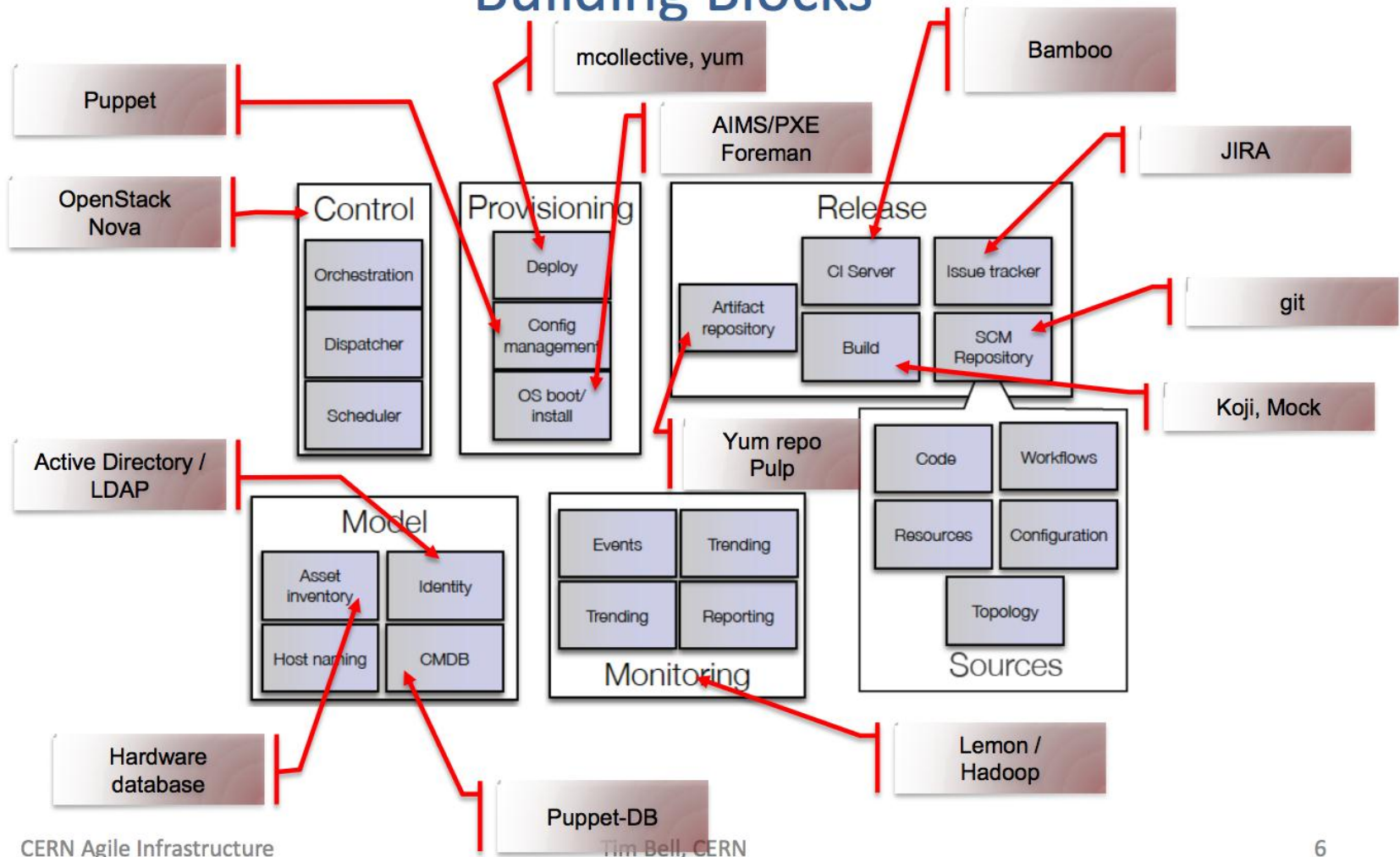
Fernando Barreiro Megino (CERN IT-ES)
Sergey Panitkin (BNL)

Summary of presentations by Tim Bell,
Doug Benjamin, Michael Ernst,
Joanna Huang & Ryan Taylor



- **HEP is no longer a special case for compute**
- Several sites/national projects are starting to adopt a tool chain model of existing open source tools and build cloud instances to expand their computing resources
 - **Ease of use, flexibility for facility management and VO usage**
- Renting out CPU/storage from commercial providers still prohibitive, but we get frequent possibilities to spill over on external resources
 - Research clouds: FutureGrid, Stratuslab
 - Projects/grants on commercial resources: HelixNebula, Amazon, Google Compute Engine
 - Hopefully experiment HLT farms
- Several facilities from all levels (T0/1/2/3) have presented their plans to move or expand towards a cloud infrastructure during the jamboree

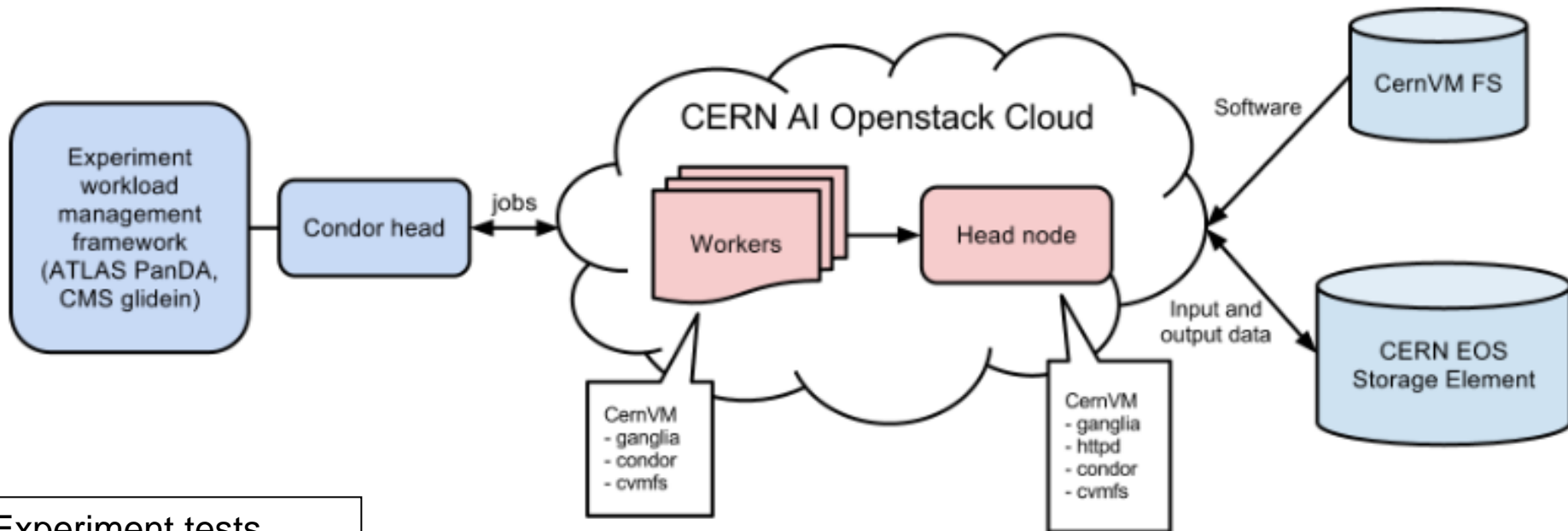
Building Blocks



- Production Preparation
 - High availability of OpenStack
 - Monitoring
- Ongoing extensions
 - Usage accounting
 - Quota management
 - External block storage (e.g. Gluster or NetApp)
- Plans
 - Production service in Q1 2013 in Meyrin data centre
 - Extension to Budapest start of 2013: aim to run 90% virtualised services
 - Target: 100K - 300K VMs on 15K hypervisors by 2015
- Migration
 - Existing services will be increasingly run on virtual machines
 - Legacy tools (e.g. Quattor) to be replaced over the next 2 years
 - CERN Virtualisation Infrastructure (CVI) and Lxcloud test bed will be migrated to the CERN Private cloud



- New centre in Budapest
- Additional 2.7MW usable power
- Hands off facility
- 200Gbit/s network to CERN

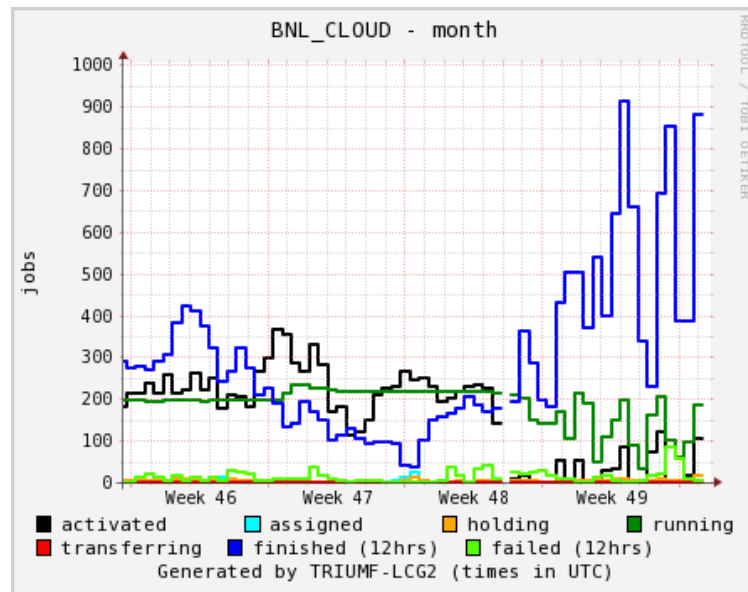


Experiment tests started end of October. Successful setup after ~2 weeks, including training of technical student (Katarzyna Kucharczyk)

6139 jobs. Click job number to see details.
States: defined:[1](#) activated:[1399](#) running:[831](#) holding:[92](#) transferring:[436](#) finished:[3282](#) failed:[17](#) cancelled:[81](#)
Users (4): c.gwenlan1@physics.ox.ac.uk:[1004](#) gangarbt:3245 sascha.mehlhase@cern.ch:[235](#) wolfgang.ehrenfeld@desy.de:[165](#)
Releases (5): Atlas-16.6.5:[757](#) Atlas-16.6.7:[835](#) Atlas-17.0.4:[768](#) Atlas-17.2.2:[2775](#) Atlas-17.2.6:[1004](#)
Processing types (3): gangarobot-pft:[2488](#) gangarobot-pft-trial:[757](#) simul:[2894](#)
Job types (2): managed:[2894](#) prod_test:[3245](#)
Task ID (4): [1031863](#):#[235](#) [1081760](#):#[1028](#) [1081764](#):#[627](#) [1087439](#):#[1004](#)
Transformations (3): AtlasG4_trf.py:[5382](#) Evgen_trf.py:[529](#) Reco_trf.py:[228](#)
Working groups (3): AP_Higgs:[1890](#) AP_Susy:[1004](#)
Creation Hosts (6): voatlas110.cern.ch:1462 voatlas111.cern.ch:1432 voatlas167.cern.ch:1843 voatlas284.cern.ch:539 voatlas284.cern.ch:539 voatlas284.cern.ch:539
Sites (1): CERN.OPENSTACK_CLOUD:[6139](#)
Regions (1): CERN:[6139](#)
Clouds (2): CERN:[3245](#) DE:[2894](#)

PanDA production jobs in 12 hours

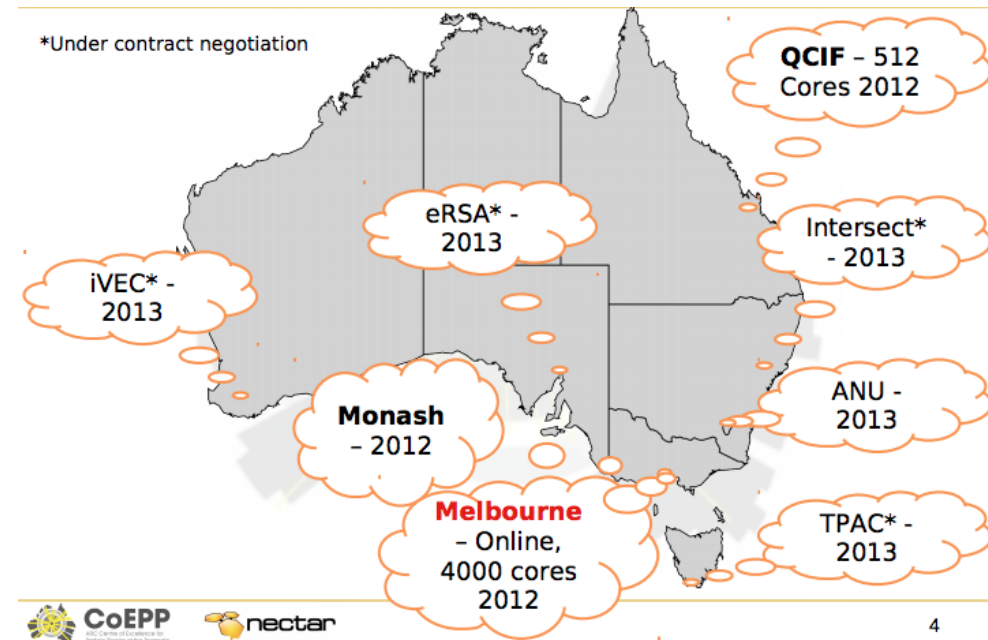
- Open Source framework selection in BNL
 - OpenStack
 - Puppet
 - Boxgrinder
- BNL_CLOUD: Standard production PanDA site
 - WAN I/O: queue can be extended transparently to Amazon or other public academic cloud
 - Steadily running ~200 prod jobs on auto-built VMs for months
 - HC tests, auto-exclude enabled
 - Performance actually better than main BNL prod site
 - Ran hybrid Openstack/EC2 cluster for a week
- US cloud resources to grow in the near future
 - BNL to 1k – 2k VMs in the next months
 - Cloud resources at 1 US Tier-2 in Q3/4 2013
 - Prerequisite
 - Scalable network architecture
 - Install everything programmatically

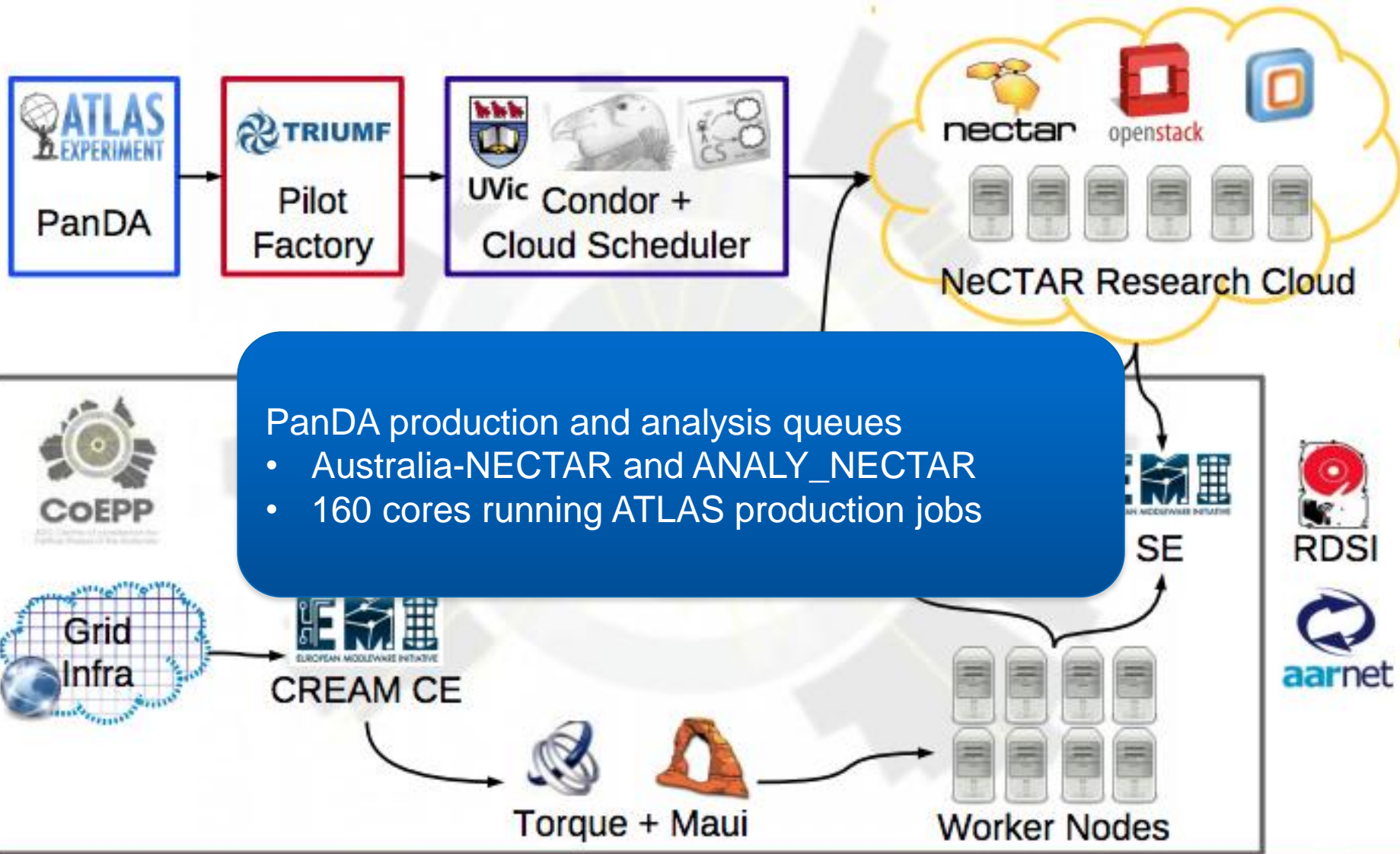


BNL_CLOUD faster than BNL_CVMFS_1 (wallclock 1740s vs. 1960s)

- Hammercloud (ATLASG4_trf_7.2...)
- Setup time (no AFS)? No shared filesystem?
- Similar spread for other tests (e.g., PFT Evgen 16.6.5)
- Anticipate using Iljia's HC framework to conduct more tests

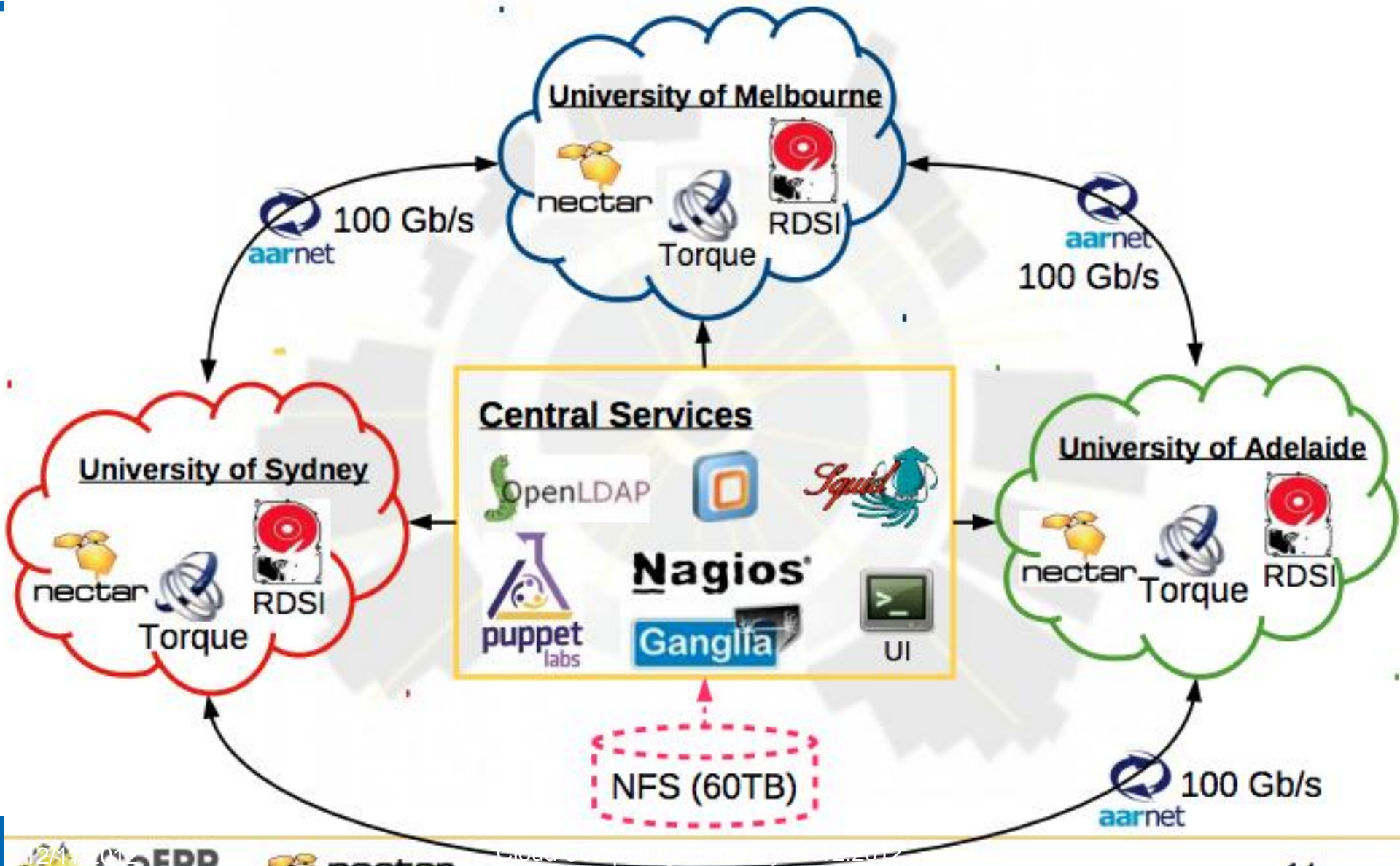
- \$47M Australian government funded project
 - Build new infrastructure specifically for Australian researchers
 - Augment Australian ATLAS Tier 2 capacity
 - Build federated Tier 3 for high throughput data analysis
- 25,000 core IaaS setup spanning 8 locations and based on OpenStack



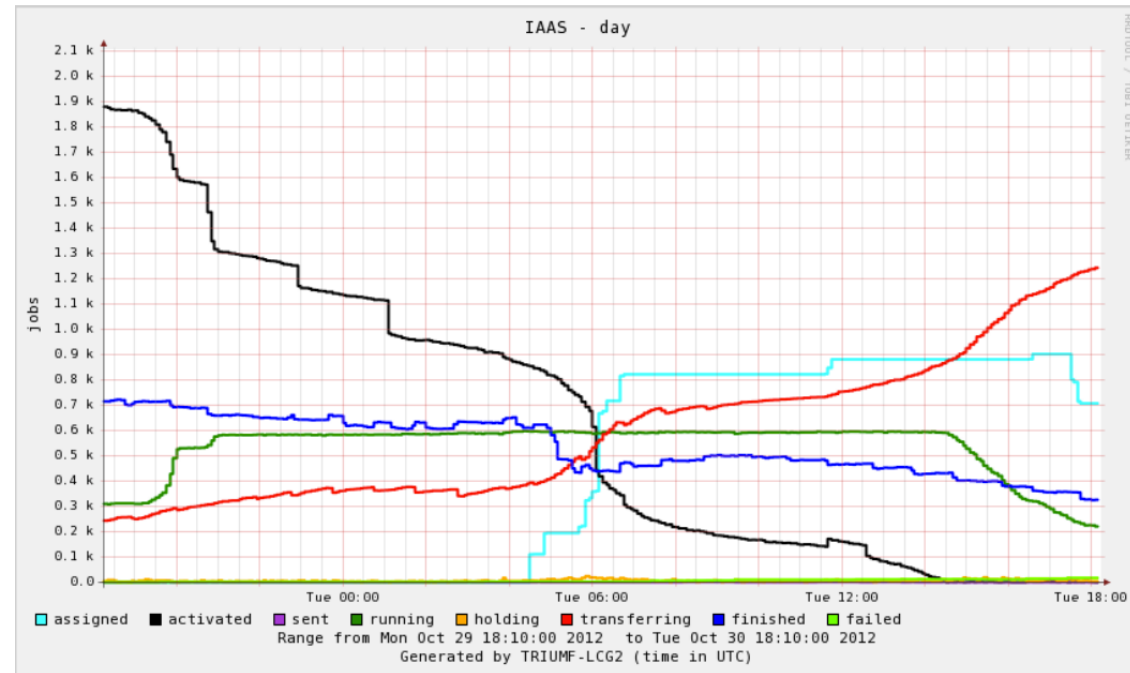
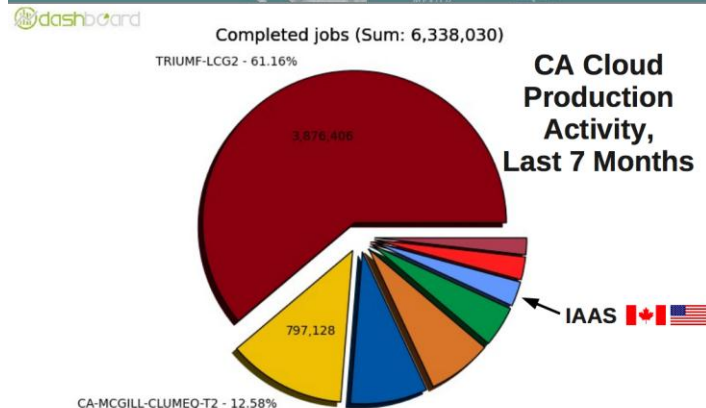
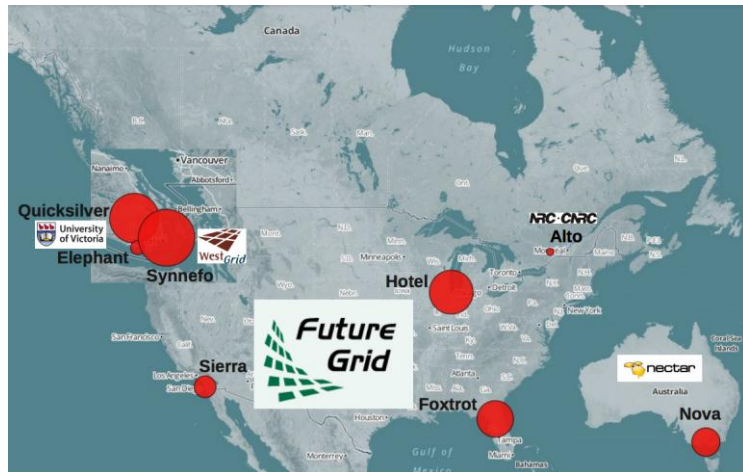


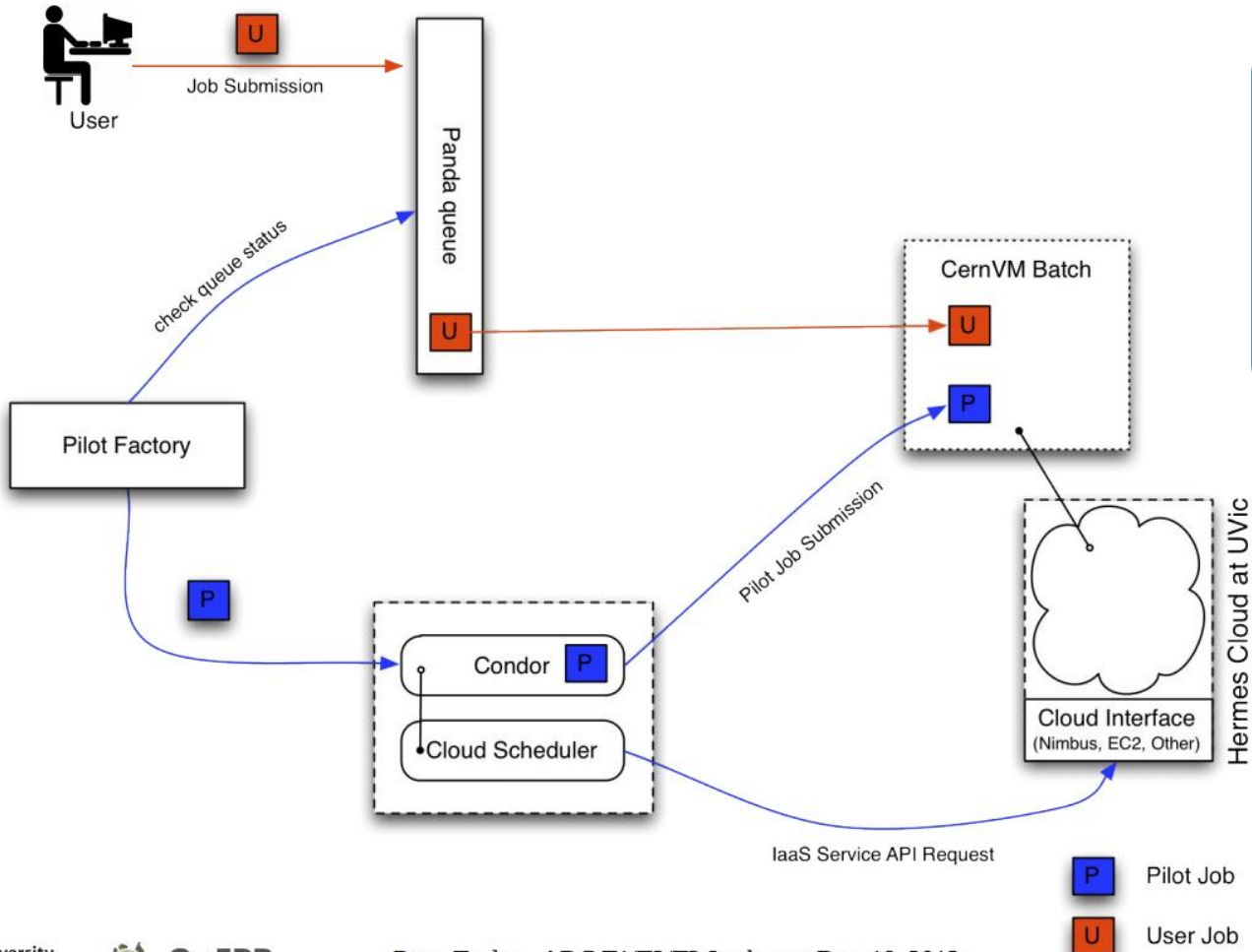
PanDA production and analysis queues

- Australia-NECTAR and ANALY_NECTAR
- 160 cores running ATLAS production jobs



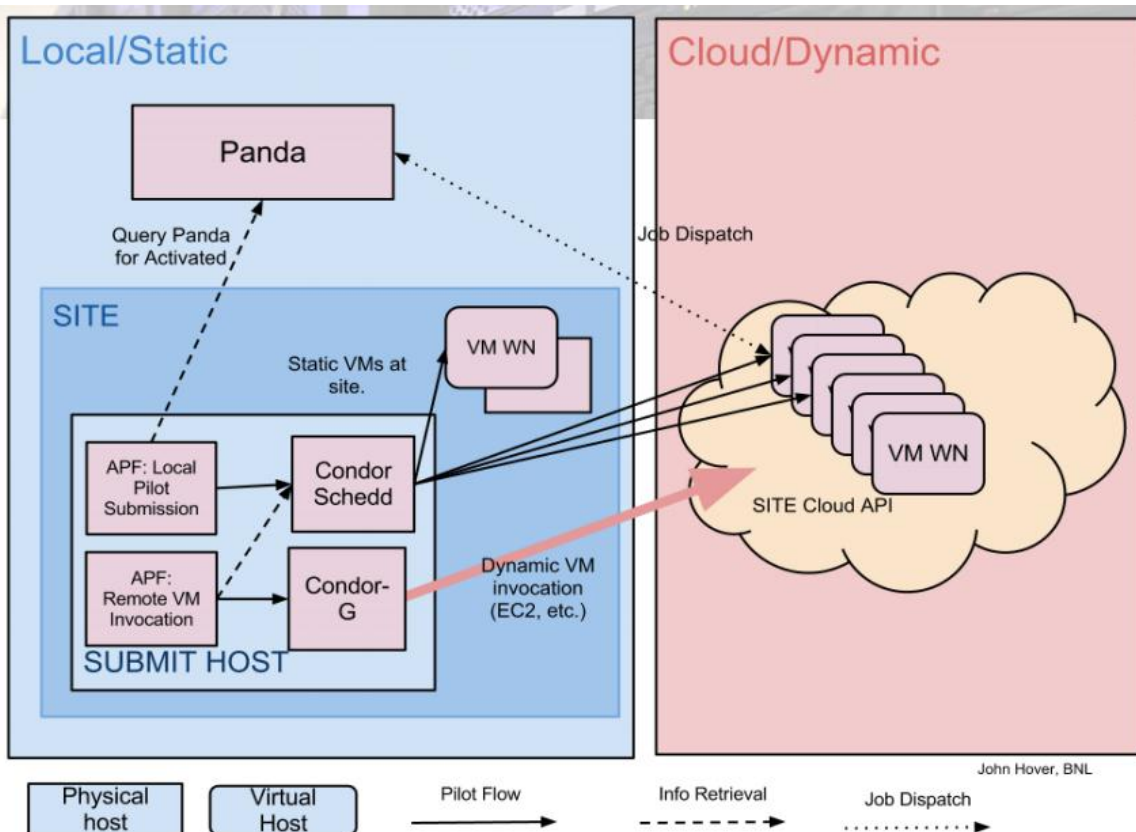
- One of the pioneers running production queues on cloud CPUs
 - Early tests Nov. 2011, standard operation since April 2012





UVic team can help you enable your cloud resources and manage them through CloudScheduler (see Ryan's presentation for steps and alternatives)

Ryan Taylor - ADC T1/T2/T3 Jamboree, Dec. 10, 2012



- Under development in BNL
- **Prototype by end February**
- Full support for VM lifecycle
 - Flexible algorithms to boot/terminate running VMs
 - Cascading hierarchies: programmatic scheduling on *local* → *private* → *commercial* clouds based on job priority and cloud cost

- LS1 opens the opportunity to use the HLT farms to extend offline computing resources with several thousand cores
- ATLAS and CMS plan to setup an OpenStack cloud
 - Too early to confirm most of the details
- GridPP built cloud instance at Imperial College
 - Test-bed for CMS in preparation of HLT cloud
 - UK-ATLAS usage of this cloud to be decided
 - Also HLT test-bed?
 - Evaluate root access to/from the cloud and storage at other UK sites?

- **US ATLAS users need to move towards central resources:** Beyond Pledged resources
- New technologies on horizon
 - Cloud computing and virtual Tier 3's
 - Federated storage and caching to reduce data management
 - Sergey and Doug independent work on data access/handling using Xrootd and the federation
 - WAN data access
 - User code must be improved to reduce latencies
 - Decouples Storage and CPU
- **Test-bed: scrounging of resources**
 - By necessity and not necessarily by choice
 - Public clouds (EC2, Google)
 - Research clouds (Future grid)
 - A stable Private cloud testbed would be appreciated

- PanDA analysis clusters on GCE+FutureGrid+AWS (Doug, Henrik, Val)
- Usage of puppet
 - Node configurations (CVMFS, xrootd, condor, PanDA, etc.)
 - Cluster orchestration
 - Investigating whether puppet can also be used for scaling the size of the cluster
- Open questions:
 - What is the scale of the virtual clusters?
 - From personal analysis cluster to multi-institution analysis cluster
 - Masterful vs. master-less setups – what is the easiest for the researcher?
 - Proxies for Panda Pilots (personal or robotic)
 - How is the data handling and caching done?
 - Initial activity used Federation storage as source
 - What is the latency incurred by this dynamic system?
 - What is the performance penalty of virtualization?

- T3 type PROOF/Xrootd cluster on Google Compute Engine (Sergey)
 - D3PD based physics analysis
 - 64 nodes with ~500 cores
 - 180TB of ephemeral storage
 - Studying data transfer using DQ2, xrdcp and direct read
 - Plan to study scalability of PROOF workload distribution system at very large number of nodes
- Interest from UT-Austin ATLAS group in personal cluster solution (Peter Onyisi)
 - Recent activity
 - Looking at very interactive use case (more PROOF than PanDA)
 - TACC Alamo FutureGrid hosted resources

- OpenStack: Currently cloud solution with most momentum



Starting to appear in ATLAS environment



HLT farms during LS1

- **Room for collaboration:** first with general user community and also within HEP
- We do have influence in future developments: Tim Bell appointed by Board of Directors to establish new User Committee

- Puppet as preferred configuration management tool
- Most deployments depend CVMFS

- Historically ATLAS was the first LHC experiment to actively explore the integration of cloud resources
 - Other experiments (CMS, LHCb) have started equivalent activities, opening possibilities for common exploration
- Running PanDA cloud production queues is a standard operation
 - Use case is best suited for the cloud: jobs have low I/O requirements
 - Examples in UVic, BNL, Australia, CERN...
- Some experience with XRootD for analysis clusters
 - Data model in analysis clusters still to be defined
 - Would be interesting to run performance benchmarks again
- CloudScheduler and soon APF to manage dynamic provisioning

- Data management to ensure scalability of above use-cases
 - We have almost no experience and also no concrete plan yet
 - BNL announced effort in cloud storage
 - Waiting for evolution of middleware
 - Data management would help to accelerate analysis use case
- APEL accounting in the cloud
 - Meeting between several interested parties will happen tomorrow
- Configuration and dynamic management of cloud resources
 - What information do we need in AGIS?
 - E.g. SW release publication for queues without CE
 - Declaration of downtimes for cloud resources, including services as Cloud Scheduler
 - Interest in creating, managing and deleting all elements (e.g. PanDA sites, DDM endpoints...) programmatically