

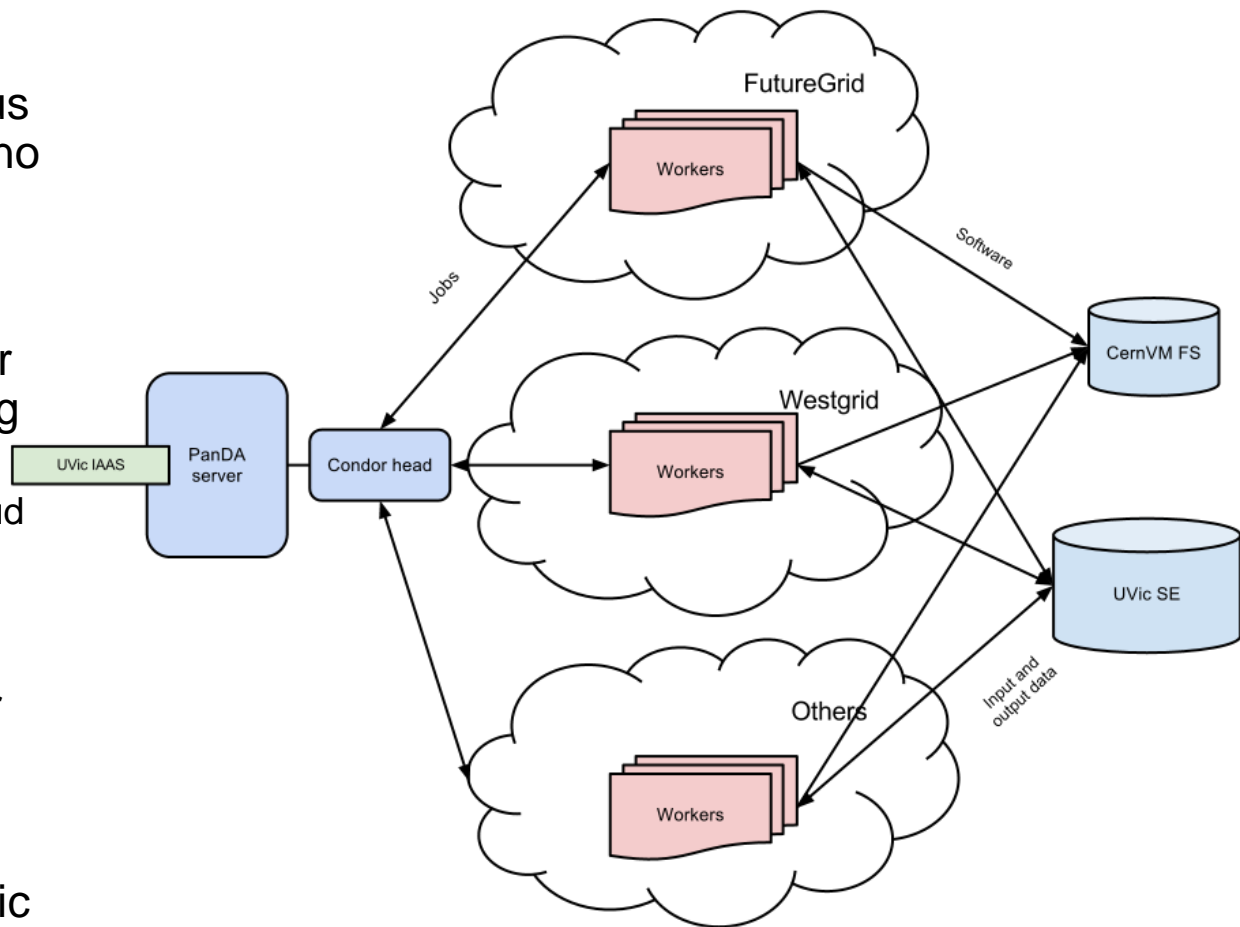
CPU accounting of public cloud resources

Fernando H. Barreiro Megino
CERN IT-ES-VOS

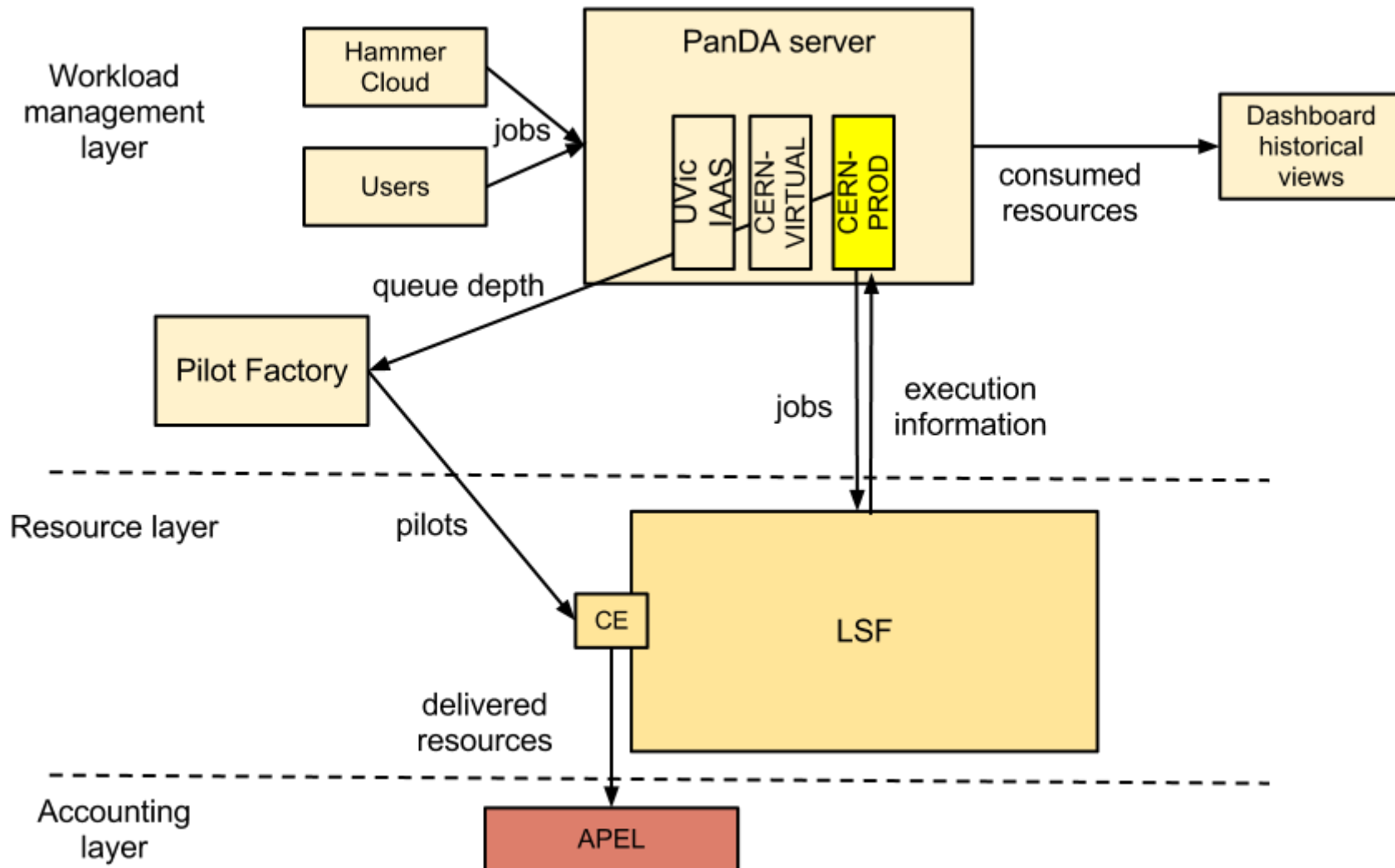


- This is a summary of a discussion held on 12.12.12 between
 - Fernando Barreiro (CERN-IT)
 - Jerome Belleman (CERN-IT)
 - Alessandro Di Girolamo (CERN-IT)
 - Sergey Panitkin (BNL)
 - Lalit Kumar Patel
 - Ulrich Schwickerath (CERN-IT)
 - Ryan Taylor (University of Victoria)
- First time the summary of the discussion is presented to the public – the ideas or implementation have not been approved by any party
- Alternative ideas and improvements are welcome

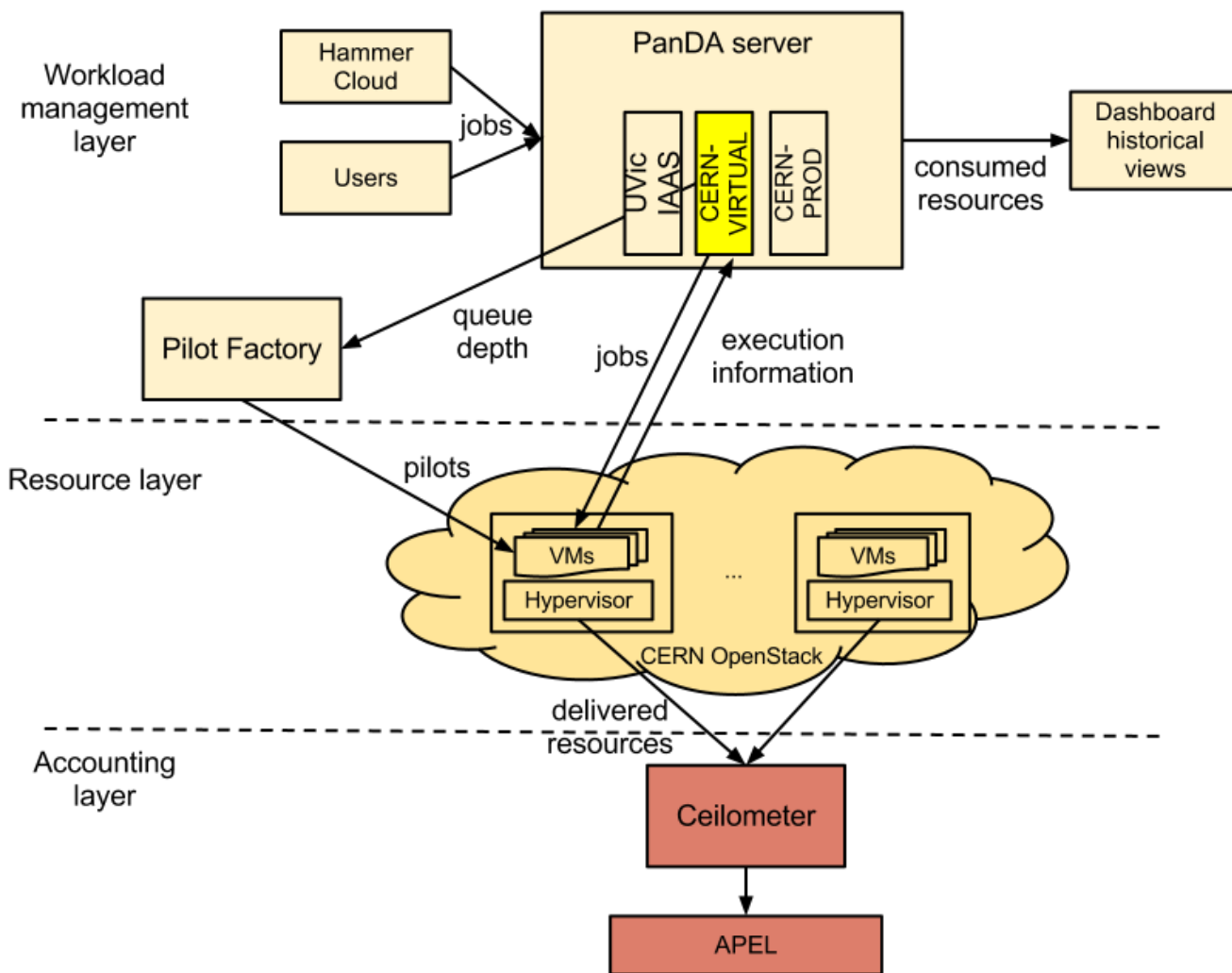
- UVic runs a production queue for ATLAS on various public clouds, but there is no recipe to account these resources yet
- Complications: there is no access to the bare metal or details about the underlying HW and its usage, e.g.
 - Are all machines in the cloud the same?
 - HS06 Benchmarking
 - CPU overcommitting
- Proposal to use PanDA (or experiments' WMS system) and Hammercloud as a complementary accounting source for public clouds



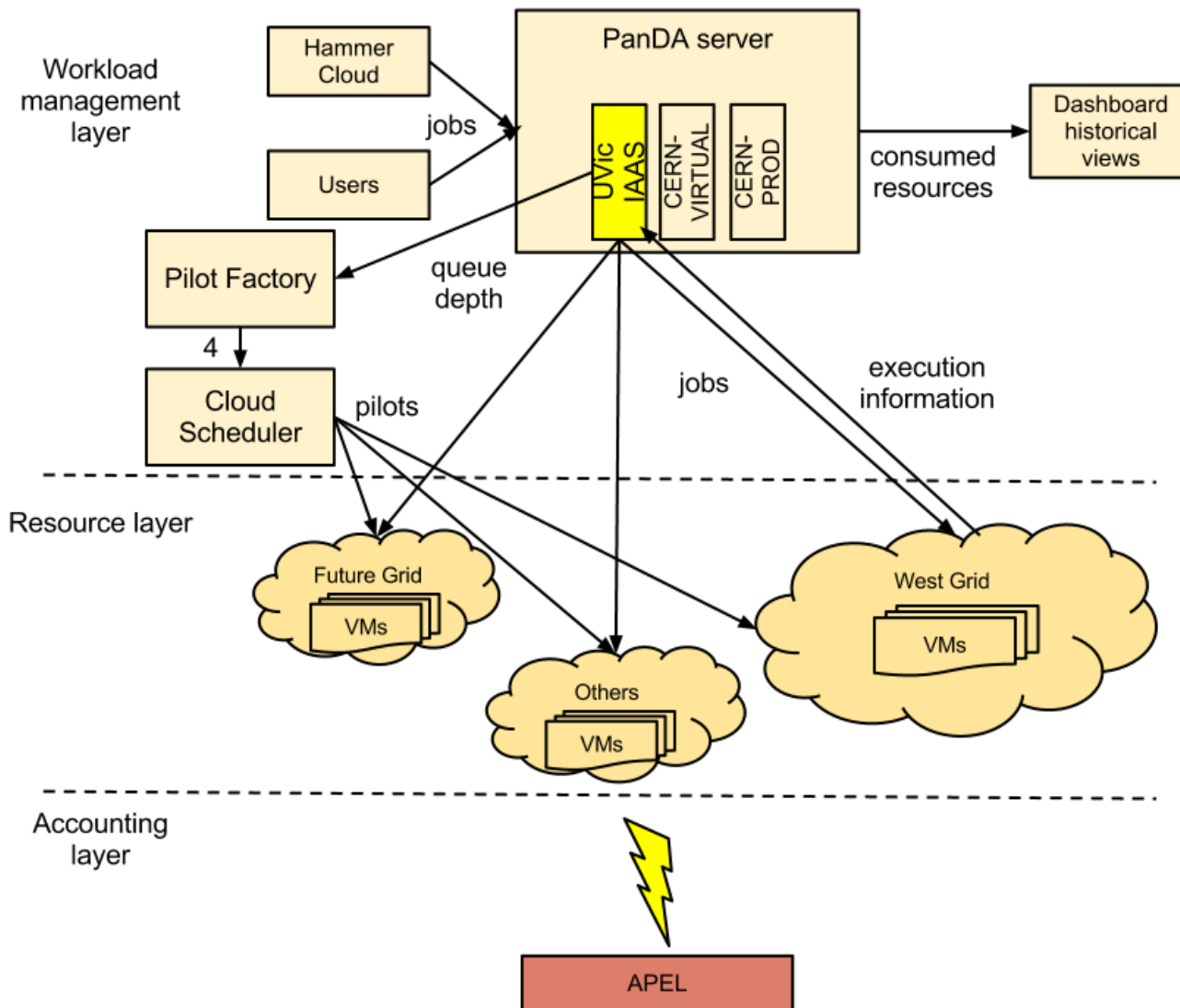
Original scenario: Batch



Solved scenario: Private cloud



New scenario: Public cloud



- Ceilometer feasible for private clouds: need direct access to the hypervisors to install the client
 - Private OpenStack clouds (e.g. BNL and Nectar) might find this a convenient solution
- Alternative idea: use the PanDA server as accounting source
 - Fed by the pilots
 - The PanDA database holds the duration of each job and the type of CPU
 - **Resources consumed by the workload** will always be lower than the **delivered resources**, since they don't include things as:
 - Boot-up/shut-down time of a VM
 - VM running partially empty (e.g. 8 core VM using only 1 job slot)
 - Using the CPU information provided by the pilot (retrieved from `/proc/cpuinfo`) will not be exact:
 - CPU can be reported as emulated CPU (e.g. QEMU Virtual CPU version (cpu64-rhel6) 4096 KB) and not the correct chipset
 - We don't have HEPSPC06 benchmarks of the VMs
 - We have to consider effects of over-commitment of CPU and others

- HammerCloud (HC) can be used to estimate some of the above uncertainties
 - HC submits same jobs to all sites
 - HC stores metrics: Events/second that the CPU of a site was able to process
 - Data mining on HC DB would reveal Ev/s rate of different sites
 1. Compare Ev/s metric and use the HEPSP06 benchmarks that are available for grid sites
 2. Estimate conversion factor
 3. Use conversion factor to approximate HS06 value for sites where this is unknown
- Aside from accounting purposes, useful data could be obtained for informational purposes to **compare the Ev/s rate and published HS06 value of sites**
 - Identify sites that may be publishing inaccurate or mis-measured HS06 values.
 - Study would show how well the HS06 benchmark represents the typical LHC experiment workload by examining the correlation with the real-world amount of work done

- No existing solution to account public cloud resources
 - Proposal to use PanDA (the WMS) and HammerCloud as a complementary source
- It would be very interesting **for all types of resources**:
 - To estimate the difference between consumed and delivered resources
 - To compare the Ev/s rates and published HS06 benchmarks of sites
 - See how well the HS06 benchmarks represent the typical LHC experiment workload