Second Price Enquiry Update

L. Field

On behalf of the IT-SDC WLC Resource Integration Team

CERN Cloud Procurement Activities

- Started in 2011 within the Helix Nebula partnership between
  - Leading research organizations
  - European commercial cloud providers

- Objective is to support the CERN's scientific computing programme
- Improve the CERN procurement process for cloud IaaS
  - Evaluate cost and benefit of IaaS
  - Transparently integrate commercial cloud IaaS
   - Into the LHC VO's computing infrastructures

Roadmap

✓ • First Procurement March 2015
  – Target a single VO, run simulation jobs
    • O(2000) VMs for 1 month

• Second Procurement October 2015
  – Target multi VOs, simulation jobs
    • O(4000) cores for 2 months

• Third Procurement Spring 2016
  – Target multi VOs, full chain data processing
    • O(2 x second procurement)

• EC co-funded joint Pre-Commercial Procurement (PCP) project
  – See next talk

First Procurement Summary
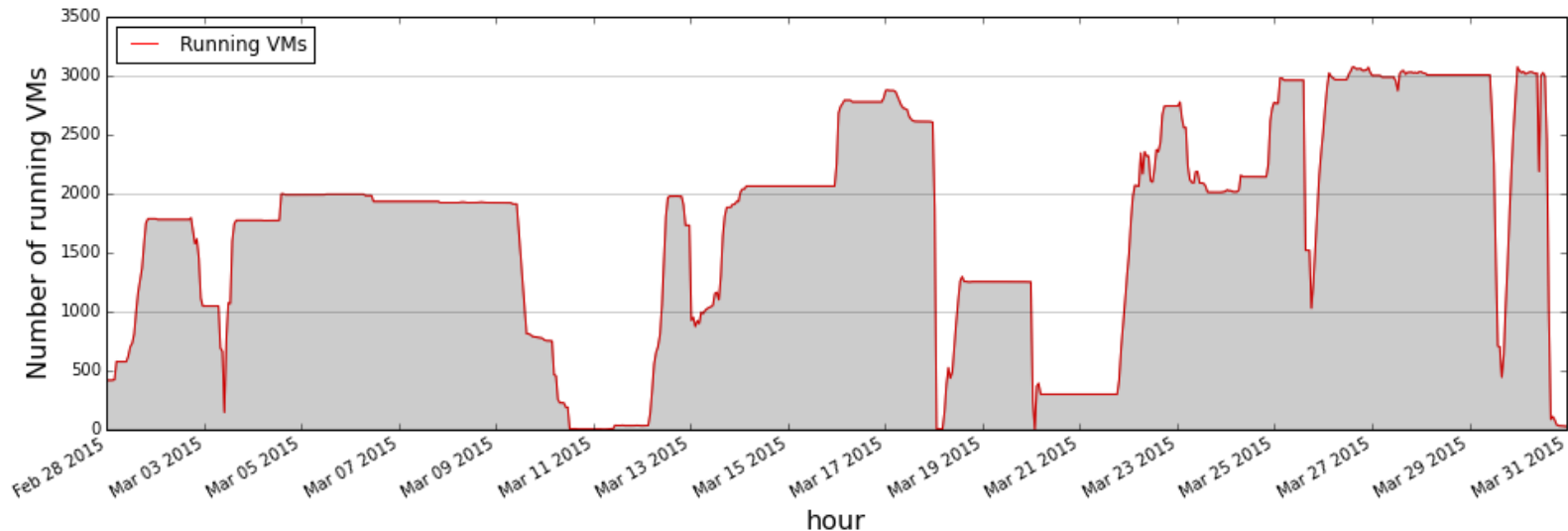
- Price enquiry announced November 2014
  - Production phase started in March 2015
- Up to 3,000 concurrent running VMs
  - 4 (+1) weeks of production
  - ~1.2 million CPU hours of processing
- ATLAS GEANT4 Simulation of tt̄ events
  - ~11.5 million events processed ⇔ ~80,000 jobs
  - ~93% CPU/Wall time ratio
    - ~9 hours single job duration
  - ~97% job wall time used for successful runs
    - Lost heartbeat is the main source of failures (~81%)
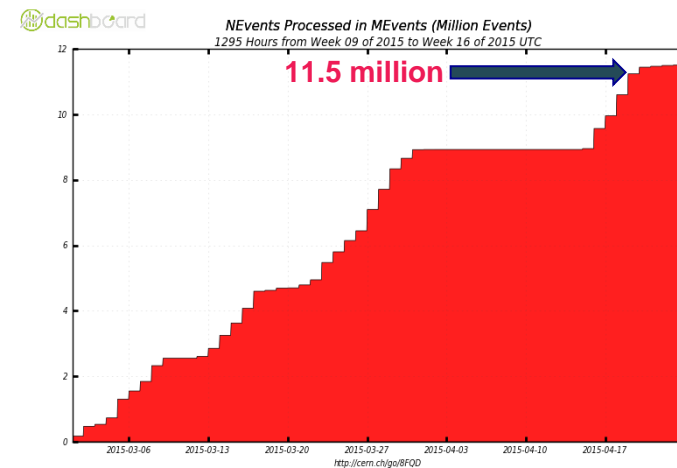
Experience Gained

- Managing VMs in commercial IaaS
    - Scaling, stability, API usability

- Monitoring, accounting and benchmarking of cloud resources
    - Client-side accounting is essential
        - Validate the invoices received from the provider for resources used
        - Ganglia monitoring is re-purposed for accounting of used resources
            - Real-time usage summaries and periodic reporting
            - Data archival for postmortem analysis
        - Needs follow up on a more tailored solution
            - VM accounting records provided by the VM
    - Monitor the usage of resources and identify inefficiencies
        - Data Analytics from Monitoring

- Benchmarking is a crucial step to commoditize cloud resources
    - Enable a "cloud" exchange
        - See for instance approach of Deutsche Börse Cloud Exchange
    - Required for the procurement procedures
        - Define technical specification, acceptance criteria and remediation options
        - Continuous monitor of the delivery
            - Run a fast benchmark after starting each single VM

Second Procurement

- Go beyond the simulation use-case of the first price enquiry
  - Simulation workload
  - Include all interested LHC VOs
  - Preparation for the EC PCP project with the other research labs
- Timeline (2015)
  - June-July: preparation of Technical Specification Document
  - Aug.   3rd: price enquiry sent to firms (DO-29401/IT)
  - Sept. 17th: closing date
  - Sept. 24th: notification to the selected firm
  - Oct. 1st: run for 2.5 months
- Addressed to single firms or combination of firms from CERN Member States

Specifications

- 1,000 multi-core Virtual Machines
  - 4 vCPU, 8 GB of RAM, 100 GB of storage, public IPv4 address
    - No persistent storage
  - Performance of the VMs qualified in terms of a reference benchmark
    - Defined in the tendering phase
    - Validated during the production activity

- Network
  - Connection to an NREN/GEANT to provide IPv4 connectivity

- API shall support:
  - CernVM or CentOS-6.x based snapshots
  - Accepted APIs
   - Amazon EC2, HNX SlipStream, OpenStack Nova, OCCI, DBCE

Benchmarking For Procurement

- Activity started in preparation for the first price enquiry

- Evaluated different alternatives based on the following requirements:
  - Open-source
    - Will be shared with cloud providers so they can run it
      - Crucial in tender phase to allow proper selection resources

  - Light-weight installation and reasonably quick to time
    - Run on many instances for fine-grained probing
      - Provide a cvmfs repository to simplify use at scale with CernVM

  - Reproducible
    - If random generation is used
      - Fix random seed in order to have always the same sequence of events

  - Correlates with experiment workloads
    - To extrapolate expectations on job duration
      - Profiting from cloud activities to study this relationship

ATLAS Kit Validation As A Reference

- Which workload to use for benchmarking?
  - CPU time/event is different for each workload
  - Measured that within ~10% the relative CPU/event performance doesn't depend on specific workloads
   - Confirmed also using a different approach: HammerCloud jobs

Study done in pre-procurement phase of the first price enquiry
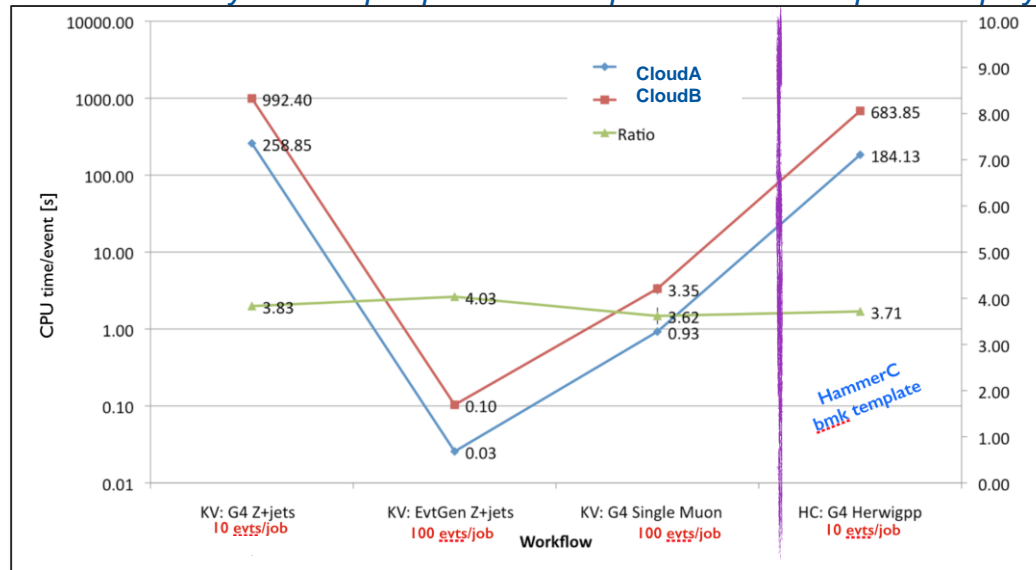
- Preferred workload: **G4 single muon**: faster running time O(few mins)
  - NB: the CPU time/event doesn't include the first event, to avoid bias due to the initialization process

Benchmarking During Procurement

- Tests executed by the suppliers using the KV benchmarking tool
  - They included the sampled performance in the Technical Annex of their bid
    - Two thresholds: Rejection (>1.5 s/evt) and Compensation (>1.2 s/evt)
    - Reminder: adjudication is on cheapest compliant bid
  - They used it to optimize their setup
    - In some cases they ran thousands of tests!!
- Benchmarking activity monitored in real time (see plot below)
- In some case very good performance but at higher prices

KV bmk results as performed by cloud providers

Rejection threshold
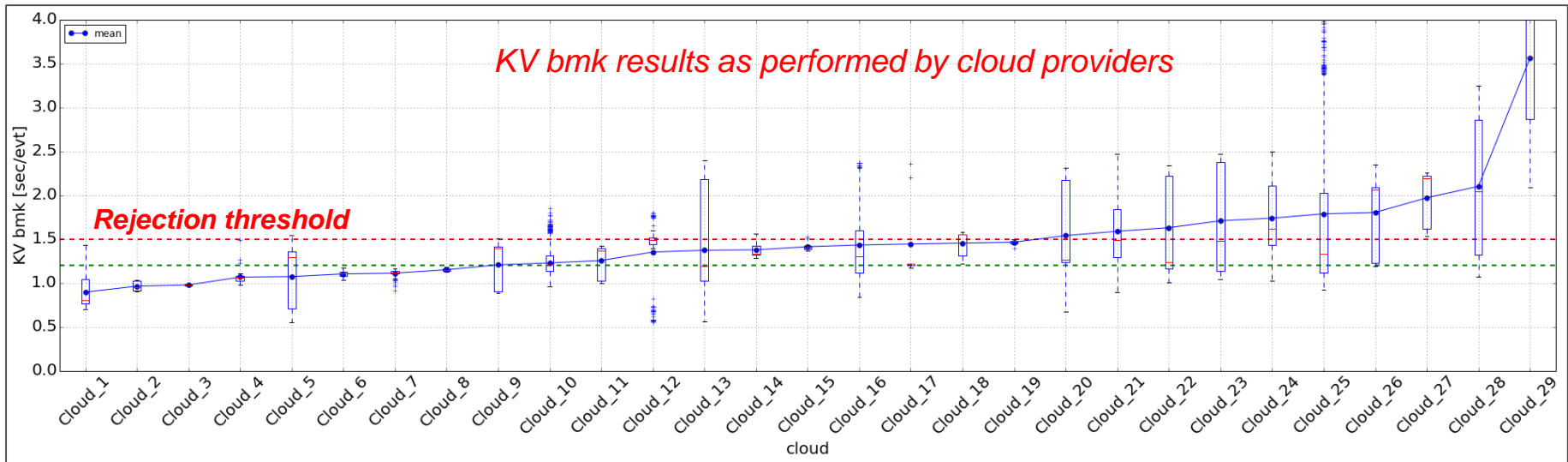
Delivery Status

- Delayed while waiting for GEANT connectivity
  - Expected to start within the next few days
- Service delivery
  - IT-SDC WLCG Resource Integration Team
    - VM provisioning, monitoring, benchmarking, accounting, etc.
  - Experiment liaisons
    - Coordinate VO specific activities
      - Job submission, monitoring, stage in/out, conditions
  - ~~Weekly~~ Daily coordination meetings
    - Deployment progress, feedback from VOs
- Encouraging preparatory phase
  - Using existing credits in commercial providers

Service Delivery
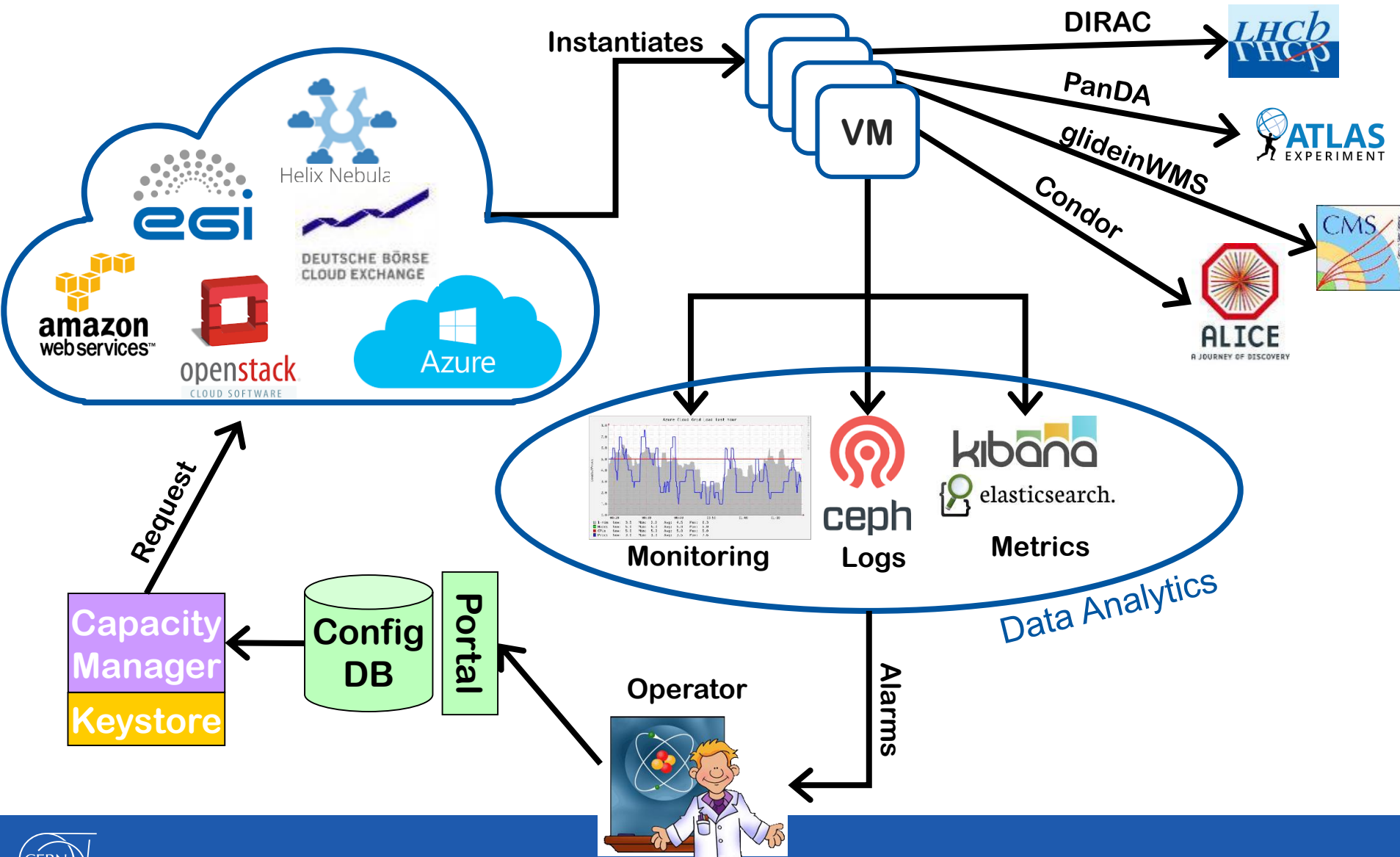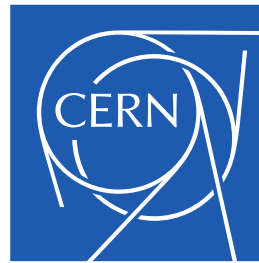
Supporting the diverse approaches
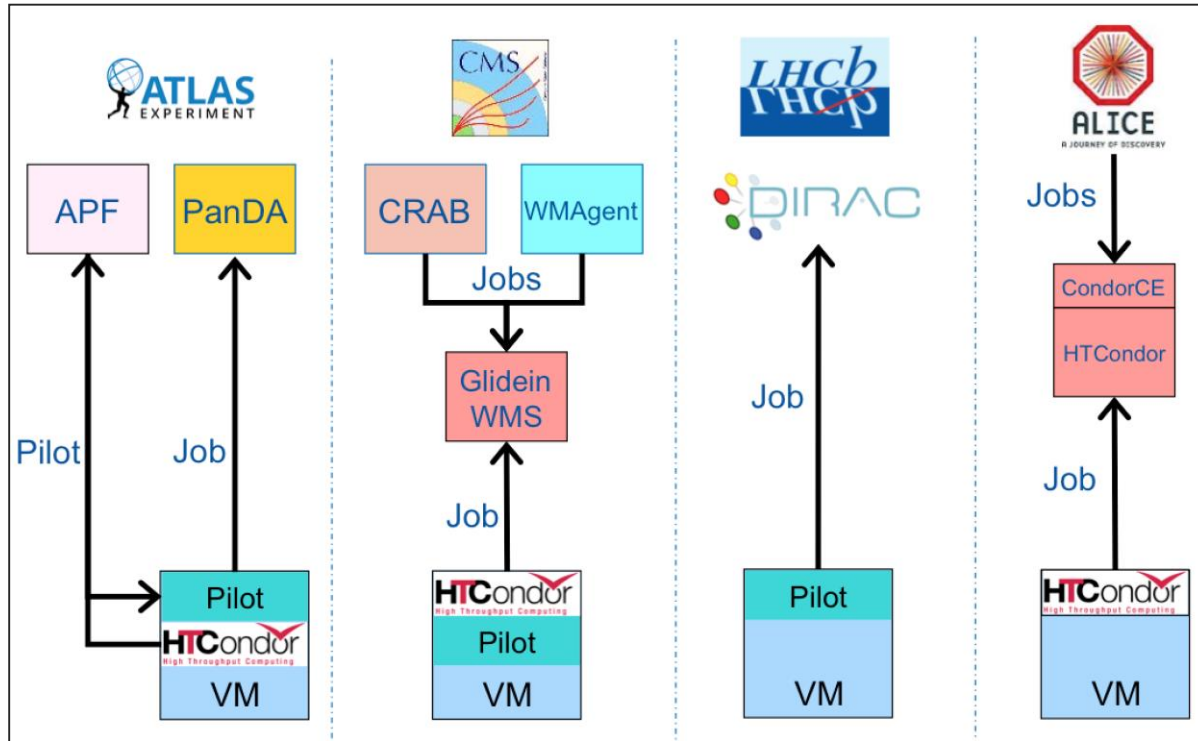
- 4 single-core or 1 multi-core job per VM (4 vCPU)
- Squid proxy cache deployed in each data centre
- Remote storage: Input data from/output data to CERN EOS
- Sharing of resources will be based on interest, requirements and readiness

Readiness

- ATLAS
  - Demonstrated working 4 core VM in multicore mode
    - Already use single core VMs for the first price enquiry
- CMS
  - Demonstrated working 4 core VM in singlecore mode
    - Condor providing four job slots
  - WMAgent to submit production jobs
- LHCb
  - Demonstrated working 4 core VM in singlecore mode
    - Four pilots run simultaneously
- ALICE
  - Single core VM ready for use
    - Following the traditional batch system approach
      - Condor Server with Condor CE deployed by IT-PES

- What still needs to be done?
  - Is there anything missing?
  - Are jobs ready to be sent?
  - Feedback needed on benefit for the experiment
    - To be compared with the cost

Conclusions

- Continuation of the price enquiry series
    - Understand our readiness for commercial clouds
- Improving the procurement process
    - Hidden but crucial aspect
- The importance of benchmarking
    - See Domenico's HEPIX talk for more details
        - https://indico.cern.ch/event/384358/session/12/contribution/25
- The importance of consumer-side accounting
    - Repurposing monitoring data but should consider a tailored solution
- The delivery of the second price enquiry will start within the next few days
    - Target multiple LHC VOs to run simulation workloads
        - No persistent storage
    - Activity will run for 2.5 months
- Service Delivery by the IT-SDC Resource Integration Team
    - In close collaboration with the experiment liaisons