CERN Flagship

D. Giordano

(CERN)

Helix Nebula - The Science Cloud:
From Cloud-Active to Cloud-Productive
14 May 2014

 The Helix Nebula project is co-funded by the European Community Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no 312301

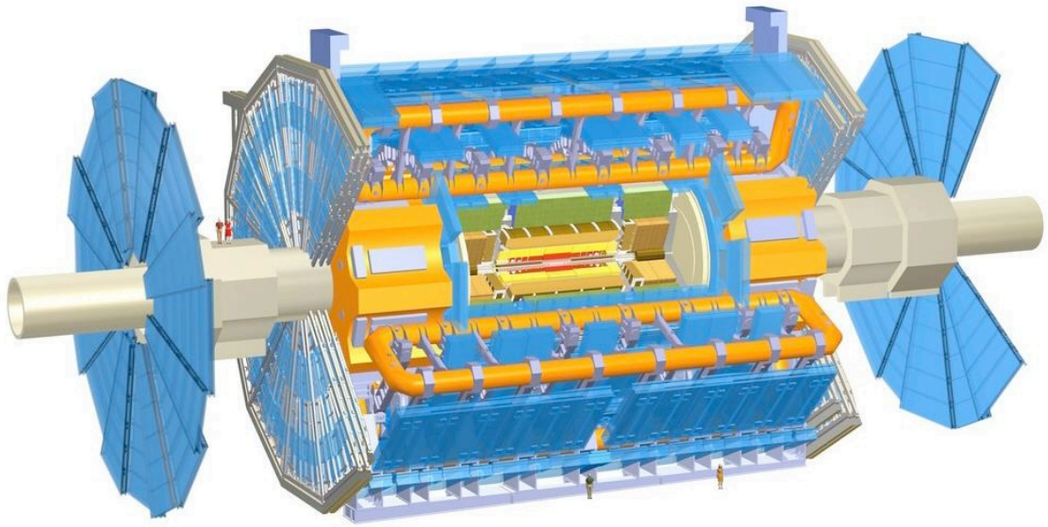
Outline

- CERN use case
- Cloud experiences @ CERN
- CERN Helix Nebula experience
- Conclusions

Helix Nebula: CERN flagship use case

Aim

- Evaluating the use of cloud technologies for LHC data processing
- Transparent integration of cloud computing resources with ATLAS distributed computing software and services
- Evaluation of financial costs of processing, data transfer and data storage
- Service Level Agreements and Governance model



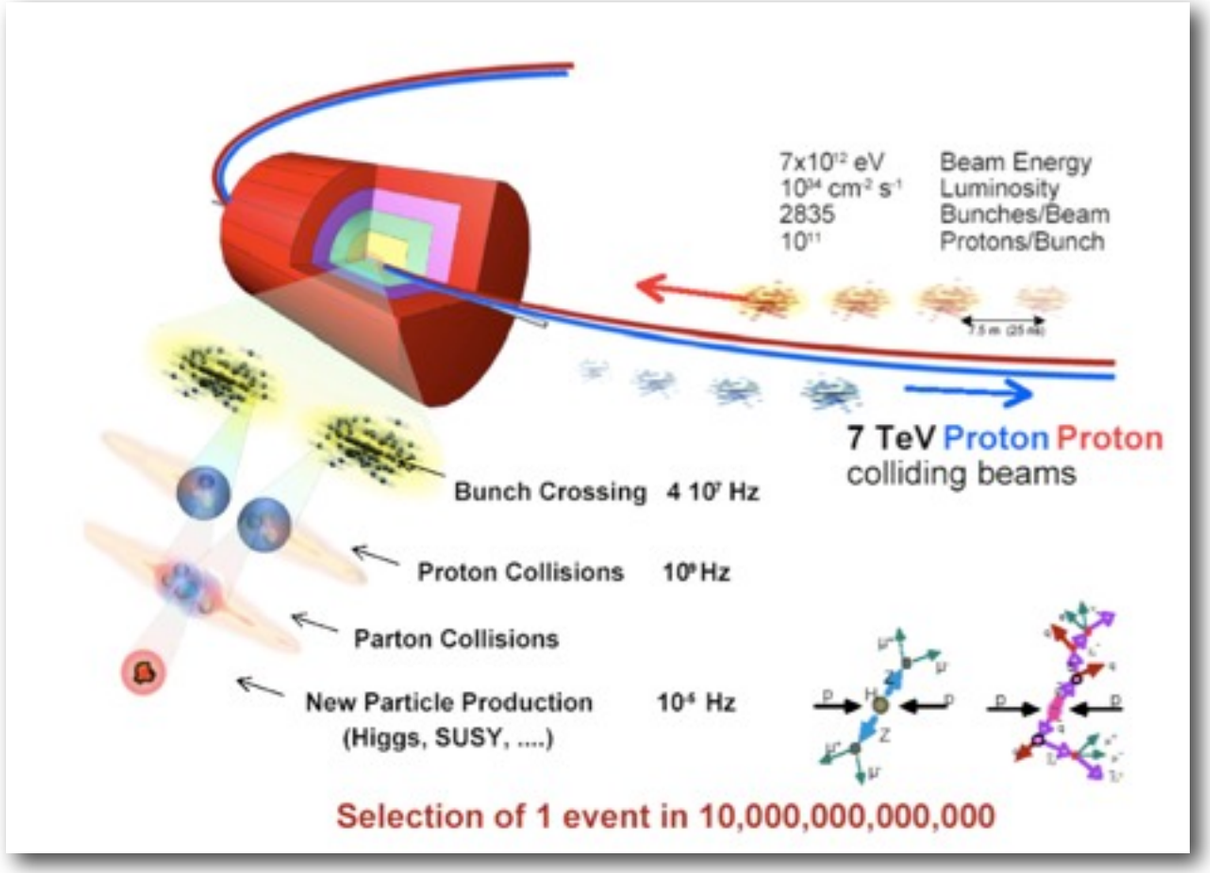
ATLAS detector

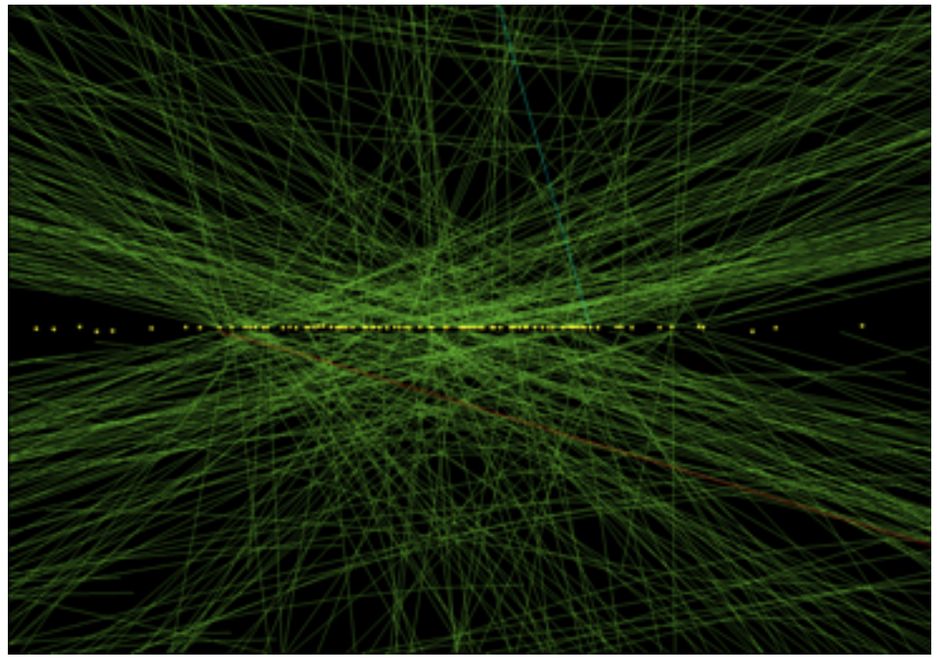
The physics and computing challenge

Billions of events are delivered to the experiments from proton-proton and proton-lead collisions in the Run 1 period (2009-2013)

- Collisions every 50 ns = 20 MHz crossing rate
- ~35 interactions per crossing at peak luminosity
- ~1600 charged particles produced in every collision
- ~ 5PB/year/experiment

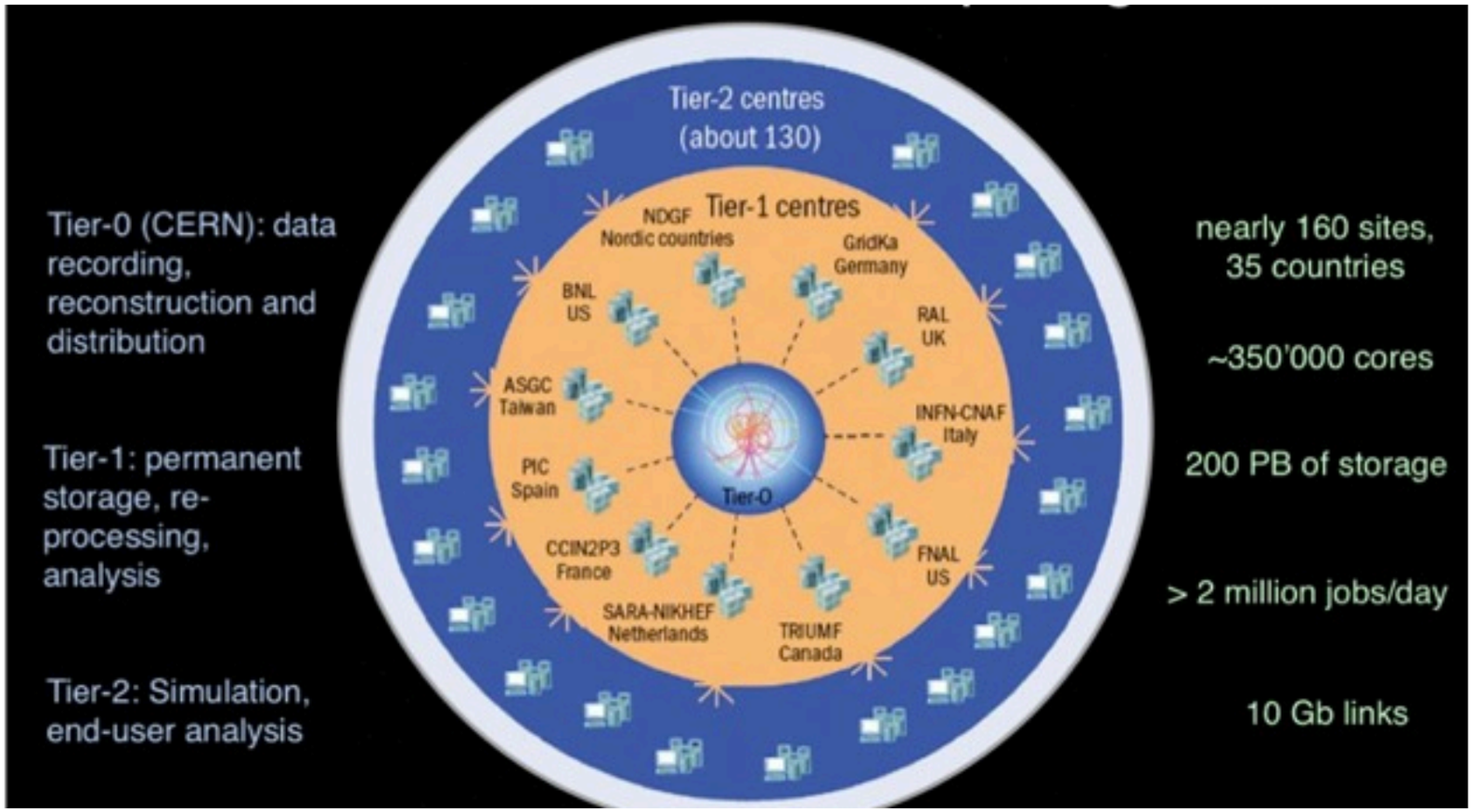
A huge computing challenge





78 reconstructed vertices in event from high-pileup in CMS detector

The Worldwide LHC Computing Grid



WLCG: an international collaboration to store, process and analyse data produced from the LHC

- Integrates computer centres worldwide that provide computing and storage resources into a single infrastructure

Cloud Initiatives

Several R&D initiatives started by the experiment collaborations to investigate and exploit cloud resources

- Utilize private and public clouds as an extra computing resource
- Mechanism to cope with peak loads on the Grid

ATLAS
"Grid of Clouds"

[S. Panitkin CHEP2013]

CERN HLT farms opportunistic usage

Exploit computing resources available in the High Level Trigger farms adopting a Cloud infrastructure

- Based on OpenStack

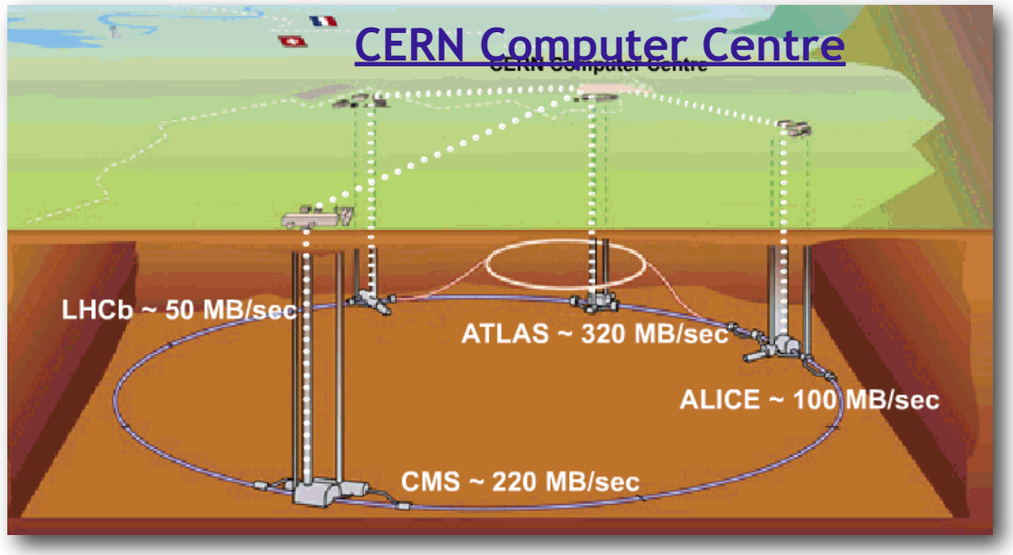
Doubling the capacity of biggest experiment Tier1s

- ATLAS: 15k cores (28k HyperThreading)
- CMS: 13k cores (21k HyperThreading)

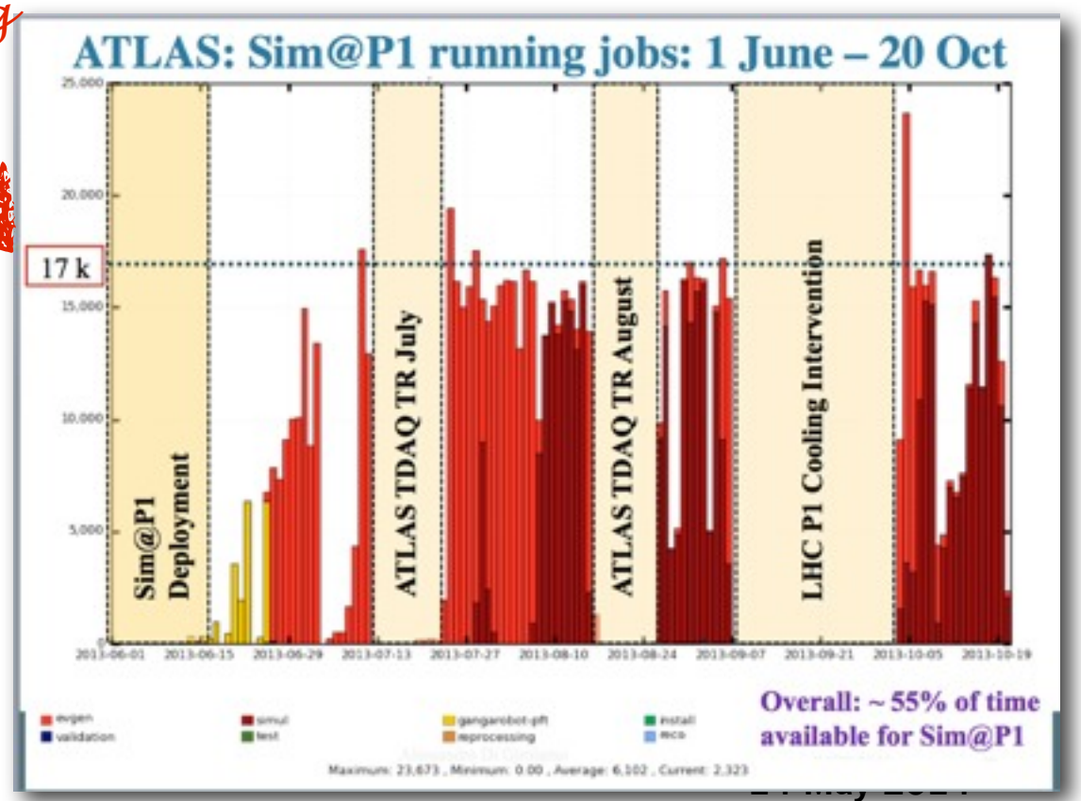
No impact on data taking, easy switch of activity

- from 0 to 17k jobs running in ~3.5h
- from 17k jobs running to TDAQ ready in ~10'

Data flow to permanent storage: 4-6 GB/sec



17k running jobs



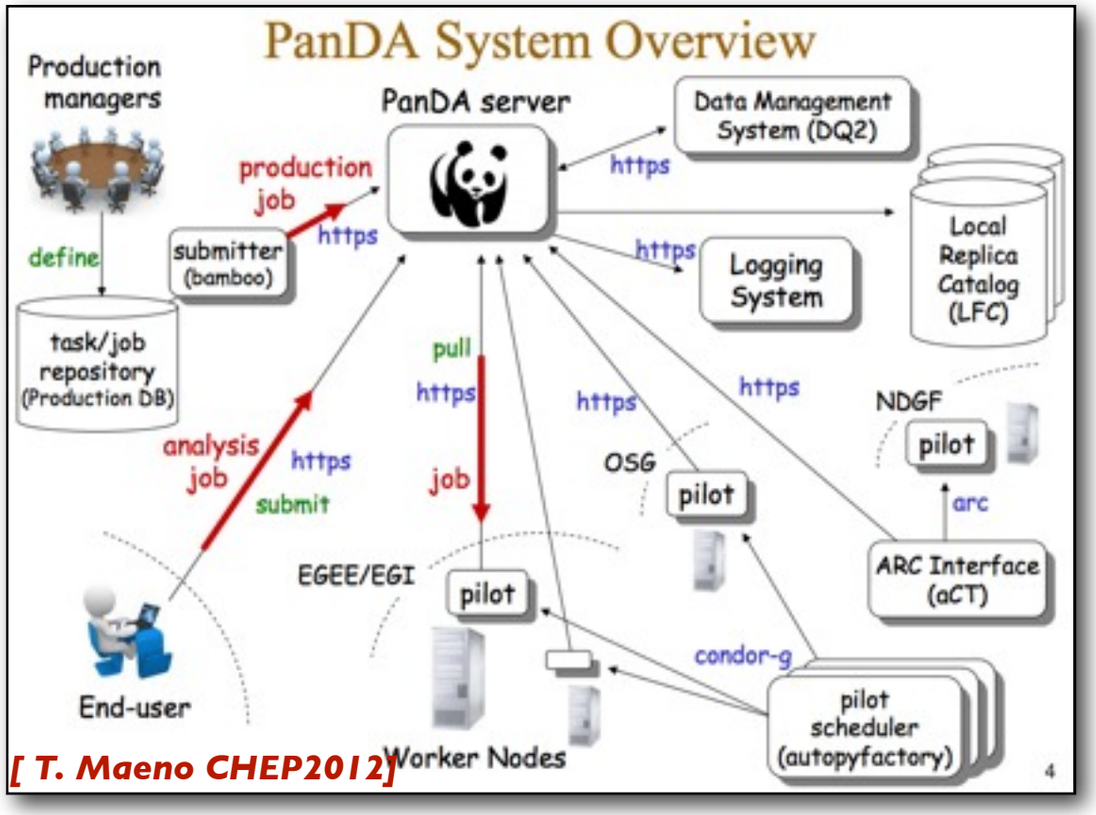
[A. Di Girolamo OpenStack User Group meeting]

CERN Flagship Results

Experiment Workflow

ATLAS workload management system is based on Production ANd Distributed Analysis (PanDA) system

- A homogeneous processing system layered over heterogeneous resources
- Use of Condor for job submission
- Use of pilot jobs for acquisition of processing resources.
- Support for both managed production and analysis



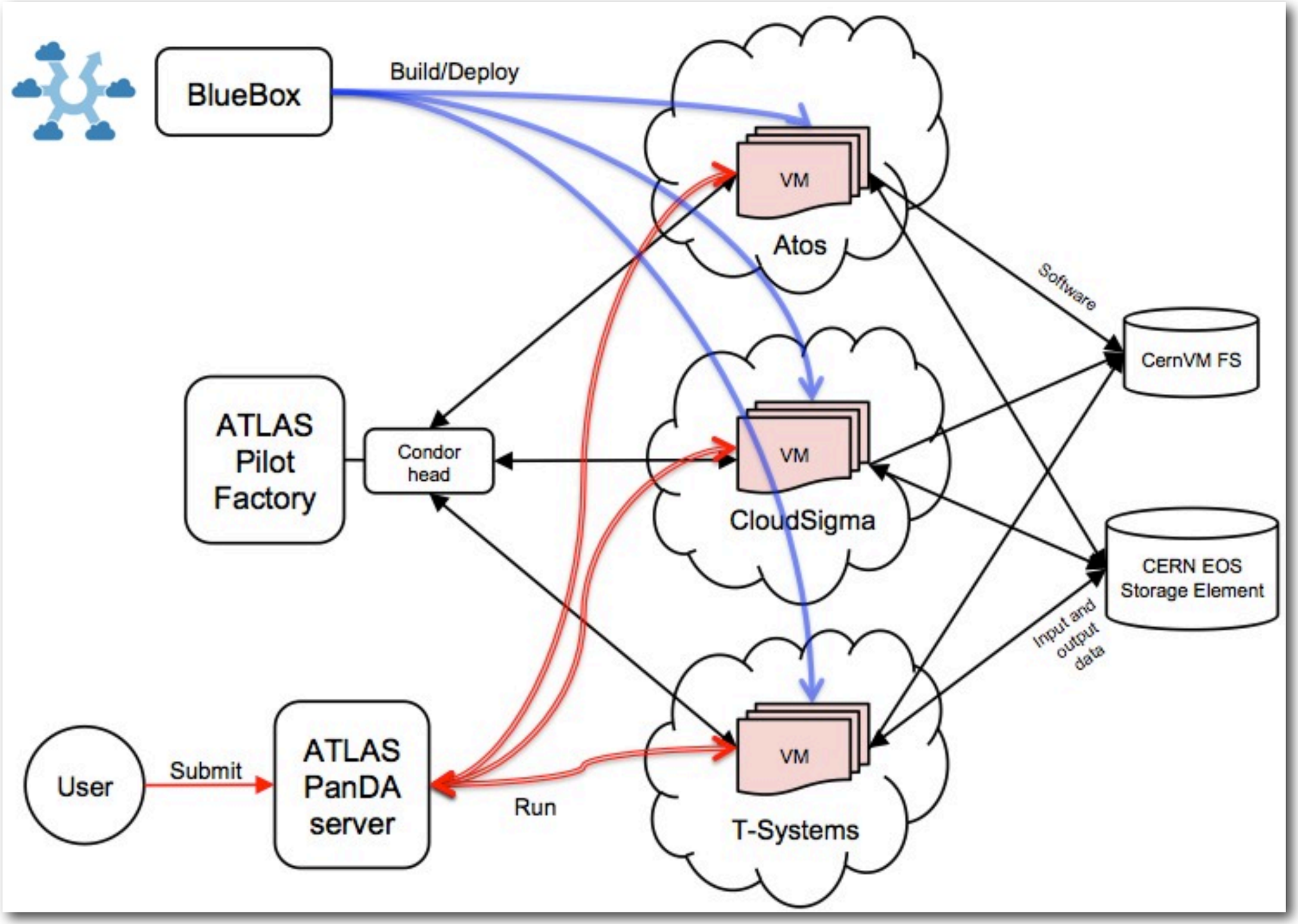
[T. Maeno CHEP2012]

Experiment workflow tested with Monte Carlo jobs

Geant4 based simulation of the particles propagation through the ATLAS detector

- Long (~4h), very intensive CPU usage, low I/O usage.
- Input: MC generator 4 vector files
- Output: ~50 MB/file of 50 events

Cloud Job Flow



Monitoring

Large variety of complementary monitoring views in order to track, log, cross-check, debug

- VM side (Ganglia), WMS side (PanDA, APF), BlueBox dashboard

Jobs:

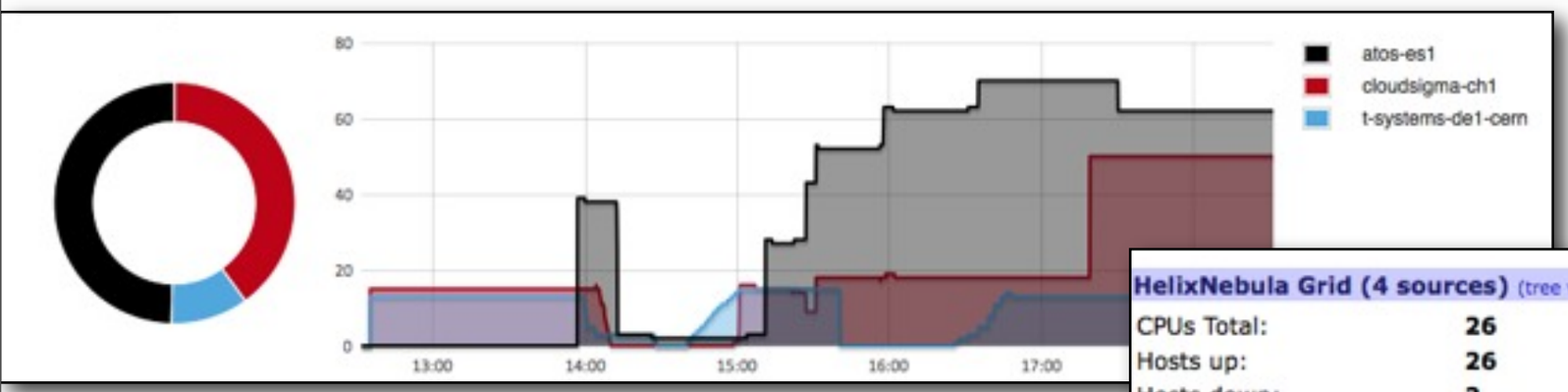
PandaID, Owner, Working group	Job	Status	Created	Time to start	Duration	Ended/ Modified	Cloud/Site, Type	Priority
2162262985 gangarbt	trans=AtlasG4_trf.py, pkg=AtlasProduction/17.2.2.2	finished	2014-05-11 10:51	1:48:55	6:15:36	05-11 18:56	CERN/CERN_HELIX_NEBULA_CloudSigma, prod_test	10000
	In: mc12_8TeV.107651.AlpgenJimmy_AUET2CTEQ6L1_ZeeNp1.evgen.EVNT.e1571_tid01001789_00 Out: hc_test.gangarbt.hc20035895.HELIX_NEBULA_CloudSigma.24							
2162262981 gangarbt	trans=AtlasG4_trf.py, pkg=AtlasProduction/17.2.2.2	finished	2014-05-11 10:51	0:11:06	6:17:37	05-11 17:20	CERN/CERN_HELIX_NEBULA_CloudSigma, prod_test	10000
	In: mc12_8TeV.107651.AlpgenJimmy_AUET2CTEQ6L1_ZeeNp1.evgen.EVNT.e1571_tid01001789_00 Out: hc_test.gangarbt.hc20035895.HELIX_NEBULA_CloudSigma.24							
2162262980 gangarbt	trans=AtlasG4_trf.py, pkg=AtlasProduction/17.2.2.2	finished	2014-05-11 10:51	0:10:19	6:58:51	05-11 18:00	CERN/CERN_HELIX_NEBULA_CloudSigma, prod_test	10000
	In: mc12_8TeV.107651.AlpgenJimmy_AUET2CTEQ6L1_ZeeNp1.evgen.EVNT.e1571_tid01001789_00 Out: hc_test.gangarbt.hc20035895.HELIX_NEBULA_CloudSigma.24							
2162262979	trans=AtlasG4_trf.py, pkg=AtlasProduction/17.2.2.2	finished	2014-05-11 10:51	0:08:03	5:02:24	05-11 16:02	CERN/CERN_HELIX_NEBULA_CloudSigma, prod_test	10000

BatchQueue (pandaq) view

Batch queue	HELIX_NEBULA_ATOS
WMS queue	HELIX_NEBULA_ATOS
Site	HELIX_NEBULA
State	test
Links	agis pandamon ssb

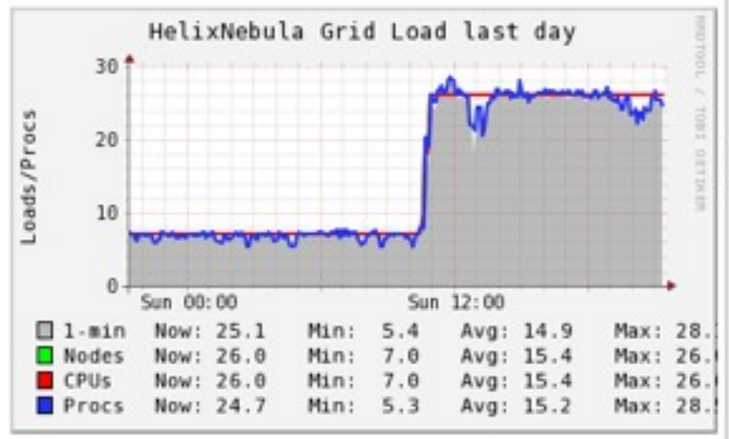
Label	factory	created	running	exiting	done	miss
HELIX_NEBULA_ATOS	aipanda002	165	7	0	0	0
HELIX_NEBULA_ATOS	aipanda009	1673	7	0	5	0
HELIX_NEBULA_ATOS	aipanda013	151	7	0	2	0

Factory	job	state	payload?	created
aipanda009	520291.0	created	-	seconds ago
aipanda002	884469.0	created	-	seconds ago
aipanda013	542345.0	created	-	seconds ago
aipanda009	520285.0	created	-	1 min ago
aipanda002	884463.0	created	-	1 min ago



HelixNebula Grid (4 sources) (tree view)

CPUs Total:	26
Hosts up:	26
Hosts down:	2



Note: Sorting by multiple columns at the same time can be activated by 'shift' clicking on the column headers which they want to add to the sort. Hovering mouse over the column headers to get descriptions of table columns.

Search:

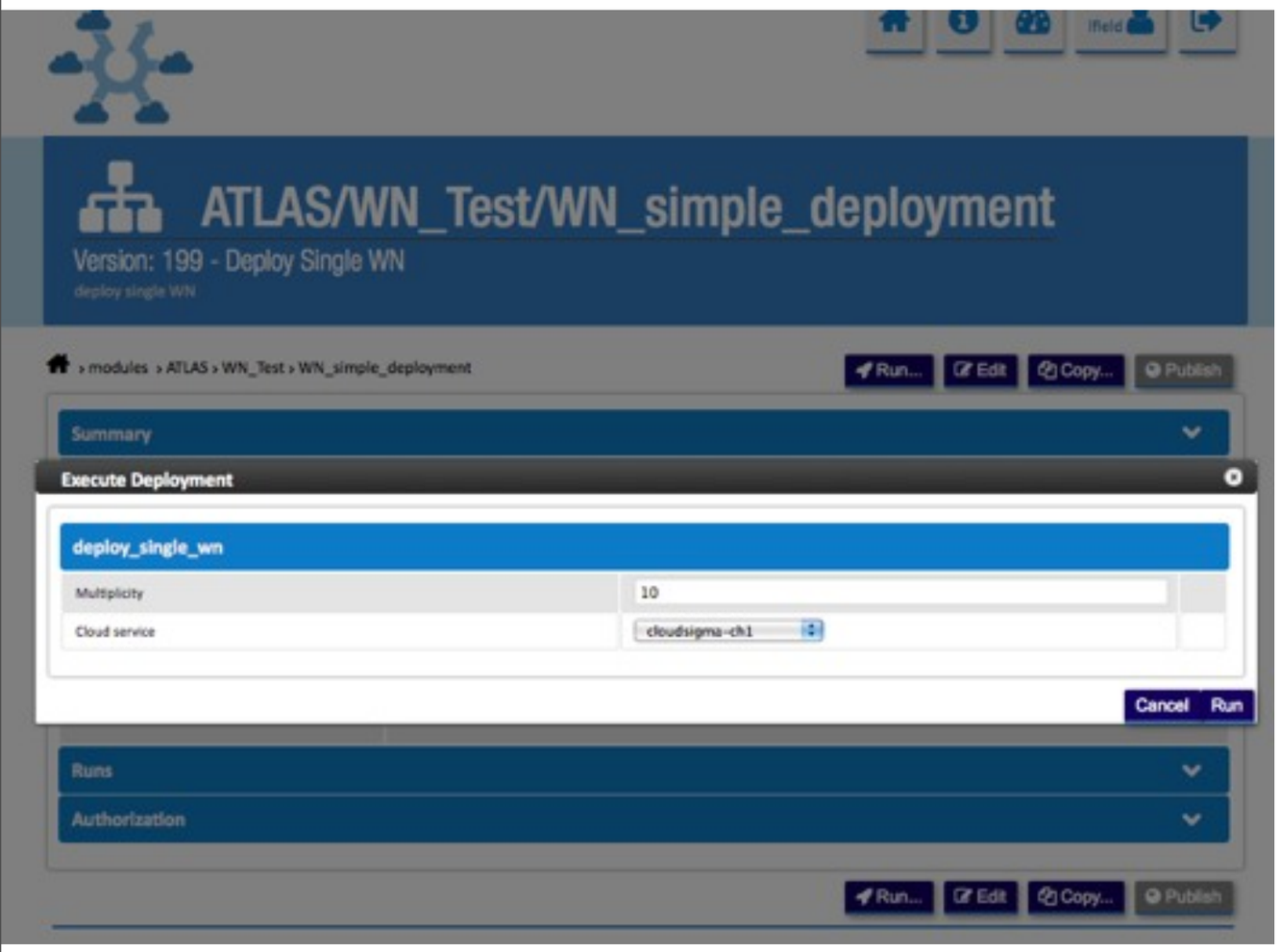
Site	Wall Duration	CPU Duration	CPU Count	Network inbound(Gb/sg)	Network outbound(Gb/sg)	Memory(GB)	Disk(GB)	Cloud Type
HELIX_NEBULA_ATOS	96.00	71.18	3.93	0.00	0.00	7.71	0	OpenNebula
HELIX_NEBULA_CloudSigma	0.00	0.00	0.00	0.00	0.00	0.00	0	OpenNebula
HELIX_NEBULA_TSystems	24.00	2.09	1.00	0.00	0.00	1.83	0	OpenNebula
Total:	120.00	73.27	4.93	0.00	0.00	9.54	0.00	

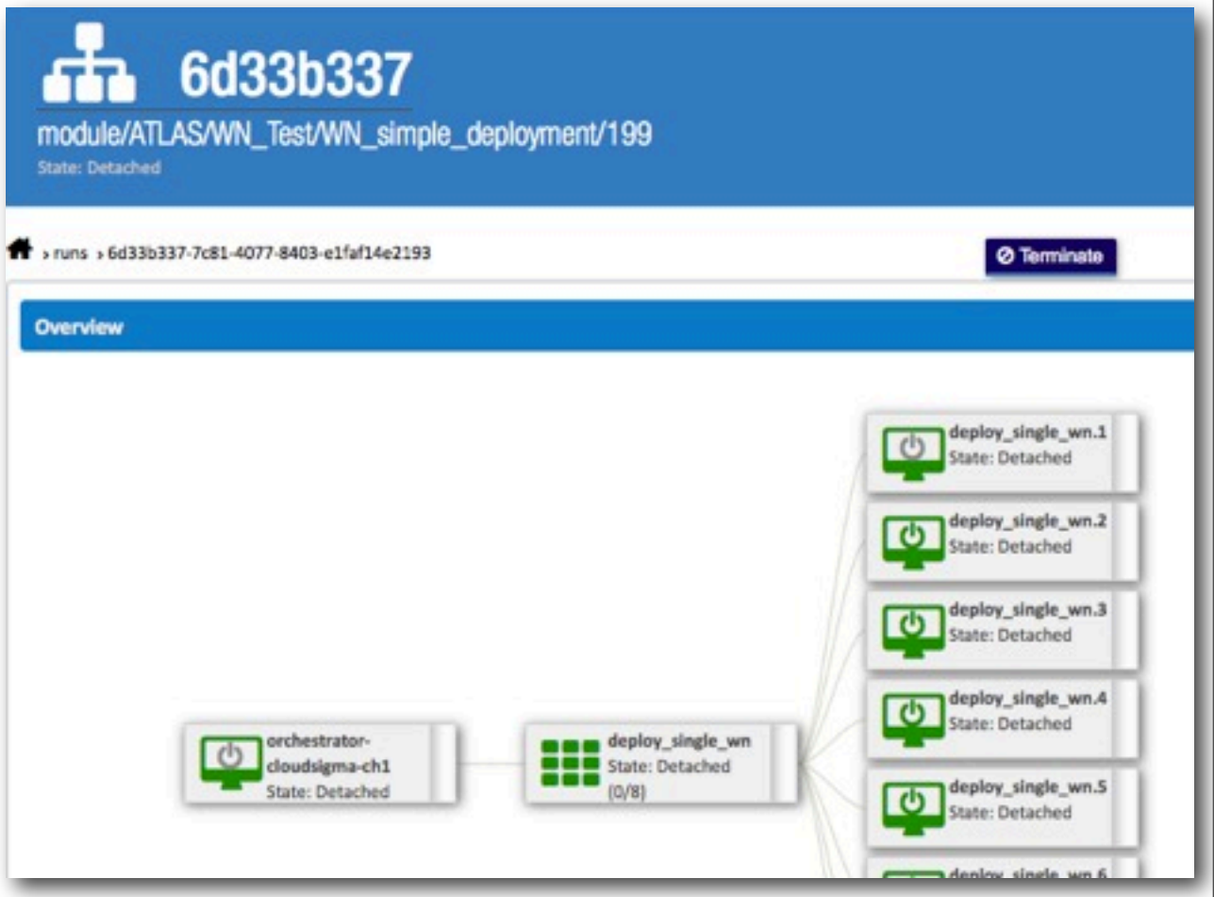
Showing 1 to 3 of 3 entries

Deployment

Multiple VMs started in a single deployment

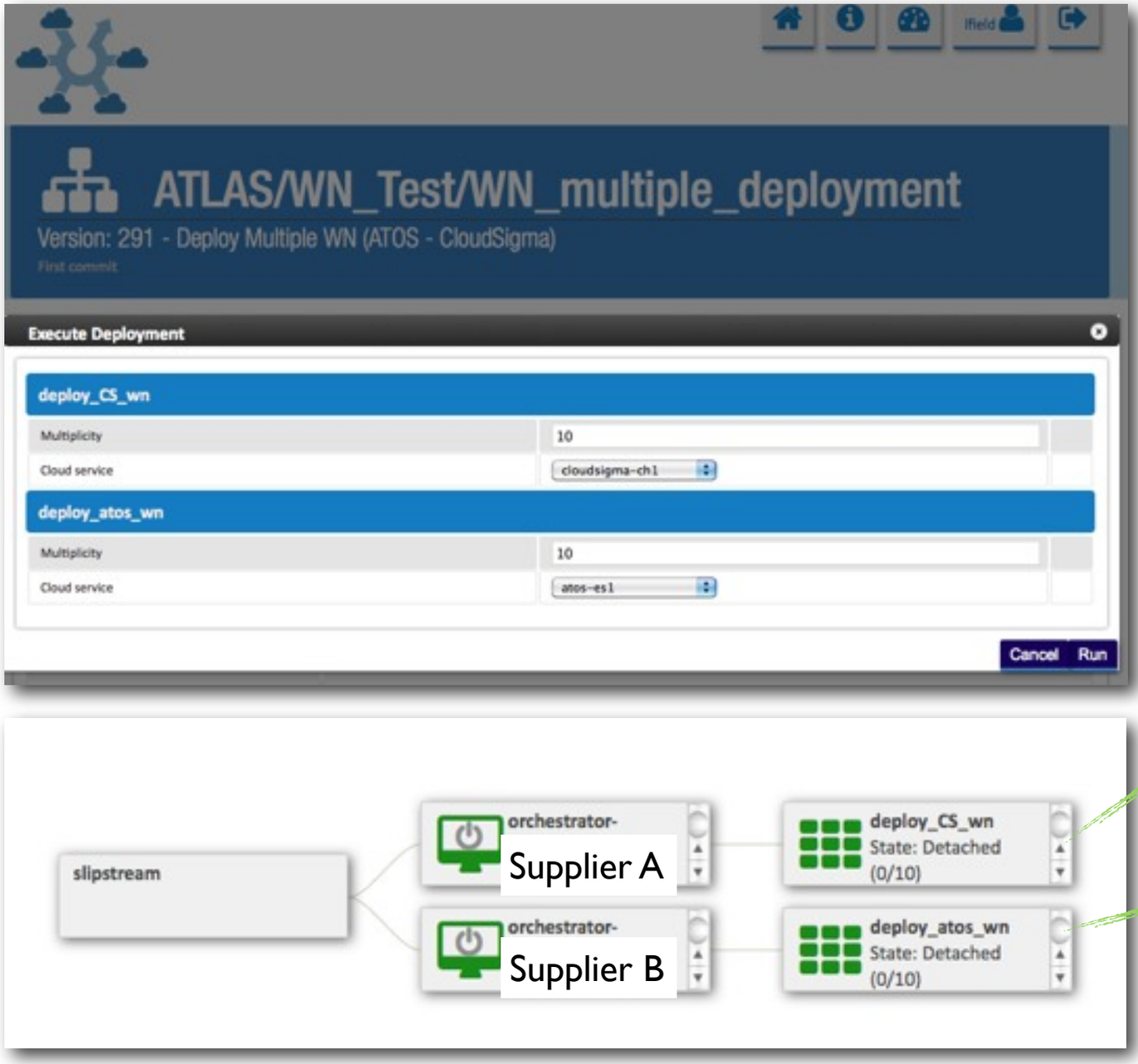
- Submitted up to 25 VMs per deployment

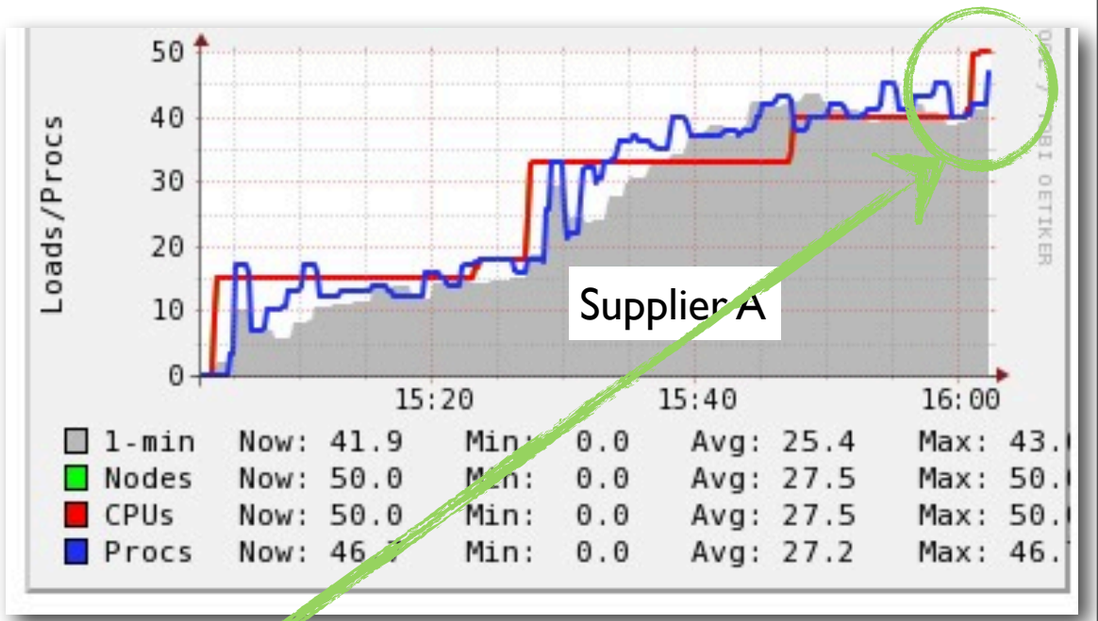


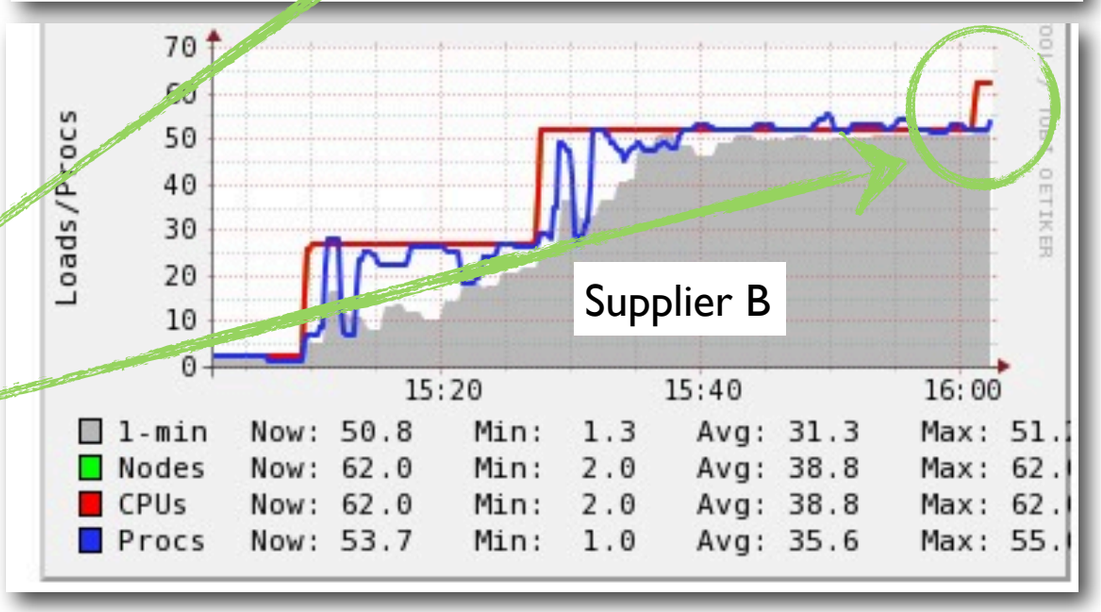


Deployment on multiple suppliers

Successfully deployed VMs in different suppliers through a single deployment

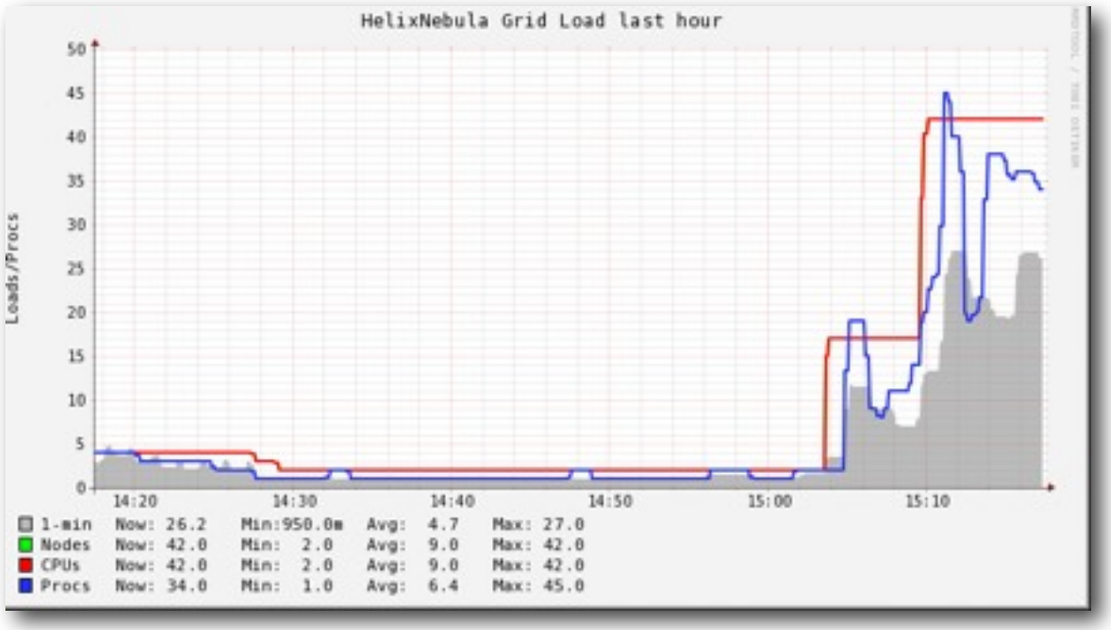


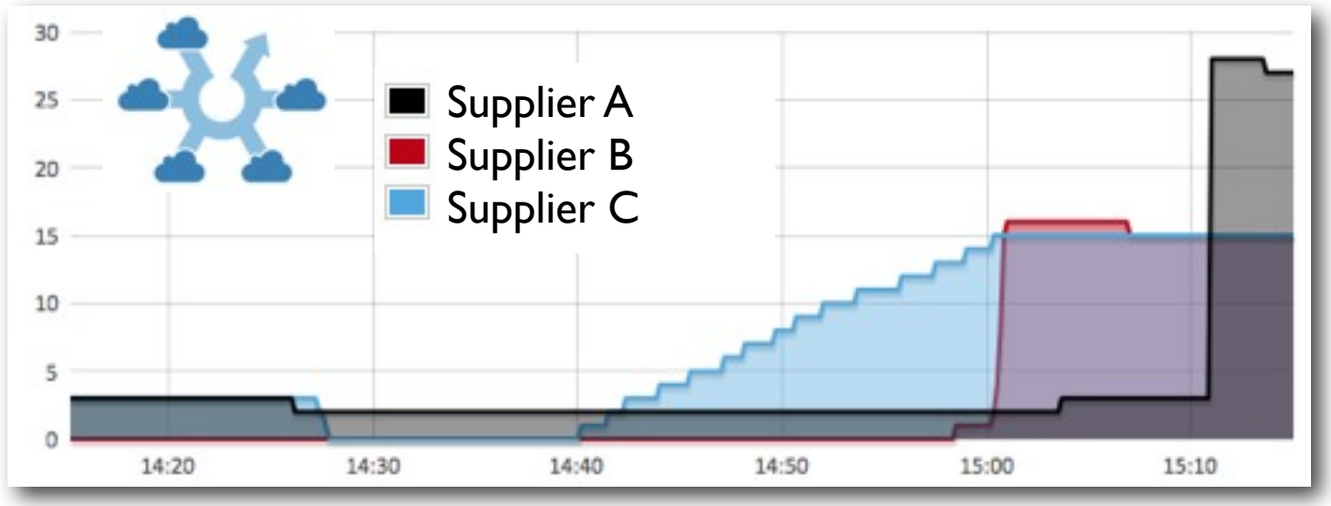




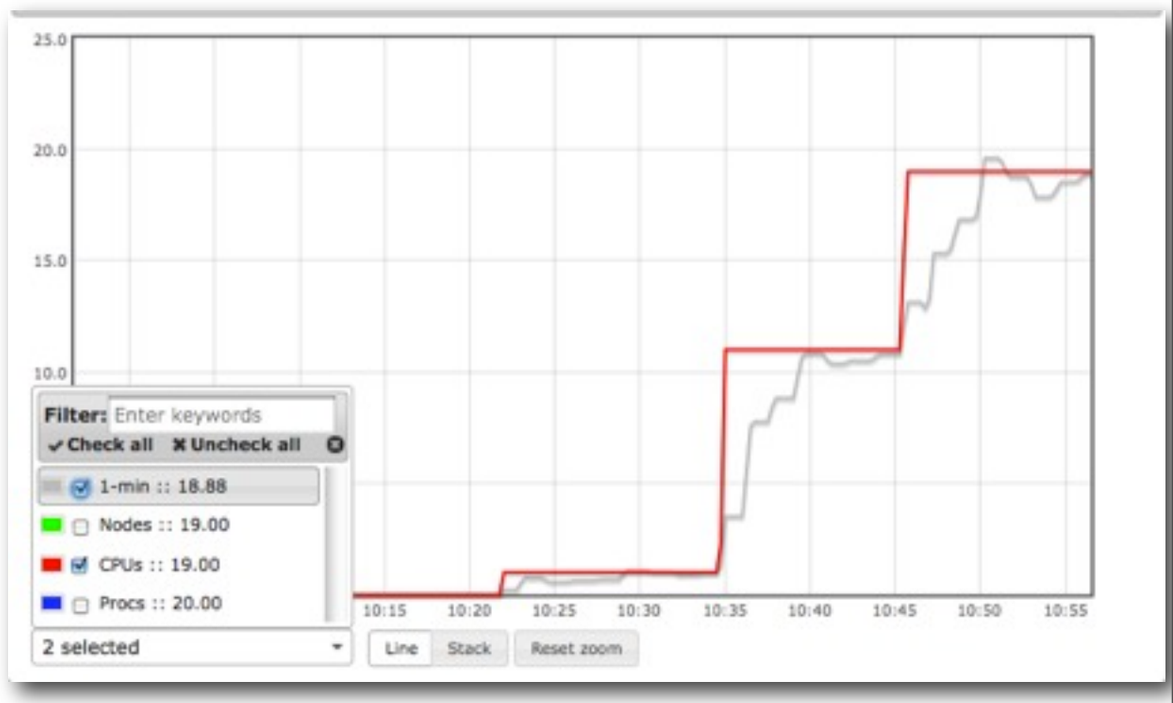
Ramping up

- Time to have VMs running from beginning of deployment depends on supplier
  - ~ 5′ - 25′

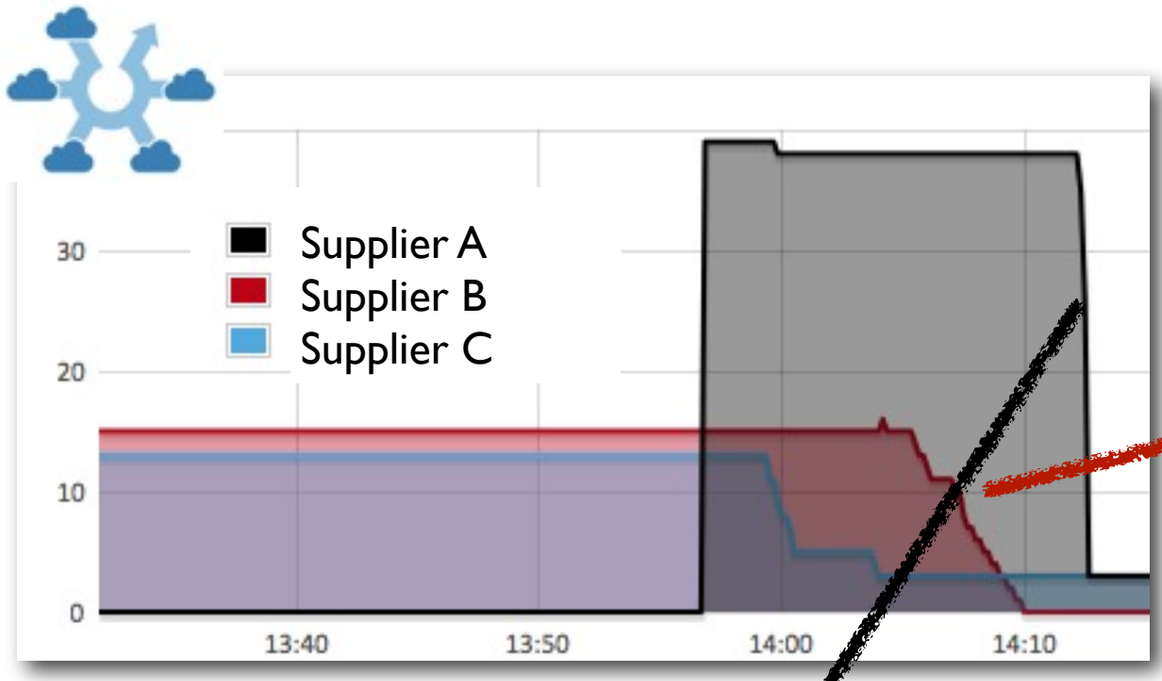


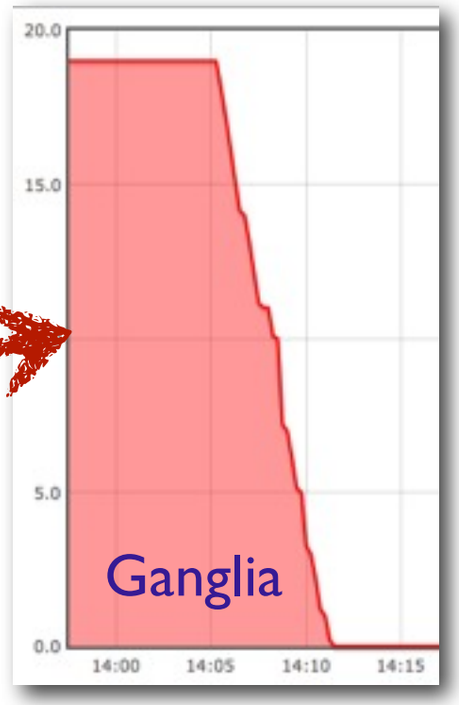


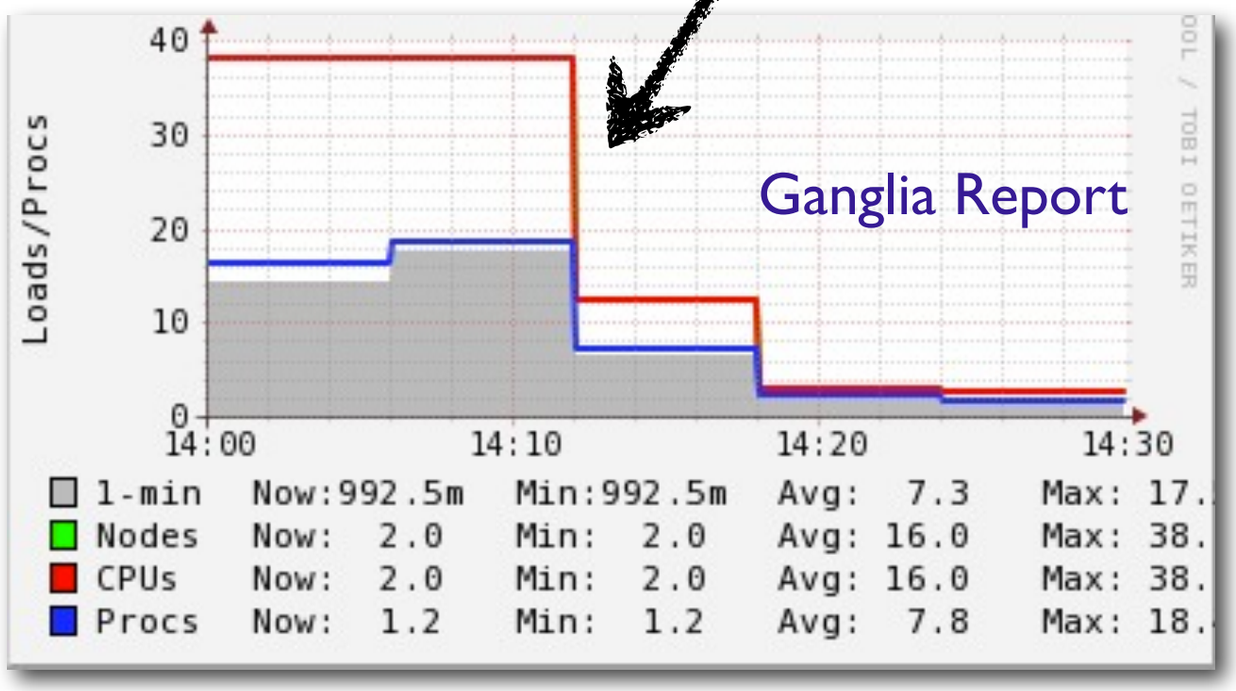
- Experiment jobs starts to run in O(5′) from the VM start



Terminate



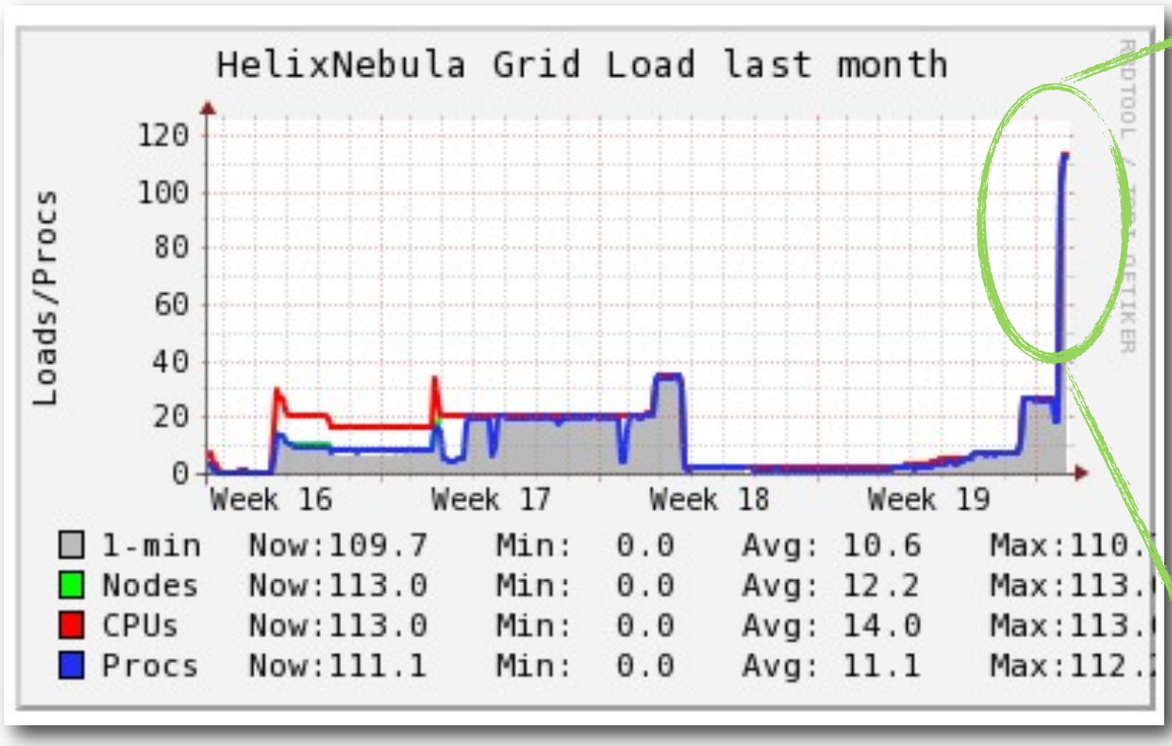


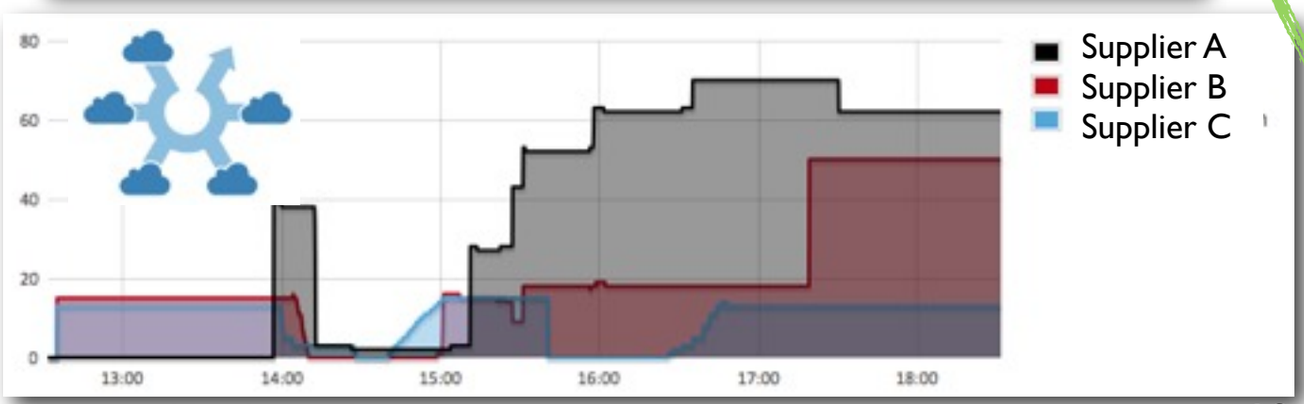


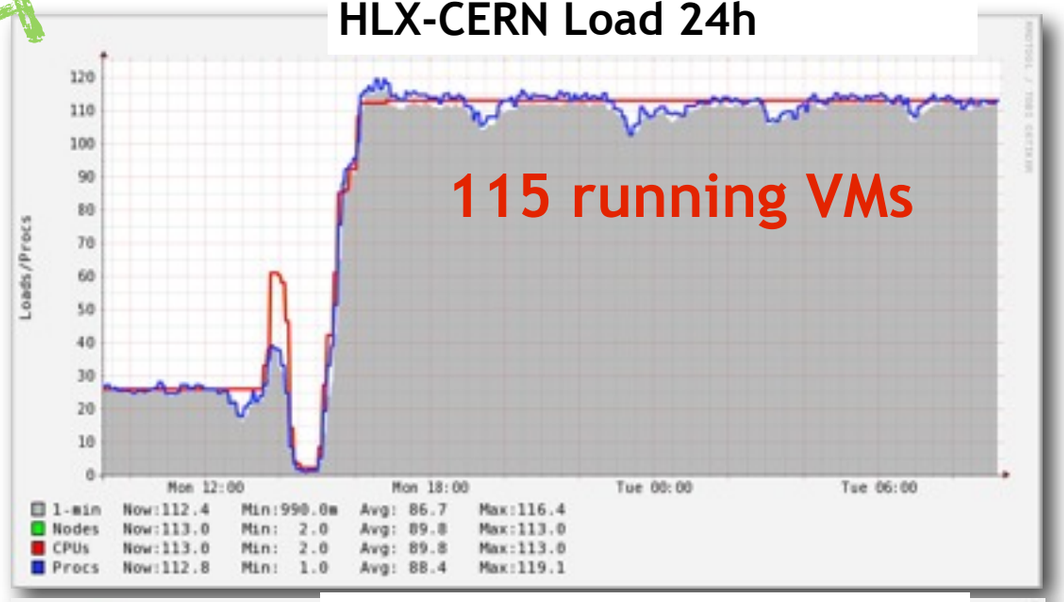
- Fast termination of machines in O(60″) from the “Terminate” command

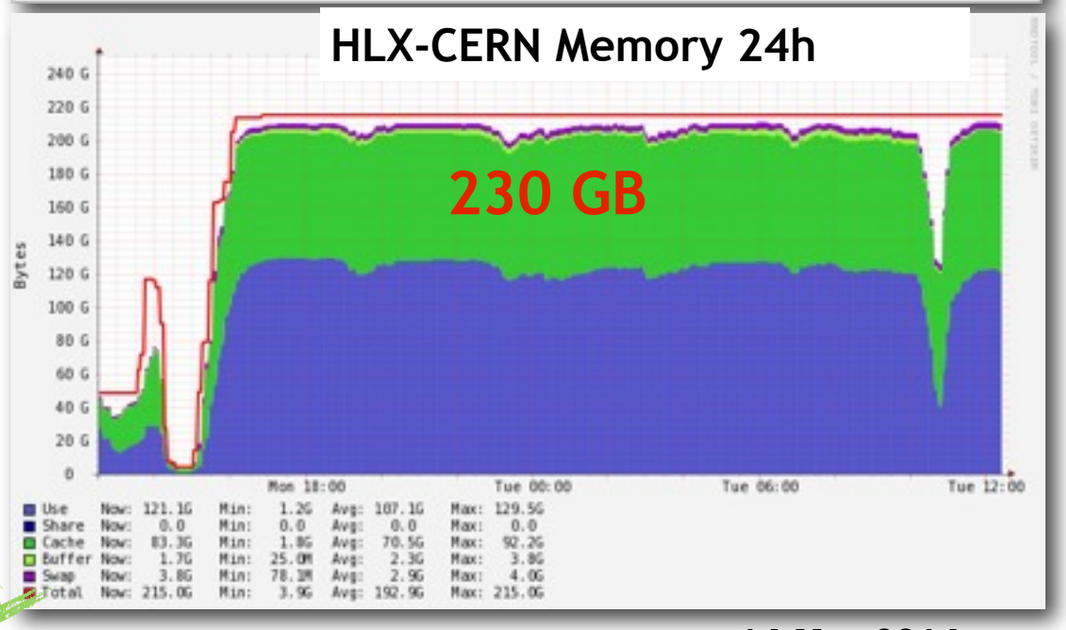
Scale tests

- Deployments show long time stability after startup
  - VMs left running for several weeks, running ATLAS functional tests
- Able to rapidly scale up to use available resources





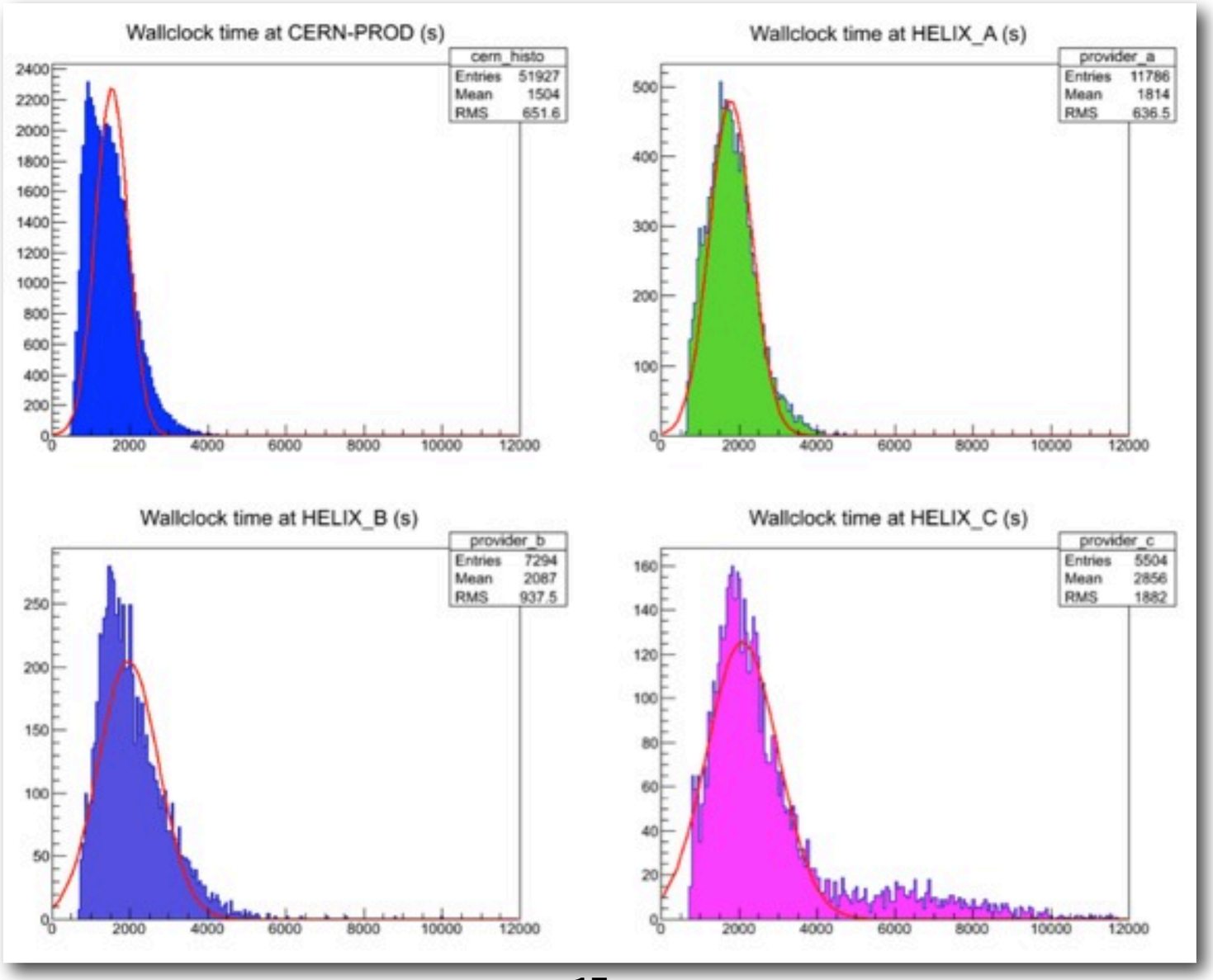




Past experience

Scale tests already performed in the past phases of the CERN flagship tests

- ~40k CPU days of processing during the pilot phase
  - Tests performed in 2013 connecting directly to each single provider (BlueBox was still not in the picture)



Conclusions

CERN computing use case is

- Embarrassingly parallel
  - Each collision event is processed independently from the others
- Huge size: process PB of data using O(10^5) CPUs

Helix Nebula CERN flagship deployed the ATLAS experiment workflow on a federation of European commercial cloud service providers

- Successfully tested primary functionalities: start/stop/status
- Successfully tested deployment of medium scale size: O(100) VMs
- Many lessons learnt in deploying experiment applications in cloud environment
- Good support received from all partners
- Encouraging experience on which to build up for the next phases