



Enhancing Grid Infrastructures with
Virtualization and Cloud Technologies

Quarterly Report

Quarterly Report QR1 (V1.2)
16 September 2010

Abstract

In the first quarter of the project, the participants have successfully laid the foundations for creating a stable, comprehensive, open-source cloud platform. The project has put in place the tools and processes to quickly build and expand the StratusLab distribution. Through dialog with targeted communities and through the partners' experience it has developed the initial architecture of the StratusLab distribution. The initial releases have been validated on the project's cloud infrastructure and the project has put in place the mechanisms for disseminating release information once StratusLab makes its first public release.



StratusLab is co-funded by the
European Community's Seventh
Framework Programme (Capacities)
Grant Agreement INFSO-RI-261552.



The information contained in this document represents the views of the copyright holders as of the date such views are published.

THE INFORMATION CONTAINED IN THIS DOCUMENT IS PROVIDED BY THE COPYRIGHT HOLDERS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE MEMBERS OF THE STRATUSLAB COLLABORATION, INCLUDING THE COPYRIGHT HOLDERS, OR THE EUROPEAN COMMISSION BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THE INFORMATION CONTAINED IN THIS DOCUMENT, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Copyright © 2010, Members of the StratusLab collaboration: Centre National de la Recherche Scientifique, Universidad Complutense de Madrid, Greek Research and Technology Network S.A., SixSq Sàrl, Telefónica Investigación y Desarrollo SA, and The Provost Fellows and Scholars of the College of the Holy and Undivided Trinity of Queen Elizabeth Near Dublin.

This work is licensed under a Creative Commons Attribution 3.0 Unported License
<http://creativecommons.org/licenses/by/3.0/>



Contributors

Name	Partner	Sections
Charles Loomis	CNRS-LAL	All
David O’Callaghan	TCD	WP3
Marc-Eliau Bégin	SIXSQ	WP4
Vangelis Floros	GRNET	WP5
Juan Caceres	TID	WP6

Document History

Version	Date	Comment
0.1	2 Sept. 2010	Initial outline of report.
1.0	11 Sept. 2010	First version for internal review.
1.1	13 Sept. 2010	Changes based on comments from reviewers.
1.2	13 Sept. 2010	Final version.

Contents

List of Tables	6
1 Publishable Summary	7
1.1 Project Context and Objectives	7
1.2 Summary of Work Performed and Achievements	8
1.3 Final Results and Potential Impact and Use	9
1.3.1 Improved Interdisciplinary Scientific Collaboration	9
1.3.2 Impact on DCI Evolution.	10
1.3.3 Improved Usability of DCI Platforms	10
1.4 Contact Information	10
2 Project Objectives for the Period	12
2.1 Objectives	12
2.2 Review Recommendations	12
3 Progress and Achievements	13
3.1 WP2: Interaction with Targeted Communities	15
3.1.1 Progress Towards Objectives by Task	15
3.1.2 Issues and Corrective Actions	16
3.1.3 Use of Resources.	16
3.2 WP3: Dissemination.	17
3.2.1 Progress Towards Objectives by Task	17
3.2.2 Issues and Corrective Actions	17
3.2.3 Use of Resources.	18

3.3	WP4: Software Integration and Distribution	19
3.3.1	Progress Towards Objectives by Task	19
3.3.2	Issues and Corrective Actions	20
3.3.3	Use of Resources	20
3.4	WP5: Infrastructure Operation	21
3.4.1	Progress Towards Objectives by Task	21
3.4.2	Issues and Corrective Actions	23
3.4.3	Use of Resources	23
3.5	WP6: Innovative Cloud-like Management of Grid Services and Resources	24
3.5.1	Progress Towards Objectives by Task	24
3.5.2	Issues and Corrective Actions	24
3.5.3	Use of Resources	24
4	Project Management	25
4.1	Consortium	25
4.2	Management Tasks	25
4.3	Issues	26
4.4	Planning	26
5	Deliverables and Milestones	29

List of Tables

1.1	StratusLab Partners	11
3.1	StratusLab Physical Infrastructure	22
4.1	Meetings	27
4.2	Metrics	28
5.1	Deliverables (Year 1).	30
5.2	Deliverables (Year 2).	31
5.3	Milestones	32

1 Publishable Summary

1.1 Project Context and Objectives

The StratusLab project is aimed at service provisioning, networking, and research of technologies that will bridge cloud and grid infrastructures to simplify and optimize the use and operation of existing distributed computing infrastructures (e.g. European Grid Infrastructure) and to provide a more flexible, dynamic computing environment for scientists.

The European production grid infrastructure has had many notable successes. It has allowed scientists from all over Europe and indeed from all over the world federate their computing resources to advance their scientific aims. More importantly, the infrastructure allows them to federate their data and expertise to accomplish more than they would be able to do singlehandedly. Common APIs and service interfaces make it possible to take advantage of these distributed resources without having to modify applications for each site.

Despite its success, the grid also has its limitations. The uniformity of service interfaces unfortunately does not extend to the underlying computing resources, where users are exposed to significant heterogeneities in the computing environment, complicating applications and increasing failure rates. Passive calculations are handled well by the grid, but many applications require active services to coordinate the distributed analyses. Either scientists must provide their own resources for such services or negotiate with a particular site to provide them. This reduces the speed at which new calculations can be done.

Virtualization technologies provide a mechanism for offering customized, uniform environments for users with negligible performance degradation. Using grid technologies combined with virtualization allows the grid to provide users with a homogeneous computing environment, simplifying applications and reducing failures. Emerging cloud technologies allow users to dynamically allocate computing resources (often in less than a minute) and to specify the characteristics for the allocated resources. The fusion of cloud and grid technologies provides a more dynamic and flexible computing environment for grid application developers.

Cloud and virtualization technologies also offer other benefits to administrators of resource centers, such as the migration of live services for load balancing or the deployment of redundant servers. Reduced costs for managing resources immediately benefit users by freeing money for additional computing resources or

by having better user support from administrators.

A combined computing infrastructure that uses grid technology's strengths for federating resources, virtualization's strengths in providing custom, uniform environments, and the cloud's strengths in dynamic resource allocation, maximizes the utility of European distributed computing resources to scientists.

The StratusLab project will create a complete, coherent, open-source private cloud distribution to allow administrators of grid resources centers to take advantage of virtualization and cloud technologies. It will also provide new ways of using existing distributed computing resources to make the infrastructure more adaptable and more useful for scientists.

1.2 Summary of Work Performed and Achievements

In the first quarter of the project, the participants have successfully laid the foundations for creating a stable, comprehensive, open-source cloud platform. The project management, in cooperation with all of the activities, has put in place the collaborative tools necessary for a successful software development project, including mailing lists, issue trackers, code repositories, and the like. Using these tools, the project members have made significant progress in collecting use cases, defining the StratusLab architecture, creating a base release, and testing it on the project's cloud infrastructure.

Project Visibility A project web site was setup to increase the visibility of the project within our targeted communities and the general public. The web site includes general information about the project as well as detailed information about the project's technical work. An initial press release and numerous technical presentations helped increase the visibility of the project. All members of the project have contributed material to ensure a visible presence at the upcoming EGI Technical Forum.

Dialog with Targeted Communities Two surveys, one for end-users and another for system administrators, were developed and used to collect feedback from those likely to deploy and use the StratusLab software. The document "Review of the Use of Cloud and Virtualization Technologies in Grid Infrastructures" (D2.1) provides a set of 25 requirements and recommendations based on the survey responses. These guided the development of the StratusLab architecture.

Agile Software Processes The project has adopted scrum, an agile software development process, for managing the production of the StratusLab software releases. This is an iterative method that begins with a sprint planning meeting, three weeks of work on the defined tasks (with daily 15-minute, "standup" meetings), and finishes with a sprint demonstration meeting and an updated StratusLab distribution. The technical coordination group reviews progress between sprints. The first few sprints have already produced a StratusLab distribution that can be easily installed manually or automatically.

Initial Architecture An initial architecture that takes into account the feedback from the targeted communities and the partners' experience with virtualization and cloud technologies has been defined. Various aspects of this initial architecture and choices for implementations (e.g. OpenNebula as a virtual machine manager, Apache for an appliance repository, etc.) have already been verified with early developments of the project.

Initial Project Infrastructure The early StratusLab distributions have been deployed on the project's physical infrastructure provided by GRNET and CNRS. Moreover, the distributions have been extensively tested using this infrastructure to reveal a significant number of bugs and issues that are being corrected before a public release. Moreover, grid services have been deployed over the StratusLab to verify the applicability of cloud resources to these services.

This foundation provides a solid basis for the first public release of the StratusLab software expected Q2 with additional features appearing rapidly afterwards.

1.3 Final Results and Potential Impact and Use

Most scientific and engineering research requires significant computing resources. Distributed computing infrastructures have brought unprecedented computational power to a wide range of scientific domains. Although, these architectures and the related software tools have been considerably improved over the years, they exhibit several difficulties, mainly due to limitations of physical platforms, which discourage adoption of grid technologies. StratusLab has the potential to profoundly change existing grid infrastructures.

1.3.1 Improved Interdisciplinary Scientific Collaboration

Cloud technologies are expected to have significant impact, both immediate and long-term, in the way scientific research is carried out. Grid infrastructures have provided a remarkable advantage over the past year offering access to vast amount of computing power and storage space, and most importantly by offering a sustainable platform for scientific collaboration enabling the sharing of computing resources and scientific data. Cloud computing is expected to take this one step further by facilitating the easy deployment of customized grid infrastructures. These infrastructures are expected to have further positive impact on the way interdisciplinary scientific research is taking place.

StatusLab focuses on the provision of scientific infrastructures over cloud computing, investigating in particular the provision of customized Virtual Machine images. This customization will be done on the user side, which means that the user can have more immediate influence on the infrastructure itself. In this way the infrastructure will adapt to the user requirements and not vice-versa. By easing the management of grid sites and the configuration of hosting services we expect to attract a broader number of scientific communities and further facilitate their collaboration.

1.3.2 Impact on DCI Evolution

Currently, there is a big shift in all e-Infrastructure projects, and related efforts in Europe, to expand their activities in order to include cloud computing technologies. StratusLab will play a key role in this landscape by providing a focused environment for development, deployment and experimentation of cloud computing services.

The projects proposal reflects an evolutionary path from the existing large-scale monolithic grid e-Infrastructures to novel, beyond the state-of-the-art, cloud-based, grid-enabled ones. Through its expected collaborations with other projects, StratusLab will disseminate its findings and drive direct impact on the way e-Infrastructures provision is currently done.

1.3.3 Improved Usability of DCI Platforms

Virtualization is the cornerstone of cloud computing and a key for achieving optimal usability of DCI platforms. Moreover, virtualized environments have the ability to adapt to different hardware platforms enabling a quick transition from one environment to another.

StratusLab operates such a virtualized platform on a variety of hardware environments. By offering customized machine images, users will be able to set-up an environment that better suits their application requirements. This will dramatically improve the current situation where current infrastructures are forced to offer a common configuration—a common denominator—that tries to do its best to satisfy many users with different runtime requirements. Another aspects where StratusLab will contribute is on power consumption efficiency (Green Computing) and the increase reliability by incorporating failover mechanisms using virtual machine snapshots and migration.

1.4 Contact Information

More information about the StratusLab project can be obtained from the project's website: <http://stratuslab.eu/>. Individual partners can also be contacted to obtain more specific information about their contributions to the project. Table 1.1 contains the list of StratusLab partners and relevant contacts.

Table 1.1: StratusLab Partners

CNRS	Centre Nationale de la Recherche Scientifique	Charles LOOMIS loomis@lal.in2p3.fr
UCM	Universidad Complutense de Madrid	Ignacio LLORENTE llorente@dacya.ucm.es
SIXSQ	SixSq Sàrl	Marc-Elian BEGIN meb@sixsq.com
GRNET	Greek Research and Technology Network S.A.	Evangelos FLOROS efloros@gnet.gr
TID	Telefónica Investigación y Desarrollo SA	Juan CACERES caceres@tid.es
TCD	The Provost Fellows and Scholars of the College of the Holy and Undivided Trinity of Queen Elizabeth Near Dublin	Brian COGHLAN coghl@cs.tcd.ie

2 Project Objectives for the Period

2.1 Objectives

In this first quarter, the primary objective was to prepare the foundations for a successful project. In more detail this involves:

- Deployment of collaborative software development tools,
- Starting dialog between StratusLab and targeted communities,
- Make the project visible to targeted communities and general public,
- Put in place the software development processes and policies,
- Define the initial architecture for the StratusLab software, and
- Deploy the initial project infrastructure.

Within this quarter all of these have been obtained providing a solid basis for the first public release of the StratusLab software in Q2 with additional features appearing rapidly afterwards.

2.2 Review Recommendations

Not yet applicable.

3 Progress and Achievements

In the first quarter of the project, the participants have successfully laid the foundations for creating a stable, comprehensive, open-source cloud platform. The project management, in cooperation with all of the activities, has put in place the collaborative tools necessary for a successful software development project, including mailing lists, issue trackers, code repositories, and the like. Using these tools, the project members have made significant progress in collecting use cases, defining the StratusLab architecture, creating a base release, and testing it on the project's cloud infrastructure.

Project Visibility A project web site was setup to increase the visibility of the project within our targeted communities and the general public. The web site includes general information about the project as well as detailed information about the project's technical work. An initial press release and numerous technical presentations helped increase the visibility of the project. All members of the project have contributed material to ensure a visible presence at the upcoming EGI Technical Forum.

Dialog with Targeted Communities Two surveys, one for end-users and another for system administrators, were developed and used to collect feedback from those likely to deploy and use the StratusLab software. The document "Review of the Use of Cloud and Virtualization Technologies in Grid Infrastructures" (D2.1) provides a set of 25 requirements and recommendations based on the survey responses. These guided the development of the StratusLab architecture.

Agile Software Processes The project has adopted scrum, an agile software development process, for managing the production of the StratusLab software releases. This is an iterative method that begins with a sprint planning meeting, three weeks of work on the defined tasks (with daily 15-minute, "standup" meetings), and finishes with a sprint demonstration meeting and an updated StratusLab distribution. The technical coordination group reviews progress between sprints. The first few sprints have already produced a StratusLab distribution that can be easily installed manually or automatically.

Initial Architecture An initial architecture that takes into account the feedback from the targeted communities and the partners' experience with virtualization and cloud technologies has been defined. Various aspects of this initial architecture and choices for implementations (e.g. OpenNebula as a virtual machine manager,

Apache for an appliance repository, etc.) have already been verified with early developments of the project.

Initial Project Infrastructure The early StratusLab distributions have been deployed on the project's physical infrastructure provided by GRNET and CNRS. Moreover, the distributions have been extensively tested using this infrastructure to reveal a significant number of bugs and issues that are being corrected before a public release. Moreover, grid services have been deployed over the StratusLab to verify the applicability of cloud resources to these services.

This foundation provides a solid basis for the first public release of the StratusLab software expected Q2 with additional features appearing rapidly afterwards.

3.1 WP2: Interaction with Targeted Communities

The major achievements of this activity include the analysis of two surveys and the development of a set of application benchmarks. The surveys targeted end-users of the StratusLab distribution and system administrators who will install it. They provided extensive feedback and requirements that will be used to guide the development of the StratusLab distribution. The application benchmarks cover a wide range of real-world analysis patterns and will be used to validate the StratusLab distribution and to provide concrete performance metrics.

3.1.1 Progress Towards Objectives by Task

3.1.1.1 Task 2.1: Interactions with Resource Providers and End-users

User and Administrator Surveys A major component of the work plan for the previous quarter was the creation, deployment, and analysis of two surveys: one for end-users and another for system administrators. IBCP has used its expertise and connections with the bioinformatics community to solicit responses to the User and Administrators surveys. Announcements of these surveys were sent to the scientific and technological lists and contacts in France and also in Europe. The results of these surveys are presented in the deliverable “Review of the Use of Cloud and Virtualization Technologies in Grid Infrastructures” (D2.1). They have provided valuable feedback for the development activities of the project and have allowed the project to build the initial lists of end-users and system administrators interested in announcement of StratusLab releases.

Contacts with ATLAS Scientists Following a presentation at the Seillac conference for LAL employees, two people from the ATLAS experiment approached us concerning use of virtualization and cloud technologies. With our help, they developed a virtual machine that contained the commercial software required for their research and successfully ran instances of that machine on LAL’s OpenNebula installation. The software uses a token-based licensing system that worked without problems with the multiple virtual machine instances.

Contacts with French Bioinformatics Community Members of IBCP have participated in meetings in collaboration with the French RENABI GRISBI community. This is designed to foster acceptance and deployment of the StratusLab Toolkit within this community and to gather their requirements.

Bioinformatics Appliance IBCP has defined what will be the first bioinformatics virtual appliance of the project. This virtual machine will be based on the grid worker node distribution, will have selected bioinformatics applications pre-installed, and will be connected to the biological databases repository of the bioinformatics grid site. This thematic appliance will be available from the project’s appliance repository.

3.1.1.2 Task 2.2: Intensive Evaluation of StratusLab Products

Application Benchmarks The core development of this work package has been the creation of a set of application-level benchmarks. These benchmarks cover CPU-intensive, IO-intensive, workflow, and parallel applications. These will be used to validate StratusLab releases and to measure the performance of the system for real scientific applications. These benchmarks are packaged and made available as part of the StratusLab releases.

Debugging During the installation and use of the OpenNebula deployments, numerous bugs and features requests were generated. All of these issues have been put into the project's issue tracker (JIRA) and will be followed up in future sprints. Issues related to the use of the StratusLab utilities for deploying machines, creating machines, etc. have similarly been put into JIRA and reported to the developers.

3.1.2 Issues and Corrective Actions

No major issues related to WP2 have arisen in Q1.

3.1.3 Use of Resources

The resource utilization is consistent with the plans foreseen in the Technical Annex.

3.2 WP3: Dissemination

The focus in the first quarter has been on promoting the project and its objectives to a variety of audiences. This was achieved through a launch press release, the project website, visible online presence and technical presentations. The EGI Technical Forum 2010 has been chosen as a high-profile event to promote the project to key target groups.

Looking forward to the rest of the first year, the project is developing its initial plans for dissemination, collaboration and standardization activities, and it has established initial aims for collaboration with related projects.

3.2.1 Progress Towards Objectives by Task

3.2.1.1 Task 3.1: Dissemination

Press Release The dissemination activities began with a press release announcing the launch of the project, which was picked up by relevant online media reaching resource providers, end-users, and the technical and general public.

Website The project website was set up (<http://www.stratuslab.eu>) and some plans have been made for future developments. A Twitter account (*StratusLab*) has been created to allow the project to develop a social network.

Presentations Project members have given presentations that describe or mention StratusLab, such as at the XtremOS Summer School, Günzberg, Germany and at the CSIC graduate course on Grid and e-Science, Valencia, Spain both in July 2010. In addition, keynotes or invited talks with slides about StratusLab were given at EuroPar 2010, Ischia; CERN, Geneva; HPC 2010, Cetraro; SWSTE 2010, Herzlia; and First European Summit on Future Internet, Luxembourg.

EGI Technical Forum 2010 The project has booked an exhibition booth at EGI Technical Forum 2010 and is preparing posters, demonstrations and other dissemination material for this event.

3.2.1.2 Task 3.2: Collaboration with Standards Bodies and Related Projects

The project is developing its “Initial Plan for Dissemination, Collaboration and Standardization Activities” (D3.1) to be published shortly. Project members have made contact with related projects (in particular the EC-funded Distributed Computing Infrastructures projects) to establish points of collaboration, via face-to-face and remote meetings. The project has joined with other DCI projects to exchange information on collaboration needs and expectations. Two project members participated in a joint DCI kick-off event in Brussels.

3.2.1.3 Task 3.3: Development of Exploitation and sustainability Plan

This task will begin in Q4.

3.2.2 Issues and Corrective Actions

No major issues related to WP3 have arisen in Q1.

3.2.3 Use of Resources

The resource utilization is consistent with the plans foreseen in the Technical Annex.

3.3 WP4: Software Integration and Distribution

During Q1, WP4 focused on the definition of the StratusLab architecture and produced the deliverable “Reference Architecture for StratusLab Toolkit 1.0” (D4.1). This major effort included an analysis of the current needs of our targeted user communities, based on the information contained in the WP2 surveys and reported in “Review of the Use of Cloud and Virtualization Technologies in Grid Infrastructures” (D2.1).

Originally Tasks 4.2-4 were scheduled to start at M4, however, they were started earlier in order to validate key concepts in the architecture and to support the work of other activities. In close cooperation with the other work packages, WP4 has put into place Scrum, an agile software development process. Within Q1 the first sprints (iterations) took place, providing the initial definition of the StratusLab distribution, definition of the contextualization strategy, creation of the initial software packages, and tools for manual and automated installation.

UCM, as part of the WP4 support activities, organized an OpenNebula Tutorial, 20-21 July 2010, and hosted a StratusLab/OpenNebula workshop on the 22 July 2010 in Madrid. These events were important for firming-up the definition of the StratusLab architecture. More generally, WP4 has provided support to the entire project concerning the StratusLab tools and OpenNebula.

3.3.1 Progress Towards Objectives by Task

3.3.1.1 Task 4.1: Definition of Reference Architecture

The D4.1 document presents the initial architecture of the StratusLab distribution, defining the foundation of StratusLab 1.0. The document is a starting point and the architecture will evolve over the course of the project, as we learn and gather feedback from users and system administrators. The architecture will be formally updated at M15 with the deliverable D4.4. While defining the architecture, sprints have been conducted and user stories selected to validate the main ideas behind StratusLab 1.0.

3.3.1.2 Task 4.2: Integration of Open-source Distribution

Early in the project and in order to support the architectural definition task (Section 3.3.1.1), it was important to validate several assumptions made during the proposal writing stage. To support this work, sprints have been conducted and user stories selected in order to validate the main ideas behind StratusLab 1.0: OpenNebula as the VM Manager, remote Appliance Repository, remote access to the cloud (via the OpenNebula XMLRPC API) and mechanisms (manual and automated) for distribution installation and configuration of the StratusLab software. This validation was successful with the production of an early distribution, which can be easily installed.

This work revealed several technical issues, which have been tracked using the project tracking tool (JIRA/GreenHopper). As the project identifies bugs and issues

with the early versions of StratusLab, including OpenNebula, both SixSq and UCM have been active at fixing these bugs and addressing the identified issues.

3.3.1.3 Task 4.3: Contextualization of Grid Services

Before grid services can be contextualized, it is important to understand well contextualization itself. In this quarter, we have defined a generic virtual machine contextualization strategy, documented on the project wiki and reviewed by all partners. This mechanism will be included in all base images managed and maintained by the project. This work included the definition of a manifest file describing the images in the appliance repository. From this foundation, specific support for grid service contextualization will be realized as this task progresses.

3.3.1.4 Task 4.4: Technical Support

WP4 has provided support for the StratusLab tools and OpenNebula to the whole project. This support has been provided via the daily standup meetings, phone calls, Skype, and email.

WP4 has defined a software development procedure based on scrum, an agile software development process. WP4 manages this process by running the daily standup meetings, the sprint demo meetings and the sprint preparation meetings. During Q1, the initial sprints (iterations) took place, putting in motion the process and producing real results. To support this process, WP4 with WP5 has put in place the tools to support continuous integration, like for example, the project's Hudson server (<http://hudson.stratuslab.eu:8080/>).

As part of its technical support, WP4 has created and maintained several wiki pages providing information on the StratusLab distribution, including installation instructions, conventions and technical notes.

3.3.2 Issues and Corrective Actions

No major issues related to WP4 have arisen in Q1.

3.3.3 Use of Resources

The resource utilization is consistent with the plans foreseen in the Technical Annex.

3.4 WP5: Infrastructure Operation

WP5 is responsible for the provision of the computing infrastructure required by the various activities of the project. Since StratusLab is mainly an infrastructure project the smooth execution of WP5 is essential for the unobstructed implementation of the whole project. During the first quarter of the project, work towards many of the initial goals of the work package progressed at a very satisfactory level.

The first priority was the establishment of the physical infrastructure necessary for hosting all the cloud and grid services of the project. Once this was completed, the next step was to start with a preliminary installation and testing of cloud middleware (OpenNebula). Further was the installation of grid services and the first trials of the provision of these services using cloud computing capabilities.

In parallel, WP5 contributed with its infrastructure to the other activities (in particular WP4) by providing the necessary physical resources for hosting various support services (e.g. the Hudson continuous integration platform) and for running various trials and testing activities. At the end of the first quarter all the necessary infrastructure is in place: a number of cloud services are already available, which host pre-production grid sites.

3.4.1 Progress Towards Objectives by Task

3.4.1.1 Task 5.1: Deployment and Operation of Virtualized Grid Sites

Commissioning of Physical Infrastructure The first priority was commissioning the physical computing infrastructure that will be required throughout the lifetime of the project, mainly for hosting cloud services on those deploying a number of grid sites based on EMI's UMD grid middleware distribution. Two project partners have committed to provide the necessary resources—GRNET and CNRS. The resources provided by each partner respectively are summarized in Table 3.1.

Installation of OpenNebula-based Cloud Sites Two cloud sites were installed in GRNET using OpenNebula (v1.4). The first site was installed on top of Ubuntu 10.04 and the second on top of CentOS 5. (Those are the initial platforms supported by the StratusLab collaboration.) Having a deployed instance of OpenNebula on each supported platform is essential for the project's testing activities. The site deployed on CentOS comprises 11 nodes (1 master and 10 hosts) and the site on Ubuntu, 12 nodes (1 master and 11 hosts). The physical nodes of these sites are monitored through Ganglia (<http://ganglia.sourceforge.net/> tool. The ganglia web monitoring front-ends for each cloud installation can be found in the following links:

CentOS site: <http://62.217.120.147/ganglia/>

Ubuntu site: <http://62.217.120.133/gang/>

Deployment of Grid Sites Two minimal grid sites, using the gLite 3.2 middleware, have been deployed on top of the cloud in GRNET's infrastructure. The first

Table 3.1: StratusLab Physical Infrastructure

Resource	GRNET	CNRS	Total
Computing nodes	26	5	31
CPU Cores	208	40	248
Storage	20 TB	5-10 TB	25-30 TB

site, named GR-42-STRATUSLAB, comprises a CREAM Computing Element (CE), Storage Element (SE), and 2 Worker Nodes (WNs). The second site, named GR-43-STRATUSLAB comprises of a CREAM CE, SE, and 5 WNs. Both sites are in pre-production phase and are monitored by the Service Availability Monitoring (SAM) tests running in NGI-GRNET (<https://sam.athena.hellasgrid.gr>). Both sites support MPI. Their small size permits reasonable feasibility tests. At the end of Sprint 1, we were able to demonstrate the stability and elasticity of the grid sites by dynamically adding and removing WNs on the fly.

Quattor Configuration of OpenNebula 2.0b LAL uses the Quattor Toolkit for the automated installation, configuration, and management of machines. To ensure that the cloud test bench at LAL can be similarly managed, LAL has developed Quattor configuration components for OpenNebula (ncm-oned) and for libvirt (ncm-libvirt). In addition, a complete set of configuration templates have been developed to allow an entire OpenNebula system (front-end and hosts) to be installed automatically from bare metal. Documentation (<http://stratuslab.eu/doku.php?id=quattorinstall>) on the StratusLab internal wiki explains how to use the configuration as well as the current limitations.

Quattor Deployment of Grid Worker Nodes The Quattor Toolkit bootstraps machine installation via PXE. LAL has verified that PXE images can be started via OpenNebula and that they can be used to install grid Worker Nodes via the standard Quattor installation mechanism. This is a first indication that automated site management tools can be used with little or no modification to control grid services running on a StratusLab cloud infrastructure.

Deployment of LAL Test Bench Using Quattor, LAL has deployed a StratusLab test bench consisting of five physical hosts. This consists of one OpenNebula front-end and four OpenNebula hosts. The front-end is also configured to allow virtual machines to run there. Currently, the test bench uses sqlite as the database implementation and the image repositories are shared via NFS.

3.4.1.2 Task 5.2: Testing of the StratusLab Toolkit

Testing of OpenNebula Deployment The LAL test bench has been used test and verify that the contextualization of images works correctly. This test bench was also used to verify the application benchmarks that have been developed within the WP2 activity. Moreover, several grid Worker Nodes were run on the LAL test bench are part of LAL's production grid infrastructure. No problems were found

related to the fact that they were running in a virtualized environment.

Performance Issues The use of NFS appears to cause significant startup delays in machine startup and non-responsive behavior with the OpenNebula daemon. These have been investigated by LAL and also by the OpenNebula developers. Optimized parameters will probably improve the performance; however, LAL has started looking into alternative configurations (using squid caching and LVM) that would remove the need for NFS and provide a more responsive and scalable architecture.

3.4.1.3 Task 5.3: Virtual Appliances Creation and Maintenance

Installation of Virtual Appliance Repository A first simple implementation of the virtual appliances repository is available from <http://appliances.stratuslab.eu>. In this first version the appliance repository is a standard Apache web server, hosted by Trinity College Dublin (TCD). The web server is accessed using Web-DAV (Web-based Distributed Authoring and Versioning), with authentication via the StatusLab LDAP server. The repository already offers a number of appliances, which have been created by other members of the project. For example, it contains VM images preconfigured with gLite software for the three basic components of a grid site (CE, SE and WN).

The appliance repository structure is designed to mirror that of a Maven repository. Information about an appliance is stored in an associated XML manifest file. The reason for following the above approach is that in the long term we plan to evaluate the use of tools such as Apache Maven for uploading and downloading of images. Initial tests using Maven have, however, proved unsuccessful with the large file sizes of the appliances currently found in the repository. This will be investigated further in the next quarter.

3.4.2 Issues and Corrective Actions

No major issues related to WP5 have arisen in Q1.

3.4.3 Use of Resources

No deviation of planned resource utilization has been experienced. The three task partners have allocated proper number of human and technical resources in order to properly progress with the planned program of work.

3.5 WP6: Innovative Cloud-like Management of Grid Services and Resources

3.5.1 Progress Towards Objectives by Task

This Work Package starts on month 4.

3.5.2 Issues and Corrective Actions

None.

3.5.3 Use of Resources

None.

4 Project Management

4.1 Consortium

The project consortium consisting of six partners (CNRS, UCM, GRNET, SIXSQ, TID, and TCD) has not changed since the start of the project. There have been no changes in the legal status of those partners.

4.2 Management Tasks

Despite administrative delays, the project participants started the project as planned on 1 June 2010. Their enthusiasm and previous collaborative experiences have allowed them to quickly deploy the needed tools, put software development processes in place, and make remarkable progress towards a first public release of software in M4. All of the partners have contributed to the project's web site (<http://stratuslab.eu/>), providing both material for dissemination and for the upcoming releases of software.

Table 4.1 contains a list of the meetings that have been planned to foster collaboration between the project participants. Not listed are the planning meetings for each development sprint and the daily standup meetings.

Management Tools The tools for managing the project and fostering collaboration between project participants have been deployed. The details are provided in the milestone “Establishment of Management Infrastructure and Metrics Definition” (MS1). This document also contains the definition of the metrics that will be used to track the progress of the project starting in Q2. These are also listed in Table 4.2.

Deliverables and Milestones The end of Q1 and beginning of Q2 has been a critical time for the project. Three important deliverables concerning the StratusLab requirements, architecture and infrastructure were due while preparations for the EGI Technical Forum and StratusLab release occurred in parallel. Despite the heavy workload, the project participants have produced the planned deliverables and milestones. Tables 5.1, 5.2, and 5.3 list all of the documents. In addition, these are available from the project website.

Conclusion of Legal Documents A consortium agreement was negotiated between all of the partners in preparation for the signature of the project's Grant

Agreement with the European Commission. This has been signed by all partners and is in force. A disagreement between CNRS and the European Commission concerning an interest-bearing bank account for the pre-financing delayed the signature of the Grant Agreement. This was resolved and the European Commission countersigned the Grant agreement on 25 August 2010. All of the signed Forms A and bank information have been received by CNRS for all of the partners except TCD. The signed TCD forms are expected shortly.

Intellectual Property Policies The Project Management Board defined the detailed Intellectual Property Policies for official document, presentations, and software. Documents and presentations will carry a common copyright with all of the partners and be released under the Creative Commons attribution license, allowing both commercial and non-commercial reuse with attribution to the project. Software copyrights will be held by the partner creating the code. The project has a strong preference for using the Apache 2 license, although other licenses can be used by partners if they do not interfere with creating a coherent StratusLab distribution of software. This will be validated by the TSCG.

4.3 Issues

Delayed Hiring For various reasons (e.g. delay in signing the Grant Agreement, administrative reorganizations, candidates finding other positions, etc.), the hiring process for StratusLab at several institutes has been delayed. Nonetheless, all of the partners are fulfilling their responsibilities with permanent or existing staff and this delay will not have a significant impact on the project schedule or results.

4.4 Planning

Two critical events will take place early in the next quarter: the EGI Technical Forum and the first public release of the StratusLab software. All of the project participants have been working hard to ensure that StratusLab has a strong, visible presence at the EGI Technical Forum and that the initial StratusLab software release will be useful to our target communities. Additionally, this period contains a deliverable to define our interactions with other projects will be important for maximizing the impact of the StratusLab results. In parallel, the WP6 tasks will start and the project will propose initial ideas for advanced management and use of cloud technologies in a distributed computing environment.

Table 4.1: Meetings

Title	Date	Venue	Comments
StratusLab Kick-Off Meeting	14-15/06/2010	Orsay, FR	Kick-off of project. Detailed planning for accomplishing objectives. http://indico.lal.in2p3.fr/conferenceDisplay.py?confid=1129
Technical Meeting	22/07/2010	Madrid, ES	Detailed technical discussions for StratusLab development. http://indico.lal.in2p3.fr/conferenceDisplay.py?confid=1189
Sprint 1 Demo	30/07/2010	Phone/EVO	Sprint 1 demonstration meeting. http://indico.lal.in2p3.fr/conferenceDisplay.py?confid=1191
Sprint 2 Demo	20/08/2010	Phone/EVO	Sprint 2 demonstration meeting. http://indico.lal.in2p3.fr/conferenceDisplay.py?confid=1192
Project Management Board	03/09/2010	Phone	PMB meeting to decide IPR policies. http://indico.lal.in2p3.fr/conferenceDisplay.py?confid=1203
Sprint 3 Demo	10/09/2010	Phone/EVO	Sprint 3 demonstration meeting. http://indico.lal.in2p3.fr/conferenceDisplay.py?confid=1203

Table 4.2: Metrics

Metric	Q2	Q3	Q4	Y1 Target	Q5	Q6	Q7	Q8	Y2 Target
No. of people on StratusLab announcement list				25					75
No. of people on StratusLab discussion list				50					100
No. of sites running StratusLab distribution				5					10
No. of sites exposing the cloud API				0					5
No. of VOs served via StratusLab sites				10					30
No. of sci. disciplines served via StratusLab sites				3					7
No. of available appliances				5					15
Availability of sites				80%					95%
Reliability of sites				80%					95%
No. of sprints				–					–
No. of releases				–					–
Delivered CPU				–					–
Delivered CPU through cloud API				–					–
Storage used				–					–
Storage used through cloud API				–					–
No. of appliance downloads				–					–
No. of views of website				–					–
No. of features (by state)				–					–
No. of bugs (by state)				–					–
No. of sites providing scale-out				–					–
Fraction of resources by scale-out of a site				–					–

5 Deliverables and Milestones

Tables 5.1, 5.2, and 5.3 list the deliverables for Y1, deliverables for Y2, and milestones, respectively. The deliverables and milestones for this reporting period were D2.1, D4.1, D5.1, MS1, and MS6.

Table 5.1: Deliverables (Year 1)

No.	Title	Version	WP No.	Lead Beneficiary	Nature	Diss. Level	Due Date	Actual Date	Status	Contractual	Comments
D2.1	Review of the Use of Cloud and Virtualization Technologies in Grid Infrastructures		WP2	CNRS	R	PU	PM2	11 August 2010	Done	Yes	
D4.1	Reference Architecture for StratusLab Toolkit 1.0		WP4	SIXSQ	R	PU	PM3	14/09/2010	Done	Yes	
D5.1	Infrastructure Specification		WP5	GRNET	R	PU	PM3	14/09/2010	Done	Yes	
D3.1	Initial Plan for Dissemination, Collaboration and Standardization Activities		WP3	TCD	R	PU	PM4				
D6.1	Cloud-like Management of Grid Sites 1.0 Design Report		WP6	TID	R	PU	PM5				
D5.2	Infrastructure Tool and Policy Specification		WP5	GRNET	R	PU	PM6				
D6.2	Cloud-like Management of Grid Sites 1.0 Software		WP6	TID	P	PU	PM11				
D2.2	Report on Evaluation of StratusLab Products		WP2	CNRS	R	PU	PM12				
D3.2	Report on Dissemination, Collaboration and Standardization Activities		WP3	TCD	R	PU	PM12				
D3.3	Exploitation and Sustainability First Plan		WP3	TCD	R	PU	PM12				
D4.2	StratusLab Toolkit 1.0		WP4	SIXSQ	P	PU	PM12				
D4.3	First Year Software Integration Report		WP4	SIXSQ	R	PU	PM12				
D5.3	First Year Infrastructure Operations Report		WP5	GRNET	R	PU	PM12				
D6.3	First Year Cloud-like Management of Grid Sites Research Report		WP6	TID	R	PU	PM12				

Table 5.2: Deliverables (Year 2)

No.	Title	Version	WP No.	Lead Beneficiary	Nature	Diss. Level	Due Date	Status	Contractual	Comments
D2.3	Survey of Targeted Communities Concerning StratusLab		WP2	CNRS	R	PU	PM14			
D4.4	Reference Architecture for StratusLab Toolkit 2.0		WP4	SIXSQ	R	PU	PM15			
D6.4	Cloud-like Management of Grid Sites 2.0 Design Report		WP6	TID	R	PU	PM17			
D5.4	Economic Analysis of Infrastructure Operations		WP5	GRNET	R	PU	PM18			
D6.5	Cloud-like Management of Grid Sites 2.0 Software		WP6	TID	P	PU	PM23			
D2.4	Final Report on StratusLab Adoption		WP2	CNRS	R	PU	PM24			
D2.5	Report on Evaluation of StratusLab Products		WP2	CNRS	R	PU	PM24			
D3.4	Final Review of Dissemination, Collaboration and Standardization Activities		WP3	TCD	R	PU	PM24			
D3.5	Exploitation and Sustainability Final Plan		WP3	TCD	R	PU	PM24			
D4.5	StratusLab Toolkit 2.0		WP4	SIXSQ	P	PU	PM24			
D4.6	Software Integration Final Report		WP4	SIXSQ	R	PU	PM24			
D5.5	Infrastructure Operations Final Report		WP5	GRNET	R	PU	PM24			
D6.6	Cloud-like Management of Grid Sites Research Final Report		WP6	TID	R	PU	PM24			

Table 5.3: Milestones

No.	Title	WP No.	Lead Beneficiary	Due Date	Achieved	Actual Date	Comments
MS1	Establishment of Management Infrastructure and Metrics Definition	WP1	CNRS	PM3	Yes	1/09/2010	
MS6	Website Operational	WP3	TCD	PM3	Yes	6/09/2010	
MS2	Contact Procedures and Supporting Tools for Targeted Communities	WP2	CNRS	PM4			
MS7	StratusLab Development, Certification and Release Procedures in Place	WP4	SIXSQ	PM6			
MS3	Creation of Virtual Appliances for Bioinformatics Community	WP2	CNRS	PM9			
MS10	Initial virtual appliance repository	WP5	GRNET	PM9			
MS14	Release of Cloud-like Management of Grid Services and Resources 1.0 Beta	WP6	TID	PM9			
MS8	Release of StratusLab 1.0 Beta	WP4	SIXSQ	PM10			
MS11	Operation of Site Running StratusLab toolkit v1.0	WP5	GRNET	PM10			
MS4	Adoption of StratusLab Software by External Grid Sites	WP2	CNRS	PM14			
MS12	Delivery of Virtual Appliance Repository	WP5	GRNET	PM18			
MS5	Opening of Virtual Appliances Repository to External Application Communities	WP2	CNRS	PM20			
MS15	Release of Cloud-like Management of Grid Services and Resources 2.0 Beta	WP6	TID	PM21			
MS9	Release of StratusLab 2.0 Beta	WP4	SIXSQ	PM22			
MS13	Operation of Site Running StratusLab Toolkit v2.0	WP5	GRNET	PM22			

Glossary

Appliance	Virtual machine containing preconfigured software or services
DCI	Distributed Computing Infrastructure
EGEE	Enabling Grids for E-sciencE
EGI	European Grid Infrastructure
EGI-TF	EGI Technical Forum
GPFS	General Parallel File System by IBM
Hybrid Cloud	Cloud infrastructure that federates resources between organizations
iSGTW	International Science Grid This Week
NFS	Network File System
NGI	National Grid Initiative
Public Cloud	Cloud infrastructure accessible to people outside of the provider's organization
Private Cloud	Cloud infrastructure accessible only to the provider's users
VM	Virtual Machine
VO	Virtual Organization
VOBOX	Grid element that permits VO-specific service to run at a resource center
Worker Node	Grid node on which jobs are executed