

Helix Nebula Marketplace for end-users: the CERN experience

D. Giordano (CERN)

The Helix Nebula Initiative & PICSE: Towards a European Open Science Cloud
26 June 2015

Large Hadron Collider

CMS

LHCb

LHC computing resources in 2014

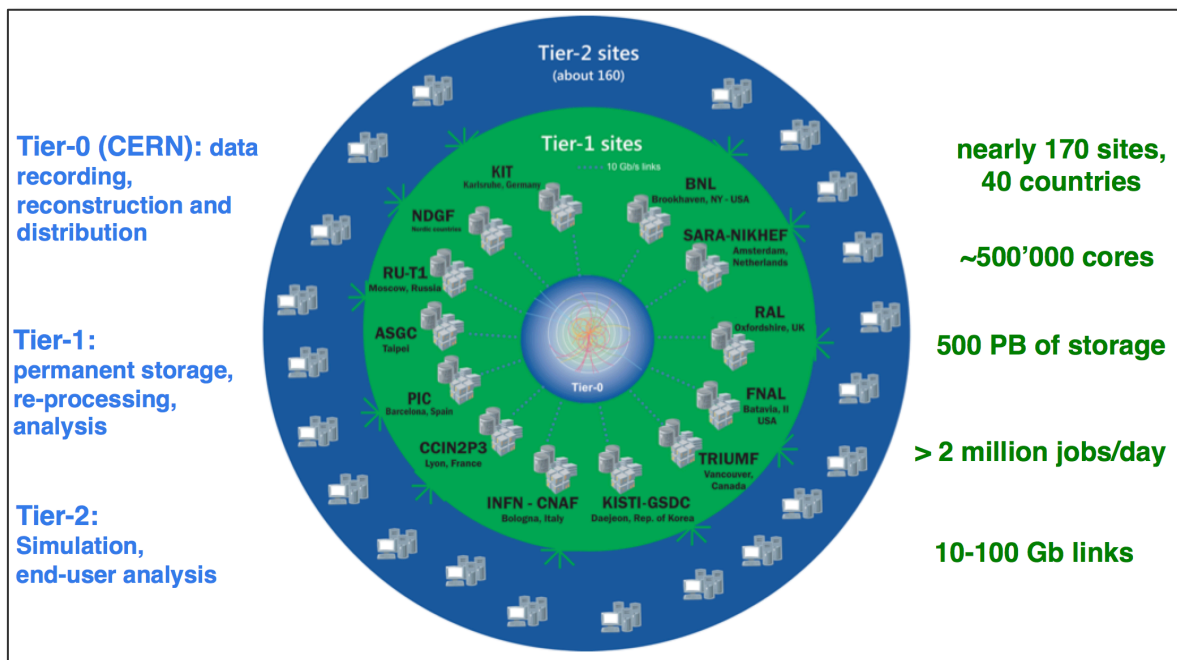
- 3.4 billion CPU hours
- 100 PetaBytes storage

ATLAS

ALICE



The Worldwide LHC Computing Grid

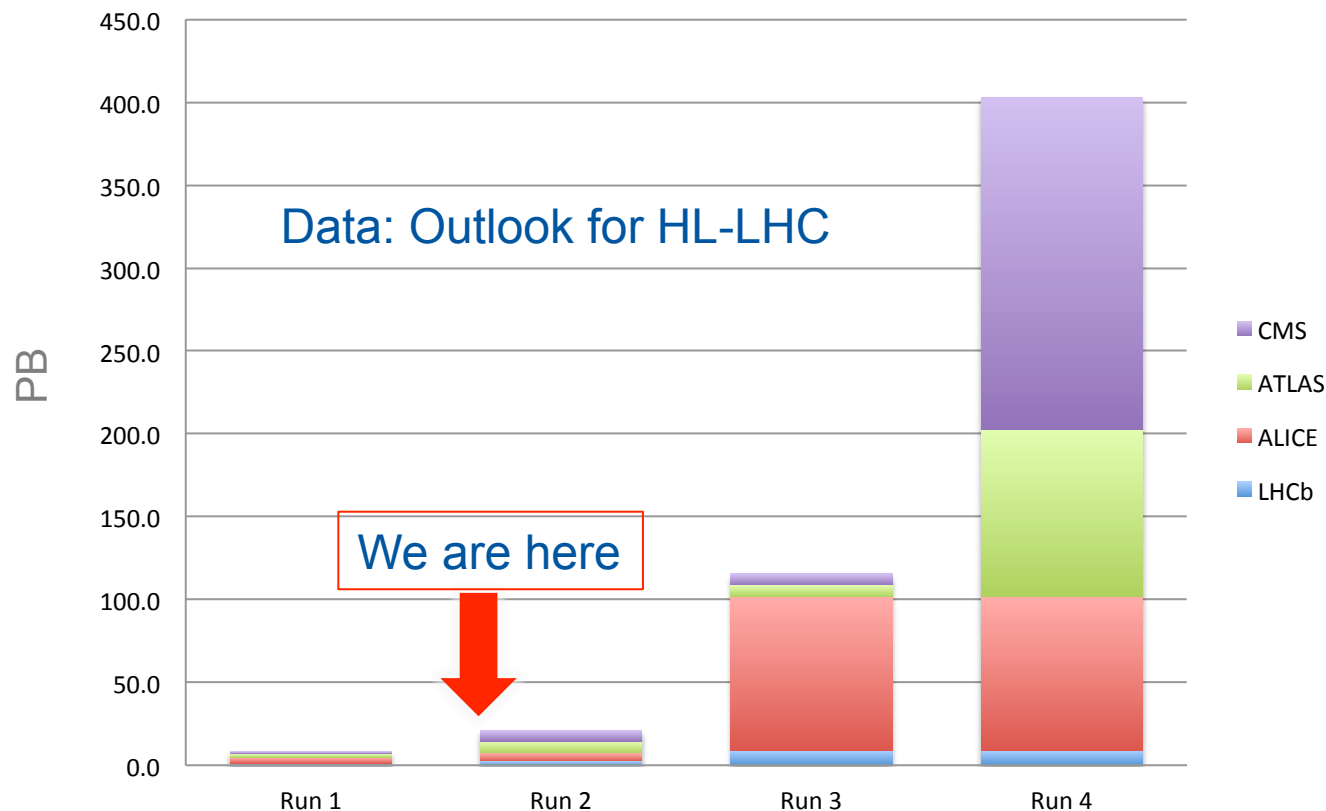


WLCG:

An International collaboration to distribute and analyse LHC data

Integrates computer centres worldwide that provide computing and storage resource into a single infrastructure accessible by all LHC physicists

Growing data and processing needs

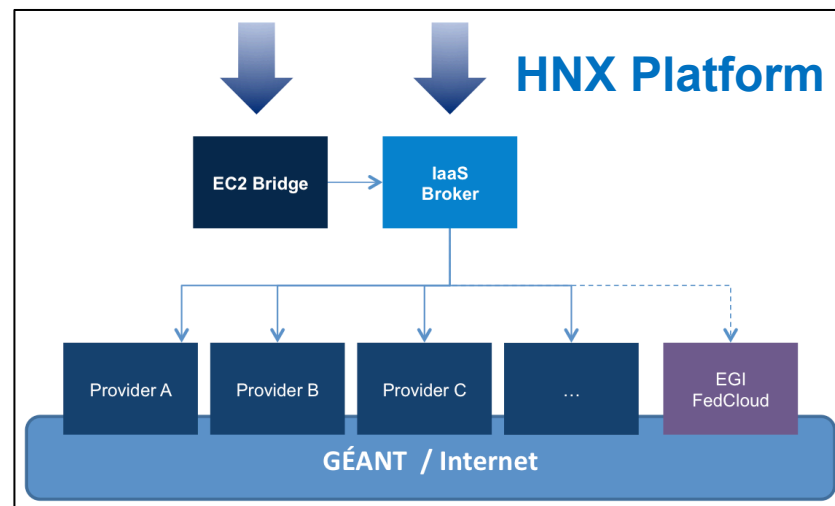
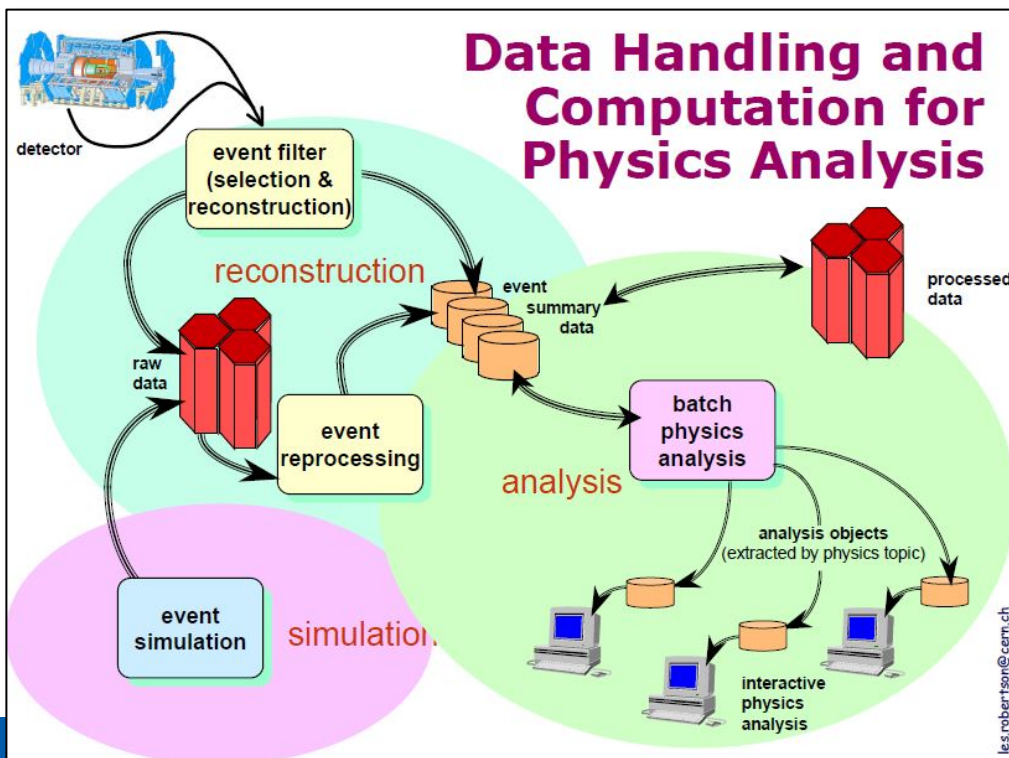


- Very rough estimate of a new RAW data per year of running using a simple extrapolation of current data volume scaled by the output rates.
- To be added: derived data (ESD, AOD), simulation, user data...

Helix Nebula: CERN use case

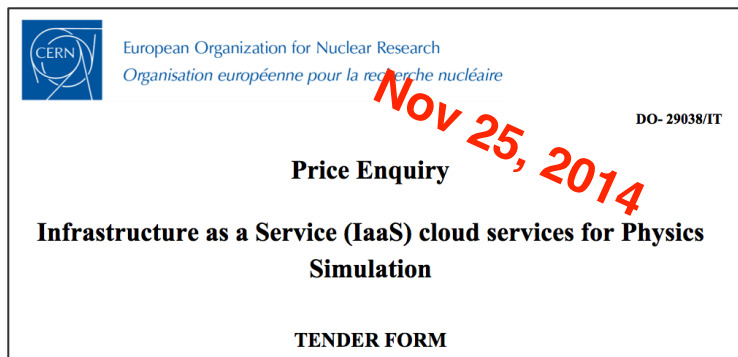
Evaluating the use of public cloud IaaS for LHC computing

- Transparent integration of cloud resources within experiments' distributed computing software and services
- Evaluation of financial costs for processing, data transfer, storage



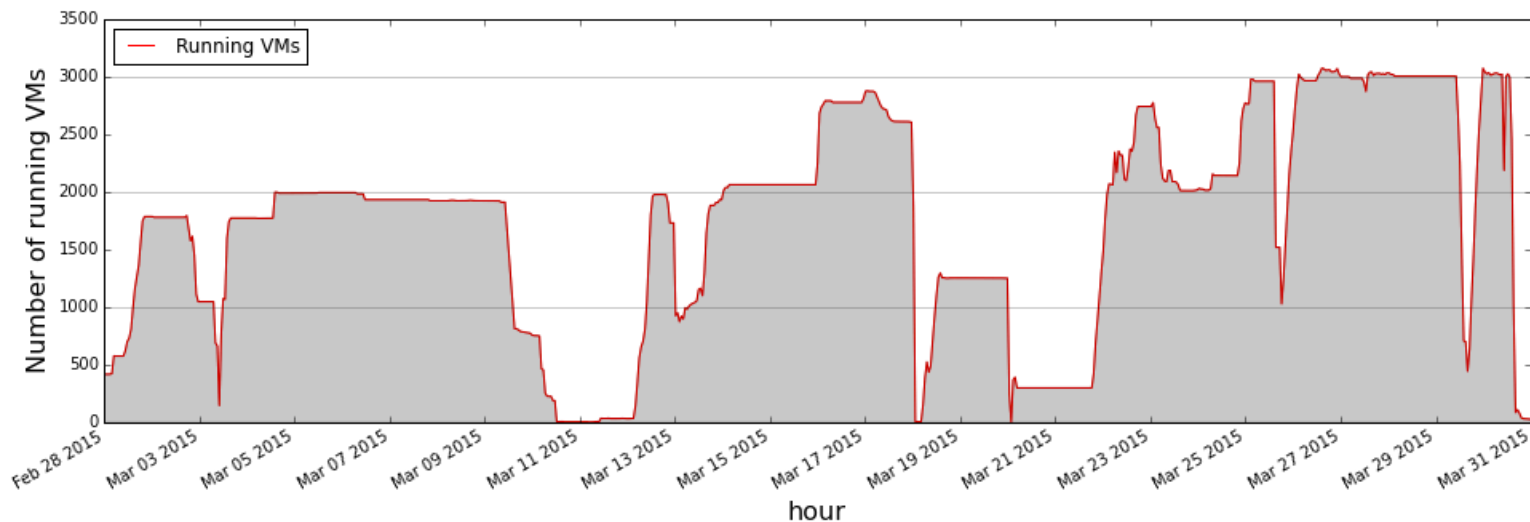
Example HNX procurement

- CERN price enquiry for a small fraction of the resources necessary to run the ATLAS experiment simulation software

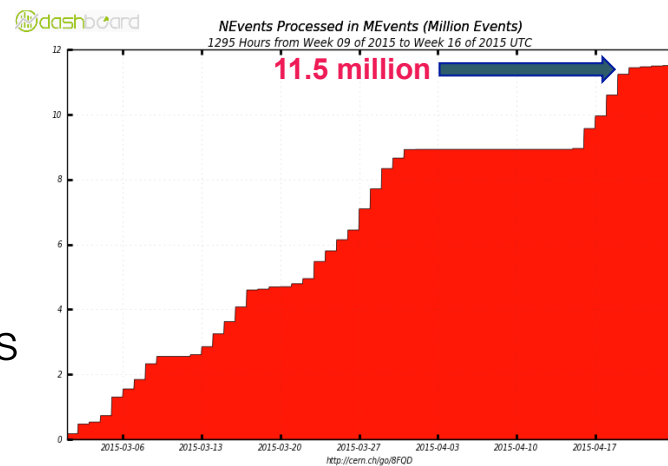


- Several offers received
 - Atos selected as contractor
 - Providing services via the SlipStream engine (from SixSq SME)
 - Services delivered over the GÉANT network

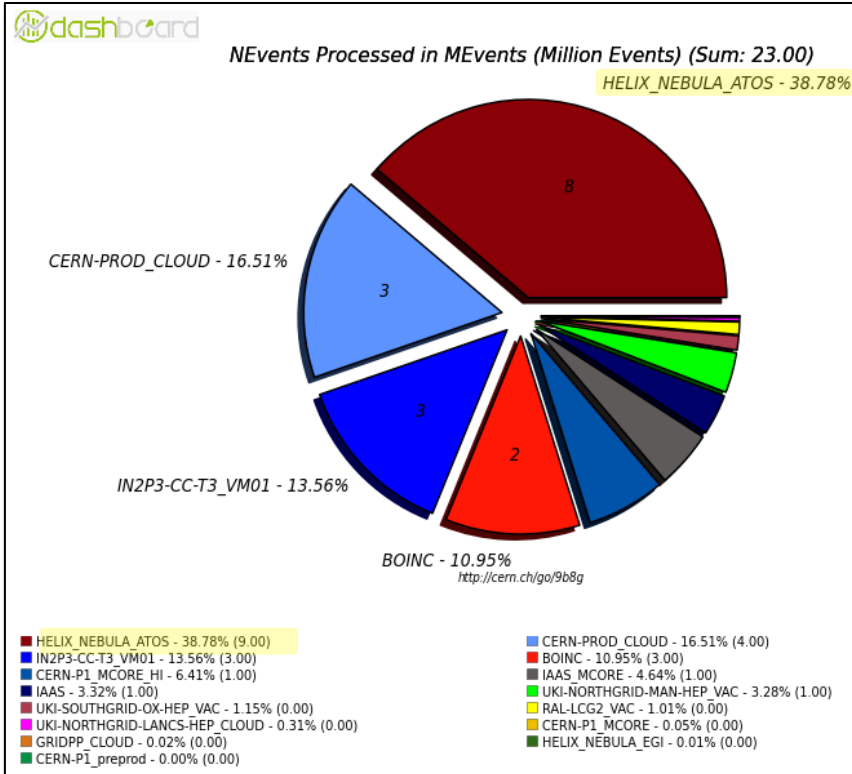
Summary results



- Production phase started in March'15
- Up to **3,000** concurrent running VMs
 - 5 weeks of production
 - ~**1.2 million** CPU hours of processing
- ATLAS GEANT4 Simulation of $t\bar{t}$ events
 - ~**11.5 million** events processed \Leftrightarrow ~80,000 jobs
 - ~9 hours single job duration
 - ~**97%** job efficiency



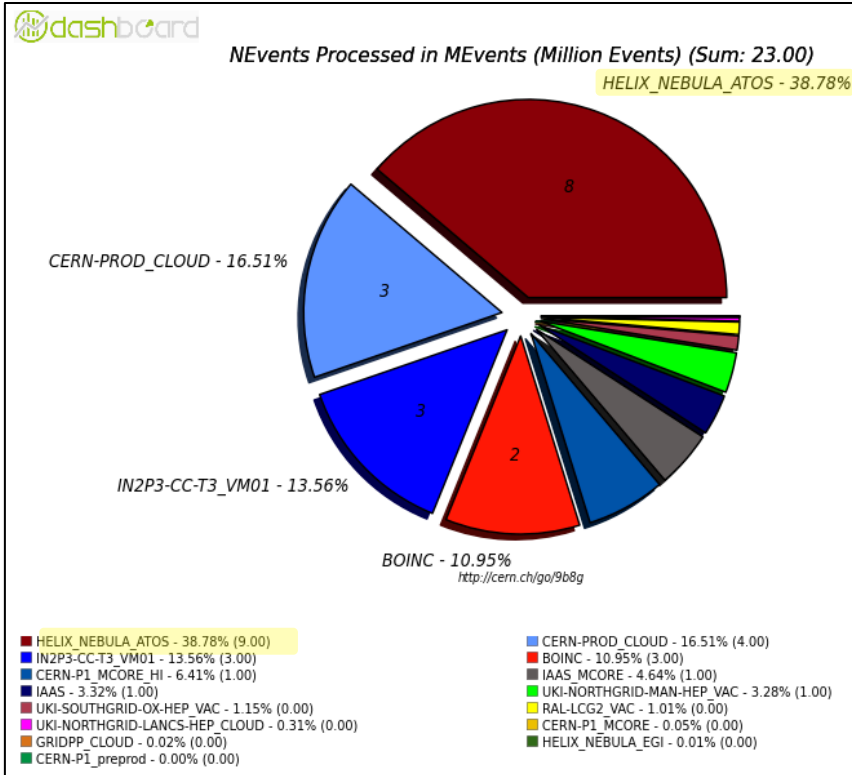
Compared with other ATLAS cloud sites (March)



- Significant contribution compared with other ATLAS cloud sites running simulation
 - **Largest** # of processed events

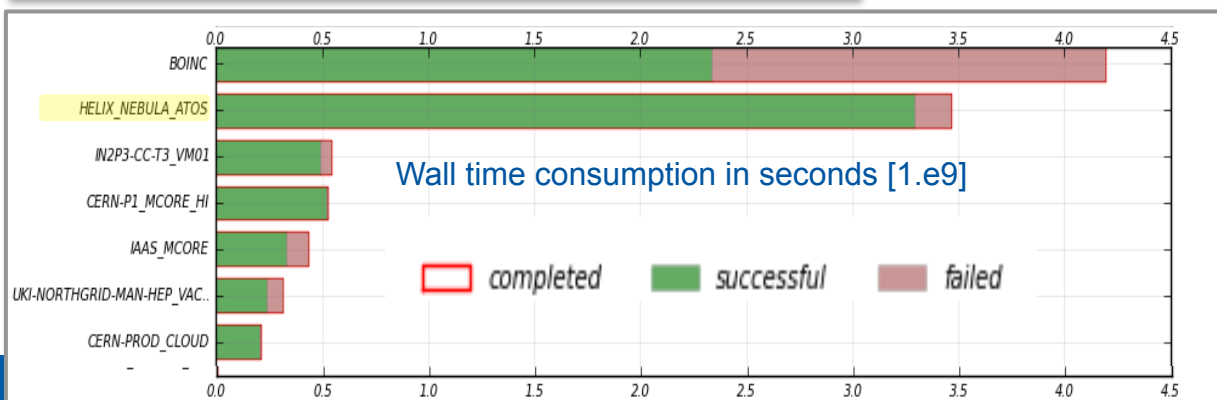
(*) CERN-PROD_CLOUD: 90% of processed events belongs to fast workloads (~24 s/evt)

Compared with other ATLAS cloud sites (March)

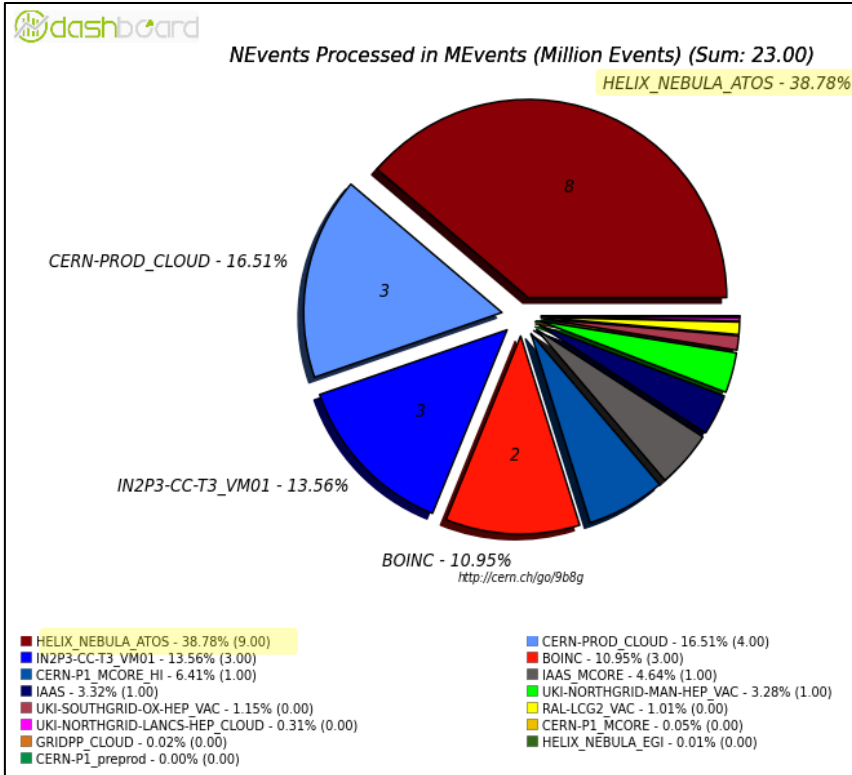


- Significant contribution compared with other ATLAS cloud sites running simulation
 - **Largest** # of processed events
 - **Longest** wall time consumption

(*) CERN-PROD_CLOUD: 90% of processed events belongs to fast workloads (~24 s/evt)

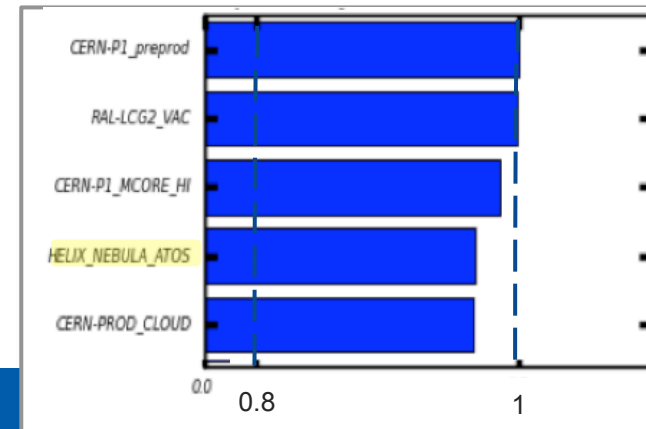
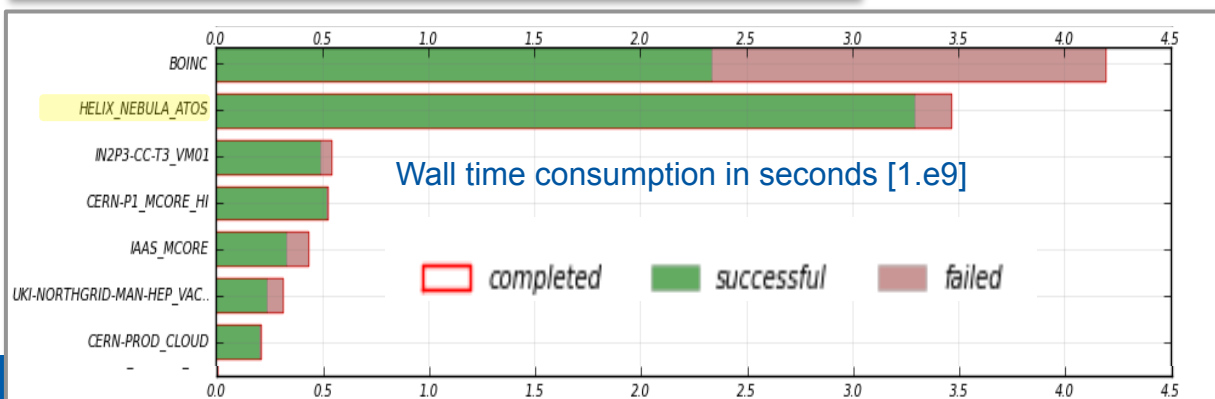


Compared with other ATLAS cloud sites (March)



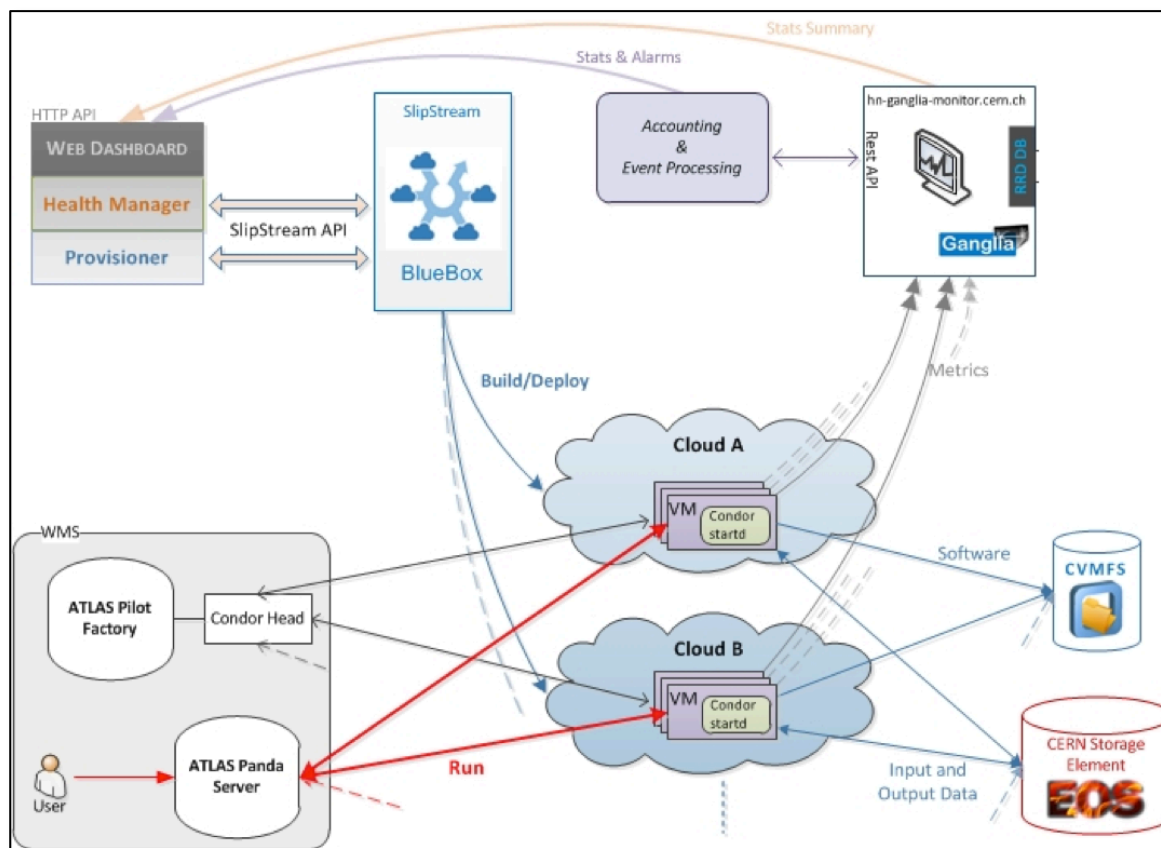
- Significant contribution compared with other ATLAS cloud sites running simulation
 - **Largest** # of processed events
 - **Longest** wall time consumption
 - **High** wall time efficiency

(*) CERN-PROD_CLOUD: 90% of processed events belongs to fast workloads (~24 s/evt)



Provisioning & monitoring chain

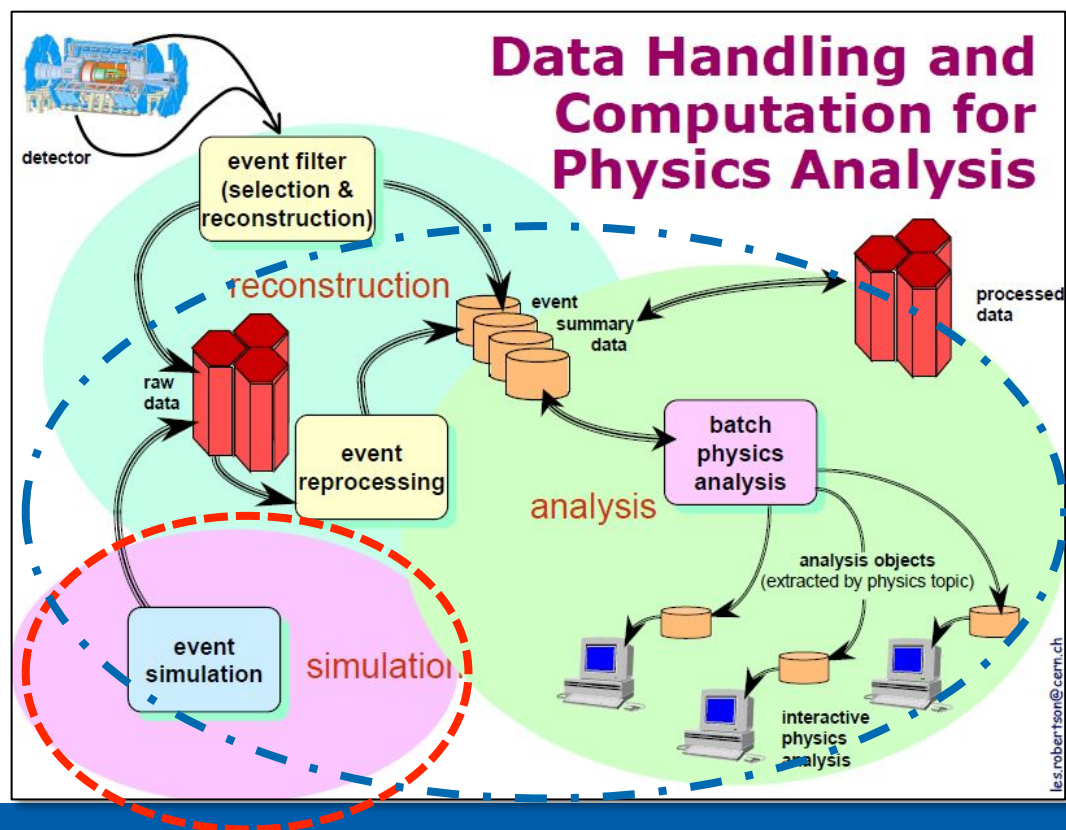
Cloud resources fully integrated in the computing system of the ATLAS experiment



Future procurement initiatives

- Go beyond the simulation use-case of the 2014 price enquiry
 - Include reconstruction and analysis workloads

- Two procurements to be launched in July '15
 - simulation use-case
 - full-chain:
 - simulation
 - + reconstruction
 - + analysis



Final remarks

- **Successful activity**
 - Unprecedented capacity acquired and maintained in HNX for CERN use case
 - 1.2 million CPU hours of processing (~11.5 million events)
- The use of commercial IaaS is technically feasible
 - Evaluated for simulation workloads
 - Next focus on workloads more challenging for storage and network
 - Variations in APIs can be overcome with reasonable technical effort
 - Independent consumer-side monitoring and benchmarking of the used resources used is essential



www.cern.ch