Helix Nebula – The Science Cloud

Blue Box Approach

Abstract: This document captures the current knowledge of the Helix Nebula Supply-side regarding the need of a federated framework to simplify discovery, access, usage and management of a federated cloud system. Alongside this objective, we aim at providing an integration framework, where current and future suppliers (i.e. cloud service providers) can easily interface their system in order to attract and receive cloud workload.

Document Information Summary	
Deliverable title:	*Helix Nebula Blue Box Approach*
Editor:	*Jurry de la Mar*
Contributing Authors:	*Phil Evans* *Mick Symonds* *Robert Jenkins* *Marc-Elian Begin* *Jonathan Graham* *Paul Parsons* *Jurry de la Mar*
Distribution:	*Public*
Version/Revision:	*1.6p*
Keywords:	*Blue Box*

Disclaimer

Helix Nebula targets scientific communities, decision makers from SMEs, industry and government as well as European Union policy makers. This document contains information on Helix Nebula core activities, as prepared by the Helix Nebula Supply-side. Any reference to content in this document should clearly indicate the authors, source, organisation and publication date.

This document is work in progress. The authors reserve the right to significantly change the underlying assumptions, system and specification described in this document.

Log Table

Issue	Date	Description	Author/Partner
V1.0	12/11/2012	Draft	de la Mar
V1.1	19/11/2012	Updates	de la Mar
V1.2	19/11/2012	Updates	Symonds, Begin, Jenkins
V1.3	28/11/2012	Updates	de la Mar
V1.4	03/01/2013	Updates	Begin, Graham, de la Mar
V1.5p	11/02/2013	Issued draft public version for review	de la Mar
V1.6p	22/02/2013	Issued final public version	de la Mar

Table of Contents

1. Introduction

This document captures the current knowledge of the Helix Nebula Supply-side regarding the need for a federated framework to simplify discovery, access, usage and management of a federated cloud system. Alongside this objective, we aim at providing an integration framework, where current and future suppliers (i.e. cloud service providers) can easily interface their system in order to attract and receive cloud workload.

This document has been written on request of the Helix Nebula demand-side to be able to assess, support and agree on the approach taken for the Pilot Phase of the Federated Cloud Services and Blue Box development. The document addresses the following topics:

- List of possible solutions that were reviewed by the supply-side
- A gap analysis for each of the proposed solutions
- A plan for how the solutions can be implemented and delivered during 2013
- The role of each partner and the resources they will be committing
- The services that will be offered via this federated cloud
- Budgetary estimates for the pilot phase
- Scale of sustainable Ecosystem for Science Cloud Computing.

The document was drafted and shared with the demand-side for iterative feedback. The content has enabled the demand-side to take a decision regarding the implementation of the Pilot Phase and demonstration of Blue Box solutions during the 2nd Helix Nebula General Assembly from 15-18th January in Frascati.

This document is structured as follows: Section 1 is this introduction. Section 2 describes the potential solutions that were analysed to support the Blue Box requirement. Section 3 contains the Gap Analysis approach and summary results. Section 4 describes the implementation plan for the Blue Box. Section 5 contains the supply-side commitments for each participating party. Section 6 is a summary.

2. Potential Blue Box Solutions

This section describes the potential Blue Box solutions that were reviewed by the partners.

Following Solutions/Building Blocks have been reviewed:

- Bonfire
- Computenext
- enStratus
- Nephos
- OpenNebula
- Slipstream

The BonFIRE FP7[1] project is designing, building and operating a multi-site cloud-based FIRE facility to support research across applications, services and systems targeting in particular but not exclusively the research community on Internet of Services. The BonFIRE vision is to give researchers in these areas access to a distributed cloud facility which supports experimentation of their systems and applications and the evaluation of cross-cutting aspects.

Computenext[2] is building the next generation of compute power through the launch and operation of a federated cloud marketplace, leveraging true multi-cloud connectors, to enable consumers and providers of cloud services to effectively participate in the search, purchase, and utilization of the most efficient cloud resources available.

enStratus[3] enables users to deploy and manage enterprise-class applications across private, public and hybrid clouds. enStratus is available as software-as-a-service, or as on-premises software that enables users to control the cloud from within their own data centers. Today, enStratus is able to support a wide range of cloud providers.

Nephos Technologies[4] is an independent Cloud Services Brokerage (CSB); providing consultancy, management and technology services to organisations seeking to develop, implement and optimise their cloud strategy.

OpenNebula[5] is an open-source project developing the industry standard solution for building and managing virtualized enterprise data centers and IaaS clouds. It is distributed and licensed for use under the terms of the Apache License, Version 2.0. The OpenNebula open source project is being managed by C12G Labs.

[1] www.bonfire-project.eu
[2] www.computenext.com
[3] www.enstratus.com
[4] www.nephostechnologies.com
[5] www.opennebula.org

SlipStream™[6] provides automated provisioning of virtual machines and coordination of service deployment, as well as image creation. SlipStream™ supports a growing number of IaaS cloud solutions, providing a simpler access to clouds. The solution is released under the Apache 2.0 open source license[7]. SlipStream™ supports a number of interfaces, including a REST interface, a portal, a CLI and a preliminary EC2 interface.

3. Gap Analysis

This section describes how the gap analysis has been performed and the evaluation results.

Based on the Preliminary Requirement Analysis done by the TechArch Working Group and the feedback received from the demand-side Organisations, the functional requirements for the Blue Box have been verified for each potential solution. It was evaluated if a solution clearly met the requirements based on current status, how it met the requirement and if a gap existed. Details of this technical and functional analysis are confidential to the Helix Nebula members at this point in time.

Other non-technical aspects that were reviewed by the supply-side were:

- maturity of the solution (documented by reference customers),
- schedule to implement the solution,
- potential Helix Nebula branding of the solution
- origin of the solution, in particular its “European” character
- based on open or proprietary standards
- closed/proprietary or Open Source software
- constraints regarding future use of the system:
    - If any IPR and licensing aspects would apply
    - If any security aspects need to be considered e.g. US regulations
- overall costs of service support for the Pilot Phase.

As has been pointed out in the recently made available documents from the Helix Nebula ServArch Working Group, the technical aspects only form the basis for further evaluation purposes. Rather than just describing how something is delivered technically demand-side requirements should be expressed in terms of services and the verifiable, comparable service levels, which the supply side commits to deliver to the demand side, and for what cost per unit. And there are aspects way beyond just enabling the use of Infrastructure as a Service (IaaS), which is anyway usually a means to an end rather than an end in itself. E.g. a key objective of Helix Nebula is to create a ‘virtual Market Place’, where data, scientists, funding bodies, SMEs and downstream industry are meeting to work along common interests. This ecosystem should represent a many-to-many relationship, quickly being established, jointly to transform data to valuable

[6] www.sixsq.com

[7] Available at https://github.com/slipstream

information. This "bigger picture" necessitates not only higher-level cloud services themselves (i.e. PaaS and SaaS) but, more importantly, a way for very different organisations to work together seamlessly to deliver those services.

The eventual selection of a "best fit for purpose" Blue Box solution should also be done taking into account these aspects.

Evaluation results
Based on a balanced evaluation of technical merits and other aspects, the partners preferred to implement both a Commercial Off The Shelf (COTS) and an Open Source (OS) based solution. It should be noted that the decision was mainly based on the requirements of the Pilot Phase for Helix Nebula in 2013 and at this point in time does not constitute any decision by the supply-side towards a future deployment.

4. Implementation Plan

The intention of installing the "Blue Box" functionality is to be able to demonstrate that the partners can build and supply a multi-supplier federated cloud environment between the current and potential future suppliers. This should ultimately have a fully-functional and coherent interface, allowing any of the users to invoke and exploit services across any of the suppliers. It is expected that this interface will be assembled in stages, starting with the three aspects (common API, VM image library management, federated identity management) which are felt most urgent by the demand side; others (such as reporting and billing) will be available with later versions of the solution. In the next sections first the general aspects of the Blue Box implementation are discussed followed by specific sections for the COTS and OS solutions.

4.1. VM image library management and conversion

VM image library management is fully integrated into the abstraction layer. It is based on APIs that specify cloud-independent interfaces for enabling an application to access resources across multiple clouds. This way the Blue Box solution can be used as a management solution for a hybrid cloud model. Automation functions include auto-provisioning, auto-scaling, automated recovery, automated backups, and smart cloud topology. And for setting-up the solution limited conversion of VMs will be needed and is usually covered by specific freeware and/or third party products and consulting support.

Following the DevOps philosophy a cloud environment should be operated in such a way that IT organizations can support today's business requirements for agility and speed without the struggle of initial or regular conversions. The traditional manual methods of system configuration and administration are slow and error-prone. The

Blue Box solution and other tools offer automated and consistent system configuration and administration – to deliver the promise of cloud computing.

In summary, three layers perform the needed vital roles: the Cloud Manager, the Configuration Manager and the Resource Manager layer. The layers can be integrated and coordinated to achieve agility throughout the application lifecycle.

The Cloud Manager Layer – represented by the Blue Box solution - coordinates the collection of resources to ensure functionality and performance of the entire application. It ensures that each VM is operating properly and that sufficient resources are available to meet application demand. It controls user access to application resources and assigns costs for those resources to appropriate user organizations.

The Configuration Manager Layer configures the resources to fulfil their role within an application. Configuration Manager products install and configure software packages to enable individual VMs to perform their role within the overall application.

The Resource Manager Layer provides the necessary resources for applications. These are fundamental resources like virtual machines with associated storage and network connections. Products such as vSphere, OpenStack, CloudStack, or Zimory typically provide Resource Manager functionality, and are commonly referred to as cloud software.

4.2. Federated identity management

The Blue Box solution provides functions for federated authentication and authorization. However, currently very few cloud providers have support for third party authentication and almost none support third party authorization and tend do so in very different ways.

The Blue Box solution enables customers to significantly improve Identity and Access Management strategy with their cloud deployments. The solution uses the existing tools from cloud providers and expands that coverage with fine-grained, role-based access control that is cloud independent. The solution supports a variety of authentication methods to meet unique requirements of enterprises. Customers also gain auditing and logging of all user actions—something that cloud providers don't make available. And, existing directories, such as Active Directory and LDAP, can be leveraged to minimize the complexity of deployments and maintain fewer points of user management.

The Blue Box solution Identity Management currently supports multiple leading public clouds and private cloud platforms. Across these clouds, the solution enables enterprises to leverage leading configuration management solutions.

Security Architecture

The Blue Box solution is built on the foundation of separation of roles in an IT infrastructure. Through the combination of role separation with the wide use of encryption and proper key management, a cloud-based infrastructure is constructed that will tolerate failures at multiple levels without damaging the overall integrity of data or ability to recover from disaster. Separation of roles starts at the organizational level:

- The user controls the operation of his applications and databases
- The Blue Box solution manages the provisioning of systems, key management, and user management
- the cloud provider manages the physical resources on which everything operates

All user interaction with the Blue Box occurs through the console zone. Access to the console is currently controlled via a user name/password system. The Blue Box enables customers to define the security profile they want for authentication, including:

- SAML Federation
- Multi-factor authentication
- Open ID with trusted providers

If a user has access to multiple accounts, the Blue Box will always force them to authenticate with the strongest authentication system configured for the accounts to which they have access.

The Blue Box provisioning systems perform the following functions:

- Monitoring of virtual servers in customer cloud infrastructures for various events
- Orchestrating the launch and configuration of systems operating in the cloud
- Processing requests by servers operating in the cloud
- Alerting appropriate individuals about key events happening in the cloud (for example, server failures)
- Tracking the encryption systems used for various resources in the cloud

4.3. Blue Box based on COTS solution

The COTS-Software will be installed dedicated for and branded as Helix Nebula on servers provided by CloudSigma in its cloud data centre. Service Support will be provided by a team consisting of Logica, The Server Labs and T-Systems. The service support will include setup, training, a help desk and 24x7 tracking of incidents. 2nd and 3rd level support is guaranteed by the vendor.

Interfaces offered to the users will be REST API, web app and EC2 interface.
A key benefit of the COTS solution is that it comes with an integrated extensive Financial Management function - that is independent of how the service is provided e.g. based on on-premise, off-premise, traditional or public component. The financial management provides full transparency and control of costs. Details will be shared with the demand-side during the Pilot Phase.

4.4. Blue Box based on Open Source solution

The Open Source (OS) solution is an automation engine able to provision cloud resources for a number of cloud providers. The OS solution is available to demonstrate Blue Box capabilities and is branded in the Helix Nebula colours and logo. Interfaces offered to the users will be REST API, command-line client, web app, preliminary EC2 interface (includes foundations for tags filtering). Atos will provide hosting of the service and first-line support, while SixSq will provide 2nd and 3rd line support.

4.5. Blue Box as part of the Helix Nebula Service Enabling Framework

For reference an overview of the Blue Box in the Enabling Framework is given below.

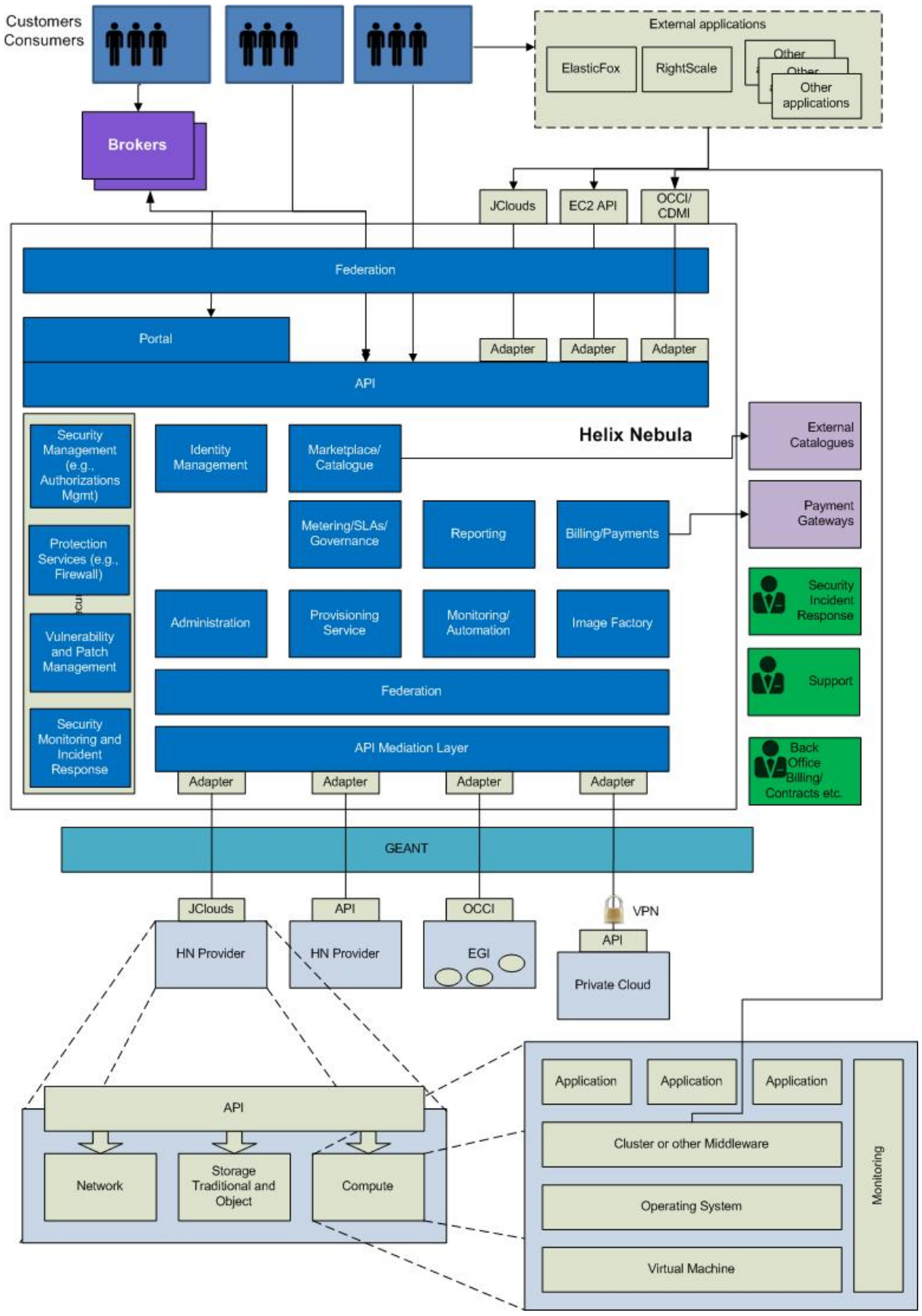

Figure 1: Helix Nebula Service Enabling Framework - overview

5. Supply-side commitments and resources

5.1. Software Development

Partners SixSq and The Server Labs are committed in providing 2nd and 3rd line support to the respective Blue Box solutions, including the different backend cloud connectors and the front-end EC2 interface. This support will include software upgrades, maintenance and bug fixes.

5.2. IaaS-Providers

5.2.1. Atos

Atos is committed to providing resources for Helix Nebula from its European Cloud environment, the relevant part of which is based in Eindhoven, The Netherlands. It is expected to be connected via a peering arrangement with GEANT in the Amsterdam IX PoP.

An initial implementation, in the order of 2,000 virtual images and a suitable accompanying volume of Storage on Demand, is in preparation, predicated on the Helix Nebula Demand-side partners confirming their ability to procure the services, if suitable.

The environment will be based on Atos' standard hardware environment, with suitable operating software for Helix Nebula installed on top, as a further-enhanced version of what was done for the Proof-of-Concept runs. This consists of various Open Source tools.

5.2.2. CloudSigma

CloudSigma will provide virtual cloud resources both in terms of compute and storage from its data centre based in Zurich, Switzerland. The CloudSigma cloud is driven from an online browser based web console and a provisioning API available via a public URL endpoint.

Network connectivity will be available over general IP connectivity and also via the Géant Network through Switch in Switzerland. The latter is expected to offer lower latency to demand-side institutions who also utilise the Géant network.

5.2.3. Interoute

Interoute will provide IaaS resources using its Virtual Data Centre Service. Compute, storage and network resources can be deployed from the London, Amsterdam, Geneva, Berlin or Paris VDC zones. The service can be used in a hosted private or public cloud

solution; accessing from the Internet or via an Interoute MPLS VPN. It can be managed online via web based GUI or using the API. The most feature rich available VDC RESTful API is the jclouds API.

5.2.4. T-Systems

T-Systems will provide virtual resources based on its DSI and VMWare vCloud IaaS-services from a data centre in Munich and/or other locations. The data centre will be connected to the Géant Network.

5.3. Service Support

Central Service Management and Support for the COTS solution will be run by a team of Logica and The Server Labs. The OS solution will be supported by Atos and SixSq for 1st, 2nd and 3rd line support.

5.4. Pricing plans and typical costs of using dynamic resources during a PoC

Pricing plans will be provided to the demand-side at the beginning of the Pilot Phase individually by each provider to maintain full competition between the IaaS-Providers and full transparency for the demand-side.

An indication of the typical costs of using dynamic resources during the Pilot Phase will vary from flagship to flagship due to the quite different type of resources required and time-span needed to run the benchmark applications. However, a close estimate based on the Proof-of-Concept Phase, the costs for use of dynamic resources would aggregate to be in a range of 10k-20k Euros for a 2-3 weeks test.

Such typical costs might also serve as good indicator for new members with interest to join the Helix Nebula Initiative and that plan to run a similar flagship.

6. Conclusion

At this stage one could assess that we have collected extensive information, achieved a much better understanding of the current state of cloud computing in eScience and federated cloud services, but also learned there is much work to be done to find the best solutions to shortcomings identified and progress beyond the state-of-art.

The Proof-of-Concept Phase clearly identified the demand- and supply-side priorities and these have led to define the Pilot Phase requirements. A solution has been agreed amongst the partners how best to run the Pilot Phase, focussing both on an assessment and evaluation of a COTS and an Open Source solution for federated cloud management. The evaluation will not only take technical aspects and features into account but, as has been rightly pointed out by the work of the ServArch Working group, will bring additional focus on the service aspects of the future Helix Nebula Science Cloud. The eventual goal is to develop an open standards Blue Box solution suitable to eScience and other applications and that can be used in a variety of environments. Ideally it would be build on Open Source but may contain various degrees of COTS elements if needed.

A key benefit of a future service will be the central platform backed by a professional service management and support by a single-point-of-contact. Current assessment from industry perspective is that an eco-system for Science Cloud Computing can be made sustainable at competitive prices when it would manage a federated cloud infrastructure that would ramp-up to a volume of ca. 10.000 concurrent virtual machines images by 2014.