Helix Nebula: delivering cloud services to scientific researchers

Mick Symonds
Atos
e-IRG workshop
Amsterdam University
March 2016

Helix Nebula: the European cloud public-private partnership

HELIX NEBULA THESCIENCECLOUD

Strategic Plan

- Establish multi-tenant, multi-provider cloud infrastructure
- Identify and adopt policies for trust, security and privacy
- Create governance structure
- Define funding schemes

CERN

To support the computing capacity needs for the ATLAS experiment

EMBL

Setting up a new service to simplify analysis of large genomes, for a deeper insight into evolution and biodiversity

esa

cnes

To create an Earth Observation platform, focusing on earthquake and volcano research

PIC port d'informació científica

To improve the speed and quality of research for finding surrogate biomarkers based on brain images

Additional User

cnrs, ECMWF, ESO, ESRF, Neutrons for Science, INFN, Science & Technology Facilities Council, SCC, SURF SARA

Suppliers

Atos, CLOUDEO, CloudSigma, Cloudwatt, DataCentred, Deutsche Börse Cloud Exchange, egi, exoscale, GÉANT, interoute, Prologue., sixsq, SWITCH, T-Systems, the SERVER LABS, ultimum technologies

Adopters

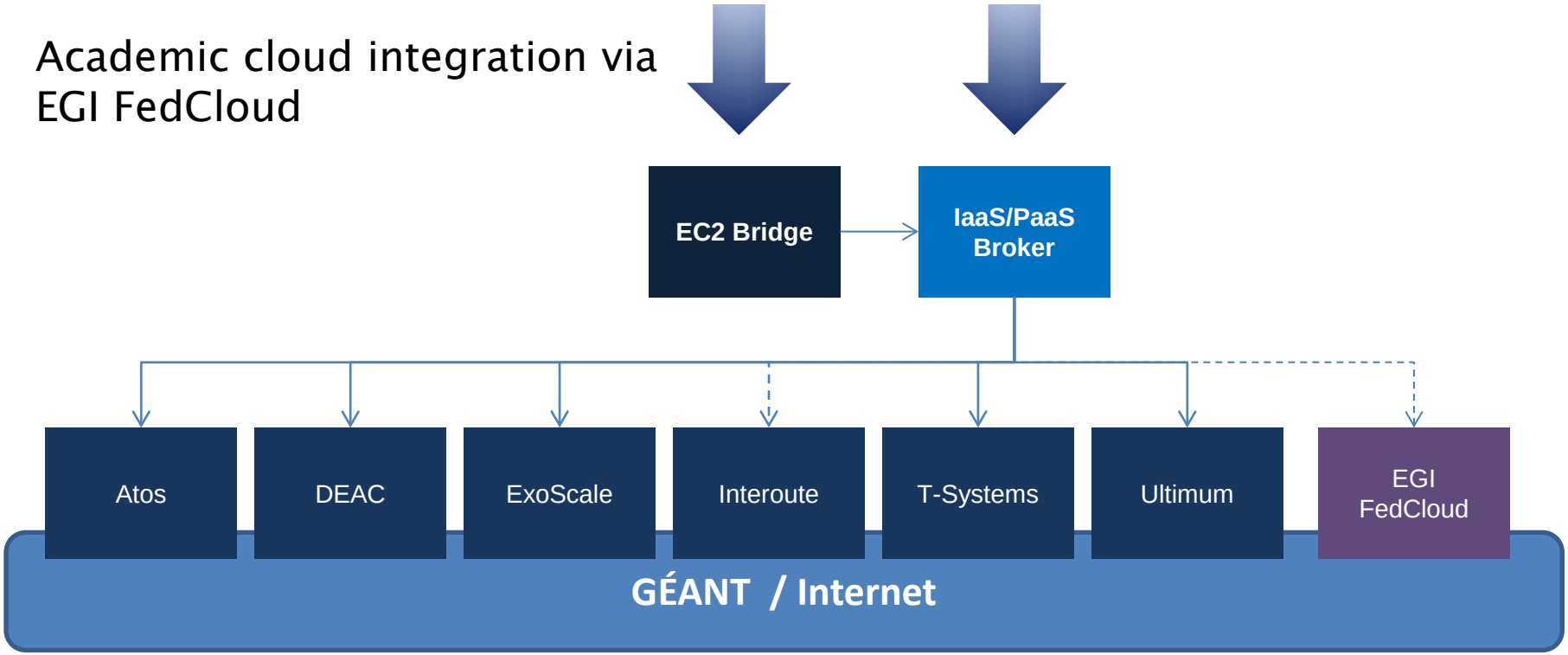
AWST, Capgemini, CSA cloud security alliance, deac, Ifremer, ITER, MEMSET, orange Business Services, RHEA, terradue, THALES, Trust-IT, vito, VU, Yandex

Helix Nebula Marketplace

- Builds upon the work undertaken as part of the EC funding project
- Form hybrid Cloud Computing Marketplace by integrating European cloud providers and existing e-Infrastructures
- Focus brokerage function on transparency, trust and impartiality
- Deliver trusted cloud services through compliance with EU regulations and legislation
- Simplified procurement process for multi-cloud services
- Offered to the global scientific community, including publicly-funded and commercial research and technology organizations, offering large-scale and HPC-type deployments from the start

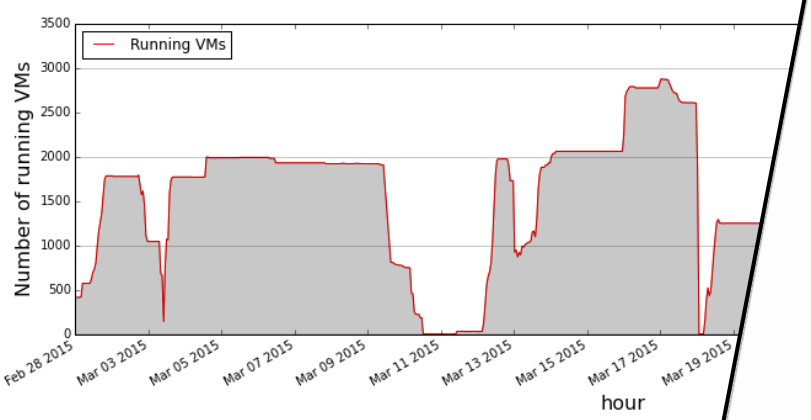
HNX Platform

- Unified access to federated IaaS resources via IaaS/PaaS UI/API
- Amazon EC2 Bridge for compatibility with any EC2-compatible third party tool such as StarCluster
- Academic cloud integration via EGI FedCloud



CERN findings on commercial cloud services

- CERN purchased resources to run LHC ATLAS experiment software
- Atos (data centre in Tenerife, Spain) was award
  - accessed via the SlipStream broker and provision
  - and delivered over the GEANT network
- >1 million CPU hours were successfully provid
  - confirming that the use of commercial IaaS is te



21st International Conference on Computing in High Energy and Nuclear Physics (CHEP2015) IOP Publishing
Journal of Physics: Conference Series **664** (2015) 022019 doi:10.1088/1742-6596/664/4/042019

Accessing commercial cloud resources within the European Helix Nebula cloud marketplace

C Cordeiro[1], A De Salvo[2], A Di Girolamo[1], L Field[1], D Giordano[1], R Jones[1] and L Villazon[1,3]
[1] CERN, Information Technology Department, Geneva, Switzerland
[2] Università & INFN Roma I, Rome, Italy
[3] Universidad de Oviedo, Oviedo, Spain
E-mail: domenico.giordano@cern.ch

Abstract. Helix Nebula – the Science Cloud Initiative – is a public-private partnership between Europe's leading scientific research organisations and European IT cloud providers. CERN contributed to this initiative by providing a flagship use case: the workloads from the ATLAS experiment. Aiming to gain experience in managing and monitoring large-scale deployments, as well as in benchmarking the cloud resources, a sizable Monte Carlo production was performed using the Helix Nebula platform. This contribution describes the Helix Nebula initiative and summarises the experience and the lessons learned from deploying ATLAS experiment application within large cloud setups involving several commercial providers.

1. Introduction

Helix Nebula – the Science Cloud Initiative – is a public-private partnership between Europe's leading scientific research organisations (CERN, EMBL and ESA) and European IT cloud providers, that aims to establish a cloud computing platform for data intensive science within Europe. Over the past three years, the Helix Nebula initiative has built a federated cloud framework – the Helix Nebula Marketplace (HNX) [1] – to provision cloud services from a range of commercial cloud providers and public e-infrastructures, such as EGI [2] and GÉANT [3]. HNX delivers access to cloud resources through a broker technology deployed by the Helix Nebula initiative. In addition it aims to comply with EU regulations and legislation to provide trusted cloud services.

CERN contributed to this initiative by providing a flagship use case: integration of Helix Nebula cloud resources within the workload management system of the ATLAS experiment. Aiming to gain experience in managing and monitoring large-scale deployments, as well as in benchmarking the cloud resources, a sizable Monte Carlo production was performed using the Helix Nebula platform. Thousands of concurrent virtual machines (VM) were run for four weeks in order to be comparable with the capacity supplied by a typical WLCG Tier-2 site.

2. Cloud resource provisioning

Figure 1 shows the architecture of the set-up used to provision resources within HNX, to steer ATLAS experiment workflows to the running VMs and to monitor the VM status and usage accounting. For the specific purpose of processing ATLAS Monte Carlo simulation events, the

- CERN see the value (rationalisation of IT
  opportunities (expand the impact of their
  cloud services
- The procurement and use of cloud serv
  and procedural questions to public organisations

nd
l
actual

Helix Nebula – learnings to date, and prognosis

March 2016
Mick Symonds
Atos-NL
HN Suppliers Chair

- **This was a development has taken much, much longer than we expected**
- **It started off as a technical challenge**
  - on the technical front, the biggest challenge was standardising interfaces, for which we deployed a “Blue Box” as a common front end
  - but quickly the service and business issues came to predominate
- **Scientific research is a greenfield market for IT Cloud service provision**
  - but suppliers have become, and remain, frustrated by the lack of take-up
  - research institutes are used to have their own dedicated infrastructure
    - scientists are not (good) customers: they do not (expect to) pay for their use of IT
  - funding and procurement models are not really suitable for cloud services
    - a project (PICSE) has been launched to address that: http://www.picse.eu/
- **Most researchers need a combination of at least three services:**
  - ad hoc and “bursty”, on-demand processing facilities
  - large-scale, long-term data storage and archiving
  - a platform for deriving valuable information on top of these infrastructure services
- **Data storage provision and location are more of an issue than processing capacities**
  - although storage and processing requirements should be addressed together to optimise performance and network costs
  - regular cloud storage services (nn cents per GB per month) are not really suitable
  - data protection and intellectual property rights need to be incorporated

Governance and cloud services

March 2016
Mick Symonds
Atos-NL
HN Suppliers Chair

- Most commercial enterprise organisations have developed a mature approach to how they manage IT services over the last decade or two
  - a move away from technology orientation and managing boxes
  - towards services and their management
- This has involved a number of changes in their management:
  - cost awareness:
    - identifying the costs per system, department and business process
  - demand/supply models:
    - at defined levels, with regular meetings and reports, to bridge from end users to suppliers, in a structured manner
  - identifying who is responsible for the supply of service to the end users
  - addressing the “turkeys and Christmas” syndrome
    - the IT Manager whose status depends upon assets and resources (data centres, staff) reporting directly to them
- Research organisations might do well to try to learn from their experience

Data protection,
with the USA Patriot Act as the prime issue

March 2016

Mick Symonds
Atos-NL
HN Suppliers Chair

- The USA Patriot Act and FISAA mean that US companies and individuals are compelled to reveal their customers' data to the
  - wherever in the world it is located
  - and without (being able to) tell the customer
- The biggest problem for international services
  - it is fear, uncertainty and doubt (FUD)
    - managers don't know whether they can
    - exacerbated by the Snowden effect and
      - and Privacy Shield doesn't seem to h
  - combined with a hint of chauvinistic prote
- When it comes to data protection, internati ee whether and how much real restrictions ex
  - so the average business manager hasn'
  - "keep it in my country, or even data ce

This does not constitute legal advice

Cost calculations

March 2016
Mick Symonds
Atos-NL
HN Suppliers Chair

- Many organisations undertake a simple exercise to determine their storage costs
  - they take the retail cloud rate (e.g. from Amazon), multiply it by the volumes and time involved
  - and compare that to the cost of buying their own storage devices
- Often they analyse their own costs in an incomplete manner
  - e.g. they ignore electricity cost, because they do not pay those bills
    - this can also apply to staffing and space costs
  - and they end up with a marginal cost for these basic facilities
- Within the EC, the e-Fiscal project was established to address this
  - see http://www.efiscal.eu/
  - but is so far neither completely developed and/or applied (afaik)
- This issue is well-known within outsourcing
  - we are well-accustomed to undertaking a consultancy exercise to help the customer understand their own true costs and how they compare to commercial provision

How to address these issues?

March 2016
Mick Symonds
Atos-NL
HN Suppliers Chair

- Most such issues were identified and addressed in the past within the context of providing outsourcing services to enterprises*
  - those techniques need to be dragged out, polished up and re-applied
  - and a manner found of getting the potential customers to address them

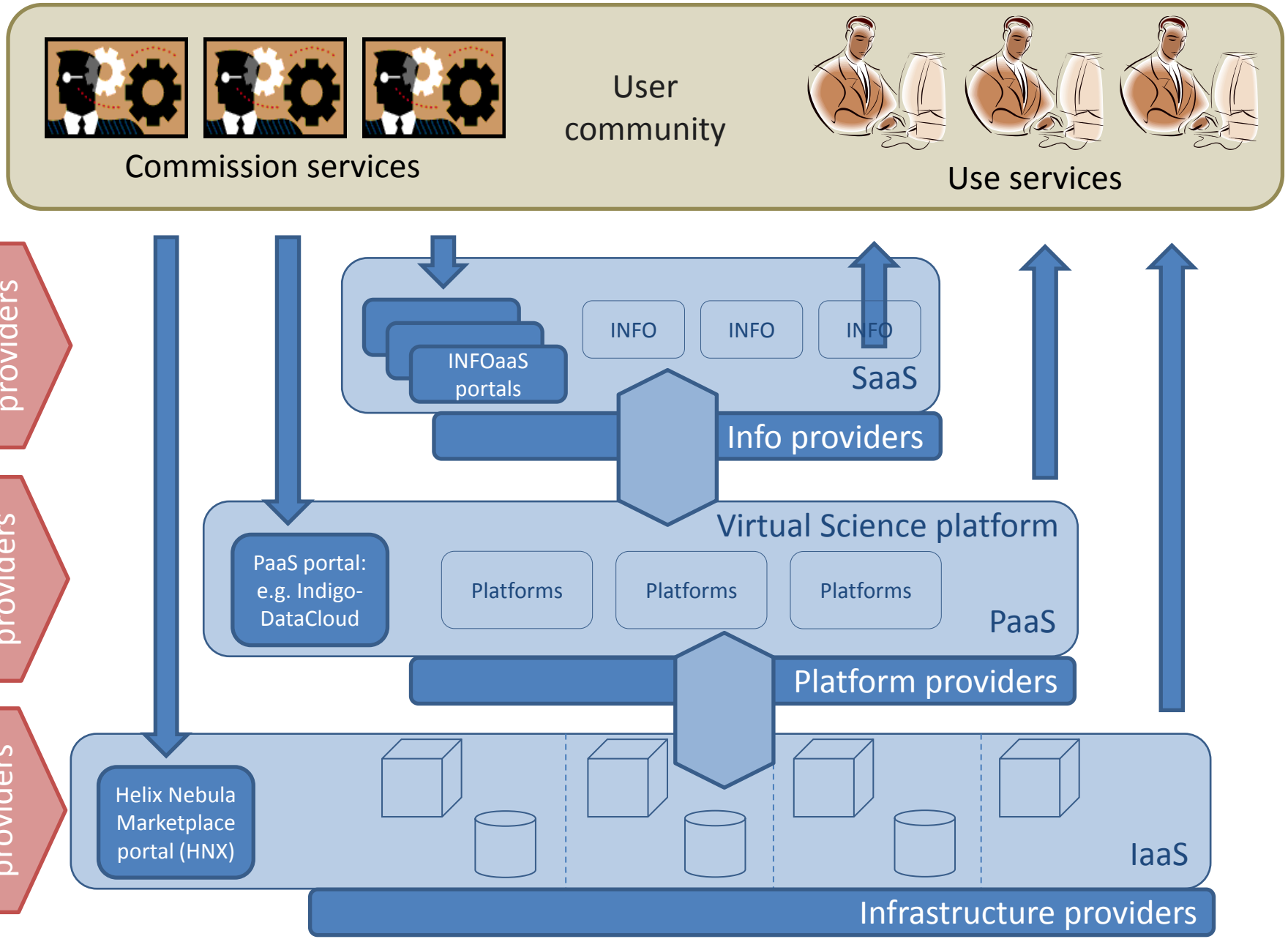
- A hybrid approach may be required:
  - deliver large volumes of storage on an outsourcing-type basis
    - managed by a data curation service (as per EUDAT)
  - alongside processing on a variable, true cloud basis
  - with a transaction-based Information service on top

- * See *Outsourcing: the Atos Origin outsourcing lifecycle – building successful outsourcing relationships*, Erik Beulen, Rene Baas, John Dain, Joyce Hudson, Ebe Reitsma, Mick Symonds and Han van der Zee, Atos Origin, 2004

Suppliers are trying to add more value,
such as the Copernicus data access proposal

March 2016
Mick Symonds
Atos-NL
HN Suppliers Chair

- In response to the European Commission high level presentation of options for Copernicus access platform(s)
  - provide a basis for the provision of a full end-to-end information as a service (INFOaaS) ecosystem
  - an opportunity to take a major step forward in deriving valuable and beneficial information from the available data
    - while safeguarding intellectual property
- By working in a federated, competitive environment, and by giving this initiative a “jump start” with our existing services to reduce time-to-market, European industry can provide both the services required and also considerable further benefits
- Europe would be endowed with high added-value assets, provided by European actors, rather than non-European ones



Source: Atos, based on CloudEO and Thales models

Documents available

March 2016
Mick Symonds
Atos-NL
HN Suppliers Chair

- White Papers:
  - Shaping the cloud: why and how you should develop a cloud strategy now: https://atos.net/content/dam/global/we-do/atos-shaping-the-cloud-white-paper.pdf
  - Helix Nebula Service Architecture: http://www.helix-nebula.eu/publications/reports/service-architecture
  - Outsourcing: the outsourcing lifecycle (Atos, 2004)
  - Establishing a Trusted Cloud Europe: http://ec.europa.eu/information_society/newsroom/cf/dae/document.cfm?doc_id=4935
  - Trusted European Cloud: http://atos.net/content/dam/global/documents/we-do/atos-white-paper-trusted-european-cloud.pdf

Contact

March 2016

Mick Symonds
Atos-NL
HN Suppliers Chair

Mick Symonds
Principal Solutions Architect/Loose Cannon
Atos
The Netherlands

michael.symonds@atos.net
m +31 651 755 779