Visual Analysis for Extremely Large-Scale Scientific Computing

D2.3 – HPC cloud infrastructure specification document suitable to the needs of e-Science

Version 2.0

Deliverable Information

Grant Agreement no	619439
Web Site	http://www.velassco.eu/
Related WP & Task:	WP2, T2.3
Due date	December 31, 2014
Dissemination Level	
Nature	
Author/s	Ivan Martinez, Miguel A. Tinte
Contributors	Abel Coll, Miguel Pasenau, Benoit Lange, Heidi Dahl, Toàn Nguyên

Approvals

	Name	Institution	Date	OK
Author	Ivan Martinez	**ATOS**	11/11/2014	
Task Leader	Jochen Haenisch	**JOTNE**		
WP Leader	Toàn Nguyên	**INRIA**		
Coordinator				

Change Log

Version	Description of Change
Version 0.1	First draft version of the document
Version 0.5	Updated version with Benoit contribution regarding HPC and Cloud characteristics.
Version 0.6	Updated version with Abel contribution in Introduction Section.
Version 1.0	First Release version generated by ATOS.
Version 2.0	Release after proof reading.

Table of Contents

1. Introduction

The main goal of the VELaSSCo project is to design and implement a platform for post-processing operations and visualization of extremely big and distributed datasets resulting from numerical simulations. These data characteristics – big and distributed – require that we design the platform using technologies from the Big Data field.

The platform addresses a real need for the scientific community: as numerical simulations handle problems of ever increasing complexity, distributed computation and data storage has become mandatory in several fields in science and engineering. Currently, High Performance Computing (HPC) is the most available option to address this problem.

Although Big Data solutions have typically been designed for and implemented in Cloud Computing infrastructures, the first prototype of the VELaSSCo platform is designed considering the HPC characteristics. The reason for this is that the scientific community performing numerical simulations is already working with HPC machines, as they are focused in intensive calculations. Thus HPC technology and hardware has traditionally evolved by considering the requirements of these user groups.

However, when planning the deployment of the finished VELaSSCo platform it is natural to consider an extension of its use from a single HPC center to clusters of computers in the Cloud. Their use is becoming more and more popular since some companies (like Amazon, etc…) provide the services to use Cloud virtual machines for computing on demand at very attractive prices.

HPC infrastructures differ from Cloud server-clusters in some characteristics. In this deliverable, the differences between the two scenarios are described, highlighting the requirements they should add (or complement) to the design of the architecture of the VELaSSCo platform, for computation as well as data storage.

It is expected that most parts of the developments done in the frame of the project (considering HPC requirements) will be reusable in a Cloud scenario. This is because an important part of the developments are related to the post-processing of data in a distributed way, and the minimization of data trafic between the storage machines and the visualization client. These two aspects are shared by both worlds: HPC and Cloud.

2. HPC Cloud infrastructure

The main objective of task 2.3 is to design an optimal HPC Cloud infrastructure based on the inputs provided by previous tasks (tasks 2.1 and 2.2) in WP2. This Cloud infrastructure is based on the evaluation of different aspects of HPC, Cloud and HPC Cloud requirements. The specification is based on different aspects: availability, data transfer or network connection. In this document, we consider inputs of the project to define the most suitable architecture. Finally, we report on existing HPC Cloud solutions comparing costs, specifications of each system, and storage provisioning in order to find a commercial/academic HPC cloud infrastructure that matches with the main requirements defined in VELaSCCo.

Information technology (IT) architectures have evolved in the modern age of computer science. One aspect of this evolution is the capability to combine several compute units in order to increase the computational capabilities of an IT system, creating Distributed Systems (DS). DS can be decomposed into several computational classes: supercomputers, clusters, grids, Clouds and Web 2.0 systems. This decomposition is presented in [13]. In this document, supercomputers, clusters, and grid systems are considered HPC systems. In the rest of this section, we will present the characteristics of Cloud, HPC and HPC Cloud infrastructures.

2.1. HPC characteristics

In this section, we will present HPC characteristics from different points of view: business model, architecture, etc.

2.1.1 Business Model

In HPC, the traditional business model is well defined. It is project-oriented, and users of the HPC community typically rent access to HPC resources for specific amounts of time. Most of these IT systems are hosted by national intuitions (see the Top 500 [1]). Academics at the host institutions will usually not have to pay extra fees to use the HPC resources. Some of these infrastructures are also open to external users, and it is possible to rent computation time to them. The cost model for these facilities is mainly focused on price per CPU hour. Other business models have been tested and some of them are used in distributed systems. Billing can be performed on load and historical information of a computation. On specific nodes, fixed cost can be used.

2.1.2 Architecture

The architecture of HPC environments is well established. These systems have been used for a long while in different contexts. To address large scale computing, the architecture is based on many computing nodes, using network resource sharing. The computation

[1] http://top500.org

capabilities of these machines are significantly better than off-the-shelf hardware, e.g., high number of threads and support of SIMD.

The architecture of a HPC system is mainly composed by only a few different types of nodes, such as 3 or 4 kinds. The goal is to reduce complexity of deployment, and increase the homogeneity of the whole system. Direct access to system resources is severely limited, and in order to start computation, it is necessary to submit the jobs to a resource manager. This solution drastically reduces the flexibility of these systems, as users are not allowed to install root software, and only have access to a specific node (called the gateway).

2.1.3 Resource management

The resources of an HPC system are limited at different levels: storage and job execution. A resource manager strictly limits the execution of code on a HPC system. Users submit their applications to the resource manager, which submits the job to the compute nodes once the queue and the required resources are available. The results of the computations are then stored in a specific repository, which can be accessed by the gateway node, and all the local resources (files, folders, etc. needed for an execution) are deleted from the active queue of the HPC system. With this method, a user never interacts directly with a compute nodes.

The storage methodology has been widely studied in the HPC context, and some storage facilities have been developed with this in mind. Two optimizations related to HPC are locality of storage and fast input/output (IO). The IO of these systems has been an important part of the recent research, in order to reduce the latency of the external storage systems. These optimizations have been implemented on specific File Systems (FS), for example NFS, GPF, and Lustre. Depending on the needs of each solution, each FS has their own advantages and drawbacks. Some of them are improved to support data locality, fast writing, fast reading, etc.

2.1.4 Programming model

An HPC environment supports the traditional programming methodologies. Computational power is increased using some HPC-specific libraries. These methods can be based on, e.g., MPI and OpenMP. The scientific community has investigated message passing systems and how to distribute a workflow on a shared architecture, and several algorithms have already been ported to HPC environments [18, 24].

2.1.5 Application model

The application model suitable for a HPC system can be decomposed in two parts: High Performance Computing (HPC) and High Throughput Computing (HTC). HPC run parallel jobs using a strategy based on distribution for example using message passing), while in HTC, the focus is on applications that need to reach a high throughput (to apply read and write operations on data stored in the memory). To achieve this, it is necessary to use SIMD programming methods and infinite band networks. These methods can be combined to reach the highest number of Floating-point Operations per Second (FLOPS).

2.1.6 Security model

In a traditional HPC environment, security is provided by the security policy of the cluster. In many cases, this security is controlled at the access level. As jobs are submitted to the compute nodes using a resource manager, direct interaction with the compute nodes is not allowed here. Thus, the most critical point of this system is located at the frontend of the HPC.

2.2. Cloud computing characteristics

In this section, we will present characteristics of a cloud system at different levels.

2.2.1 Business Model

In the cloud environment, the traditional business model is based on consumption. This business model is similar to that of traditional utilities: a user is charged according to their electricity, gas or water consumption. The cloud business model uses the same paradigm – typically a user is charged according to their CPU and storage usage. This business model allows the provider to exploit the economy of scale. This methodology aims to drive down prices of IT facilities.

Most Cloud environments charge the users regarding to hour, and compute usage. Storage nodes are also charged according to their storage usage, etc. If you need to add more resources to an experiment, you no longer need to worry about installing new hardware – you simply have to swipe your credit card. Due to the topology and scalability of a Cloud IT system, it is possible to virtually extend these infrastructures to the infinite.

2.2.2 Architecture

In a Cloud environment, the architecture is drastically different than HPC. The infrastructure of a cloud system mainly refers to a large pool of commodity nodes. The architecture of the global system is customized to fit a project's needs related to storage and computation. These resources are accessed using standard protocols (for example a Web 2.0 interface based on a REST API). Cloud systems can be classified in different architectures such as Infrastructure, Platform, and Software as a Service (IaaS, PaaS, and SaaS, see D 2.1). The choice of a specific architecture depends on the user needs. Each of these architectures implies a specific infrastructure and set of software.

A cloud environment is based on more traditional IT equipment compared to HPC. And in most cases, these providers use commodity hardware. Moreover, to ensure a highly flexible IT system, these solutions are massively based on virtualization. In these environments, a user has more hardware flexibility.

2.2.3 Resource management

In a Cloud environment, virtualization is one of the most important ingredients. Virtualization is used to unify the operating environment as an abstract layer above the available hardware. This strategy improves security, manageability and isolation of resources for separate users. Virtualization also provides a more efficient model in which to use available resources. It ensures quick recovery from failure and a high level of responsiveness by using automation.

In Cloud systems, storage is centralized with a virtual file system (FS). The choice of the virtual FS is based on different features: mapping, partitioning, querying, caching and replicating data. The location of datasets is also important for algorithms used on Cloud infrastructures. For example, with the MapReduce programming model, the locality of data can have an important impact. The storage solution used with this programming model is tightly linked to computation: computation is performed (and potentially even moved) where the data is located.

2.2.4 Programming model

Running an application in a Cloud infrastructure can be more complex, and less intuitive than HPC. With Cloud computing, two programming models exist: a simple one and a complex one. For the simple strategy, a top language is used, and the complexity of programming is smoothed by the available toolkits (mainly for SaaS or PaaS). In the second approach, the developer is in charge of the computation decomposition. This decomposition entails an independent computation on all nodes, where the data is stored. It is the main characteristic of the MapReduce (MR) programing model. With the Map function, results are extracted locally on all the compute nodes. Then the Reduce function merges data into one single instance. With this programming model, different features have been implemented such as automatic partitioning, fault tolerance, etc. Different solutions have been presented in the literature, and one of them has become a standard: Hadoop. It is an open source implementation of the MR paradigm, with several extensions which simplifies the programming of certain jobs. For example, Hive brings SQL to Hadoop.

2.2.5 Application model

A Cloud system can use different application models in its infrastructure. This extensibility is enabled by the different available architectures: SaaS, PaaS, IaaS. Within a Cloud system, it is possible to deploy a HPC workload and even surpass it. Unfortunately, current Cloud implementations are not able to reach the same performance as HPC system, as they have more latency and less CPU capabilities.

In a Cloud environment, the submission model is very different than for HPC: users share resources concurrently. The resources are divided among users and new issues arise, such as latency and Quality of Service (QoS). Thus algorithms and methods have to be redesigned with these in mind.

2.2.6 Security model

In a Cloud environment, the concept of security is more complex than for a traditional IT system. Security management and evaluation is not a trivial task, and several levels of security are present. Two points have to be considered: access point and provider. In most cases, security is provided through a simple web form. Thus, the security is managed by web queries. A first security hole can appear in the communication with the server. A second important vulnerability can be introduced at the virtualization level. But in most cases, big companies are able to ensure a high level of security through state of the art solutions.

The security of a Cloud also depends of its classification as private, public, community or hybrid. For private Clouds, a single institution owns the IT system, and a set of specific users can access to the infrastructure. For a public Cloud, public institutions (academics, national laboratory, commercial companies, etc.) are in charge of the systems. They are often open to the public, and a variety of users. A community Cloud is a system owned by a specific community, only users of the community could have an access to this Cloud. Finally, hybrid Clouds are mixed version of the previous methods.

2.3. HPC Cloud characteristics

The Cloud paradigm is always evolving, as new features and services appear. One of them is HPC Cloud or HPCaaS (HPC as a Service).

Amazon is a pioneer in the HPC cloud. They provide one of the first HPC infrastructures available as a Cloud service. This specific Cloud system is composed of a dedicated set of nodes, with GPUs and high-end CPUs, which provides a low latency bandwidth. Several optimizations are proposed by Amazon in order to increase the computational capabilities of these nodes. The service offered by Amazon provides different instances of nodes for the HPC-cloud.

We consider HPC Cloud a new class of distributed systems, combining features from both HPC and Cloud computing. It combines the flexibility and affordability of Cloud infrastructures with the low latency and high-performance of HPC systems.

2.4. Specificities of Big Data in HPC Cloud

Cloud system is a promising approach to reduce cost of an IT system. It is not anymore necessary to maintain and update a complex IT system, and cloud-HPC systems have the same advantages than clouds. The company in charge of a cloud infrastructure performs all the maintenance and updates. This new HPC area is characterized by some specificities [24, 25].

Cloud infrastructures enable to reduce the exploitation cost of HPC systems, and this hardware also becomes more agile. It is possible to increase the number of nodes easily. Moreover, due to the services offered to the customers, it is possible to extend the storage of the system. These platforms also support the features of traditional HPC: job submission using a queue system and acceleration programming methods (using MPI or OpenMP, for instance). But with this specific IT system, developers have to deal with an important problem: locality of nodes. Communication patterns are not similar to traditional HPC. Depending on the cloud HPC provider, users can avoid spending time to boot machines, deploying operating systems etc. This is provided by companies which deliver bare metal compute nodes. For virtual nodes, these issues are mandatory, but allow a better control on nodes (which OS, root accesses, etc.).

Before switching to a cloud HPC, it is also necessary to answer some questions, to ensure a real interest in deploying a specific application in the cloud. These questions are: will moving to the cloud enable new business model? Will increases in agility and elasticity overcome the cost of moving the applications to the cloud?

After this step, it is necessary to identify if an application can be ported to the cloud. One of these requirements is full automation. An application needs to be fully automatic: from the deployment to the execution. The human interactions are forbidden in this specific case. Applications have to be standard, and not fit with very specific features of architecture. Clouds bring to the HPC four important features:

- On-demand computing: the necessary resources can be provided on demand to the users.
- Resources pooling: the provider pools resources to serve multiple users.
- Rapid elasticity: the system can be easily extended or released.
- Network access: nodes can be accessed using a standard network system.

Thinking now in the generic ecosystem for Big Data, and taking into account the above comments on HPC and clouds, we can establish some specificities of Big Data over an HPC cloud.

Considering that Big Data is focusing on data, while HPC is focusing on computing we will try to take advantage of both parts to create a Big Compute integrated platform. This platform will be able to deal with dense compute (HPC) and dense data storage (Big) coming from simulations. To integrate the Big Data framework into the HPC cloud some important aspects should be taken into account related to network, storage or resource management.

Related to network integration we have to consider that common Big Data ecosystems were designed for Ethernet environments using sockets as communication mechanism. In that sense, the HPC ecosystems require a specialized interconnect mechanism for high performance (as InfiniBand or High Performance Ethernet). To address this issue, RDMA for Apache Hadoop provides an excellent high speed, low latency interconnect option for Big Data platforms on HPC. An approach based on “RDMA support in conjunction with several optimizations to HDFS, MapReduce, HBase and other components for Infiniband or 10 Gigabit networks” is proposed in [8].

Related to storage the generic Big Data ecosystems provide a pluggable file system abstraction that interoperates with any POSIX compliant file system. Thanks to this characteristic, most parallel file systems can be used with Hadoop, although in these cases

Hadoop has no information about locality of the data maintained at parallel file system level. Examples of that parallel file system supported on top of Hadoop are Lustre [5] provided by Intel and GPFS [6] provided by IBM. In addition to that "*another optimization concerns the MapReduce shuffling phase that is carried out via the shared file system*" is proposed in [7]. Moreover, as the processing of data in HDFS, is done locally minimizing the movement through the network, the horizontal scalability of parallel file systems is often a restriction compared to HDFS. Thus, parallel file systems are more reliant on fast interconnects than HDFS.

Finally, related to resource management system, the Hadoop-level scheduler can be deployed on top of the system-level scheduler. Examples of resource managers supported on Big Data ecosystems are Condor or SLURM. Similarly to the case of storage, the main disadvantage here is the loss of data-locality from the system-level scheduler perspective. In the case where HDFS is selected, before data can be processed it needs to be copied into HDFS, which implies a significant time consuming. Some resource management systems, such as SAGA-Hadoop [2], JUMMP [4] or MyHadoop [3], can be applied on HPC environments.

The approach that we will follow in the VELaSCCo platform for integrating Big Data and HPC system will be based on Hadoop-YARN, probably with Lustre on top of HDFS.

2.5. Different HPC Cloud options

As stated in the previous section, cloud systems have evolved in the previous years. Different solutions have been proposed and regarding to needs, it is possible to deploy a cloud infrastructure in some scientific case. Cloud is not anymore limited to some specific requirements. Scientific community can now have access to these cloud resources, and achieve computation on this specific hardware.

The evolution of usage in cloud systems has been ported by their high flexibility. Virtualization has been an important paradigm of cloud enabling to deploy virtual super computer suitable for specific needs. HPC-cloud is a specific feature of cloud infrastructure. As different provider already exists, here we will present some of these and we will depict their advantages and drawbacks. This section is a summary of the document [21].

Penguin computing

Penguin Computing is a company, which is specialized on delivering HPC and cloud solution based on a Linux environment. This company can provide different compute resources regarding to needs of compute and user profiles. Penguin is mainly dedicated to deliver a cloud-HPC solution; penguin offers a bare-metal hardware, similar to a regular HPC cluster. With this provider, virtualization is used but only at the login level. This virtualization method does not impact performances of the final cluster. Thus, execution, and computation are directly applied to the hardware nodes. To fit with requirements of different users, this provider proposes a large set of HPC servers. As the computational model, networking is performed on a real hardware, and virtualization is not used. This provider is an old provider in term of cloud HPC. And it is seems to be the most experimented provider in this domain.

R-HPC

R-HPC is another company, which provides some HPC resources to their customers. They provide HPC as a Service, and customers can rent HPC resources in the cloud. This provider offers two different services: a shared cluster and a virtual private cluster. The shared instance is similar to a traditional HPC instance but in a cloud infrastructure. This company proposes different HPC instances. This provider has different hardware composed by multi- and many-core resources. This company provides direct hardware access, and no virtualization methods are used. An interesting feature with this provider is the windows support. The pricing of computation depends on submitted jobs (the economic model is a pay as you go).

Amazon

Amazon is one of the most well known providers of cloud services. In the earlier age of cloud, Amazon only provides two cloud facilities: a storage and a compute one. But now, to fit with modern needs of scientist, they build a specific HPC cluster in their infrastructure. This cluster was listed on the top500 at the rank 101. It is a self-made cluster specially designed to run HPC application. This instance offers a huge amount of compute resources that are more suitable with scientist requirements. Their product is decomposed in three distinct instances: one dedicated to heavy storage, one devoted to GPU computation, and the last one concerns improvements on IO operations.

Univa

Univa is grid Provider Company. This company delivers different services one related to a cloud environment, and the other one related to HPC environment. The HPC feature is delivered with the Grid compute engine. Their solution enables to deploy a HPC system over a Grid system in one click. This system enables to easily deploy and manage cloud servers. A template of servers improves this deployment. The HPC system provides by this company is based on virtual machines.

Sabalcore

Sabalcore offers professional High Performance Computing services over the Internet. They offer a full cloud HPC infrastructure. This company offers an on-demand cloud system, with preinstalled OS and applications. No virtualization is used, and computation is directly applied on the metal. This provider also gives access to a large set of preinstalled applications. Resources of this system are interconnected using an Ethernet or an infiniband network. Intel or AMD CPU and NVidia graphics card compose the nodes of this system.

SGI

SGI is another important service provider; traditionally, they deliver very high powerful computation nodes. It is a super computer leader in super computing. SGI has developed a cluster named Cyclone[2], based on their specific super computer technology. This system is based on a specific shared memory model, which enables to reach a large amount of shared resources. This IT system offer a direct access to the hardware with dedicated improvement provides by SGI compute nodes.

In the Scientific community

The HPC Cloud infrastructures have also been evaluated by the scientific community.

In [15], Rehr *et al.* present an evaluation of the Amazon HPC instance to run a typical HPC application. They use FEFF[3] to evaluate whether Amazon can be used as an HPC system. They present an execution benchmark of a serial and a parallel version of this software deployed on an EC2 cluster. This evaluation also provides a study on inter- and intraconnection of nodes using a MPI (Message Passing Interface) benchmarks. The result of this study shows scientific computation in the cloud is a viable alternative, though the Amazon service has some drawbacks. The current version of the HPC Cloud system does not support 64-bits. Another important point is focused on compiler; Amazon does not provide a high performance compiler and the automation tools provided by it needs to be enhanced.

In [16], the author presents a new Cloud infrastructure. This platform uses the Azure ecosystem, provided by Microsoft). This Cloud system was used to provide HPC resources where heavy computation is necessary. But this specific deployment needs to rethink the application workloads.

In [17], the authors deployed a CFD solver on the Amazon HPC Cloud. Their tool runs on both GPUs and CPUs. In their evaluation, they show that the Amazon cluster provides 70% of parallel efficiency on 8 cores and 50 percent on 256 cores. On the GPU instance, they are able to reach 75% of efficiency on 16 GPUs. All these results show that the Amazon HPC cluster seems to be a viable infrastructure for HPC.

In [18] the authors proposed an evaluation of scientific workload on the Amazon EC2 infrastructure. They show that the performance reached on a cloud infrastructure is similar to performance available on a HPC system. Their evaluation also considers the cost of the solution, concluding that a cloud storage system is a viable strategy for storing data and reducing the cost of a compute solution.

In [19], the authors provide an interesting comparison between HPC, Grid and Cloud systems. They propose a new architecture composed by a hybrid approach to achieve a better efficiency and predictions on execution workloads. It is a specific HPC solution tailored for the Cloud environment. This solution is designed using three methods: scale-up and down (increase and decrease the size of the cluster), scale-out (distribute the workload on external services), and personal virtual clusters.

Edinburgh Compute and Data Facility (ECDF)

ECDF is the institutional service provider of High Performance Computing and Research Data Services to the University of Edinburgh and associated institutions. One of the principal

[2] http://www.sgi.com/company_info/newsroom/press_releases/2010/february/cyclone.html
[3] http://feffproject.org

services provided by ECDF is a high-performance cluster of computers named “Eddie”.The Eddie cluster is composed of worker nodes that run an industry standard Linux-based operating system, specifically Scientific Linux. Currently, the Eddie cluster comprises two main parts:

- Mark2Phase1 - 130 IBM dx360M3 iDataPlex servers, each with two Intel Xeon E5620 quad-core processors. All these nodes are connected by a GigabitEthernet network with a 10 Gigabit network core.
- Mark2Phase2 - 156 IBM dx360M3 iDataPlex servers, each with two Intel Xeon E5645 six-core processors. All these nodes are connected by Gigabit Ethernet network, and 68 of them are also connected by a QDR QLogic Infiniband network for Message Passing Interface (MPI) jobs.

Eddie can also handle distributed memory parallel tasks where fast inter-node communications arerequired. The Infiniband equipped part of the cluster is especially well-suited for this role. There are 68 nodes in this section giving the capability to run up to 816 way parallel jobs.

3. Comparison of Classic HPC vs HPC Cloud

In this section, we present the current state-of-the-art related to classical HPC. Then we provide a comparison with the new HPC Cloud paradigm in order to find which features are the most suitable for our project, as appointed in Task 2.3.

3.1. Classic HPC

High Performance Computing (HPC) can be defined as a paradigm where a set of machines, clusters or supercomputers are gathered into specific networks to work together in order to solve specific tasks or problems which require high computing capabilities. These types of infrastructures are usually expensive to maintain due to their specific software, architecture deployments, algorithms implementations, etc. In this context, the appearance of Cloud computing techniques adds a new potential to HPC environments which could facilitate their adoption by a wide range of users.

Some researchers have already argued about this new scenario [10], exposing how cloud computing new approaches like Infrastructure, as a Service (IaaS[4]) and Platform as a Service (PaaS[5]), allow scientists and researchers to deploy their HPC applications on cloud environments, thus reducing considerably the costs and disadvantages of owning the infrastructures. This main outcome is allowing new actors like industries or big companies for business and analytics to use HPC applications in Cloud infrastructures, in addition to classic academic laboratories for scientific research.

Finally, the main advantages that new Cloud computing infrastructures offer to HPC applications are the benefits of virtualization, elasticity of resources and reduction of cluster configuration cost and time to HPC applications users. However, some disadvantages have been detected, like poor network performance, performance variation and OS noise, as reported on [9].

Table 1: Grid vs. HPC vs. Cloud

Category	Grid	HPC/ Cluster	Cloud
Size	Large	Small to medium	Small to large
Resource Type	Heterogeneous	Homogeneous	Heterogeneous
Initial Capital Cost	High	Very High	Very low
Typical ROI	Medium	Very high	High
Network type	Private Ethernet-based	Private IB or proprietary	Public Internet Ethernet-based
Typical Hardware	Expensive	Very expensive - "top of the line"	Usually VM's atop of hardware
If I didn't know any better:	"faster workstations"	"supercomputer"	"bunch of VMs"
SLA requirement	High	Strict	Low
Security Requirement	High	Very low - but typically high	Low

Figure 1. Grid vs. HPC vs. Cloud comparative table [11]

[4] Infrastructure as a Service
[5] Platform as a Service

Figure 1 (taken from [11]) shows a comparative table between the main computing paradigms which coexist currently. As shown in Figure 1, the computing paradigms compared are Grid, HPC and Cloud Computing. The differences identified among these approaches are the following:

- Grid computing is usually more expensive because it can be customized and adapted with proprietary hardware and software. It is designed to perform on single tasks or problems at the maximum of its capacity on a private network or set of nodes, because reusability and flexibility is not a key concern in this case.
- High Performance Computing typically requires a large number of machines, nodes or supercomputers, so its cost is high and could be as expensive as GRID. HPC services are also offered on private networks and dedicated nodes, so they still lack the versatility of new Cloud approaches.
- Unlike GRID and HPC, Cloud computing is composed by Virtual Machines deployed over physical nodes which can be assigned dynamically, depending on end-users requirements. This versatility and flexibility is the main advantage regarding the previous paradigms because it provides a multipurpose hardware infrastructure and hence, it reduces considerably your initial investment.

3.2. HPC in a Cloud

As described in [12], HPC and Cloud computing main features and characteristics do not exactly match at first sight, so they could be considered different approaches for different scenarii. Despite this, they share some required features like scalability and reliability as they confer to the system the desirable aspects to allow good performance during computing, whereas other aspects such as hardware and software customization, availability, etc. show completely different characteristics. Figure 2 display the intersection between the two paradigms:

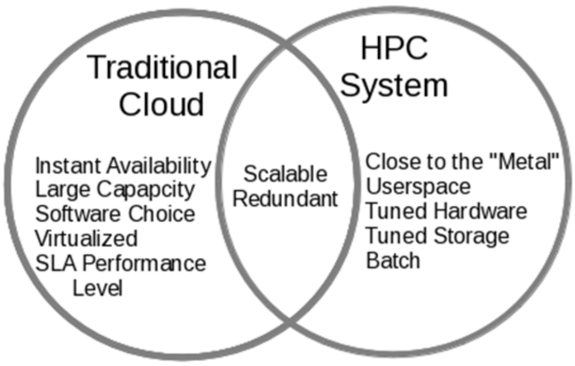

Figure 2. Cloud services and an HPC system[6]

[6] http://clustermonkey.net/Cloud/Grid/will-hpc-work-in-the-cloud.html

Regarding scalability, it could be defined as the ability of a system to increase its computing capacities dynamically and safely. Cloud environments usually accomplish this easily, creating new virtual machines instances, redeploying old ones already created or replacing some obsolete instances. HPC applications scalability is more related to the number of cores or processors which can be assigned to single tasks. In both cases, cloud and HPC clusters use these mechanisms to allocate the computing resources required, but manage this scalability using different approaches. Whereas Cloud infrastructures are set by adding, deleting or replacing VMs instances, the HPC clusters use a resource scheduler to provide the physical resources to the end-user application. Due to this dynamic principle, cloud computing is usually able to offer more compute capacity than large HPC clusters.

Redundancy aims to control where applications are running, in terms of hardware and storage resources. The main objective is detecting broken or failed hardware parts, dependencies among the nodes, etc. in order to avoid undesirable redundancy and application failures. Both technologies are able to handle and manage their physical resources to do so.

4. Data Storage in the VELaSSCo HPC Cloud

After studying the HPC Cloud commercial and academic options in Section 2.5, ECDF has been selected as the HPC Cloud solution for the VELaSSCo project.

VELaSSCo will use the Eddie cluster provided by UEDIN which is connected to a large amount of high performance disk storage via a number of dedicated machines. These machines then share the data across the cluster using a parallel file system. ECDF will allow the support staff to run the infrastructure, perform data ingestions, and help with workflow integration.

Initially, data storage in Eddie Cluster will be done by means of the raw device mapping (RDM) approach as it shown in Figure 3.

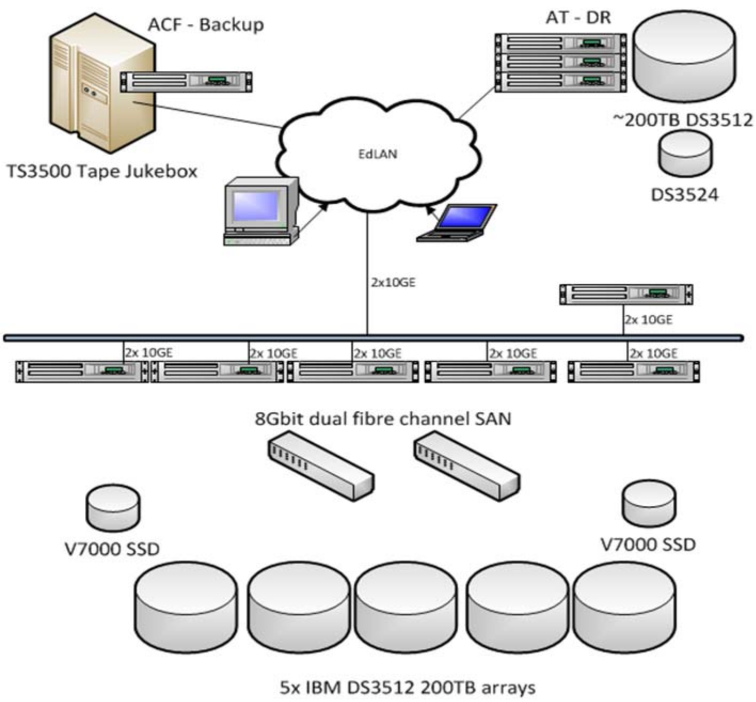

Figure 3. RDM storage in Eddie Cluster

Taking the definition of RDM provided in [20], "*RDM is an option in the VMware server virtualization environment that enables a storage logical unit number (LUN) to be directly connected to a virtual machine (VM) from the storage area network (SAN).*

RDM is one of two methods for enabling disk access in a virtual machine. The other method is Virtual Machine File System (VMFS). While VMFS is recommended by VMware for most data center applications (including databases, customer relationship management (CRM) applications and enterprise resource planning (ERP) applications, RDM can be used for configurations involving clustering between virtual machines, between physical and virtual machines or where SAN-aware applications are running inside a virtual machine.

RDM, which permits the use of existing SAN commands, is generally used to improve performance in IO-intensive applications. RDM can be configured in either virtual compatibility mode or physical compatibility mode. Virtual mode provides benefits found in VMFS, such as advanced file locking and snapshots. Physical mode provides access to most hardware functions of the storage system that is mapped."

For the VELaSSCo platform, we plan up to 200TB (currently 50 TB) of data storage as a working capacity for visualization of data sets, with data backup for appropriate project golden copy data and a dedicated LTO6 tape jukebox to ingest the provided data sets. The use of a dedicated tape jukebox will allow easy ingestion of data, while a high bandwidth connection with the ECDF "Eddie" cluster and the HECToR and ARCHER UK supercomputers will allow easy transfer of the data to and from the high-performance HPC storage.

5. Conclusions

In this deliverable, we have presented a summary of the efforts provided in Task T2.3 related to the “Data storage design for the HPC cloud infrastructure for engineering analysis”.

We presented the characteristics of cloud, HPC and HPC Cloud infrastructures considering relevant aspects such as Business Model, Architecture, Resource Management, Application Model, Programming Model and Security Model. Subsequently, the specificities of Big Data for HPC Cloud are described as well as the different HPC Cloud options explored in terms of industrial and academic providers.

Once the HPC Cloud characteristics and options have been studied and reported, we describe the current state of the art of the classical HPC paradigm and compare it with the new HPC Cloud approach in order to find the main topics, highlights, features, etc. which will be more suitable within VELaSSCo platform.

Taking into account the HPC Cloud commercial and academic alternatives reported in Section 2.5, ECDF was selected as the HPC Cloud solution for VELaSSCo.

Finally, the specific data storage approach to be followed in the VELaSSCo HPC Cloud is reported in Section 4.

6. References

[1] The Apache HBase Reference Guide: http://hbase.apache.org/book/book.html

[2] SAGA, "SAGA-Hadoop," https://github.com/drelu/saga-hadoop, 2014.

[3] myHadoop, "myhadoop," https://portal.futuregrid.org/tutorials/running-hadoop-batch-job-using-myhadoop, 2013.

[4] W. C. Moody, L. B. Ngo, E. Duffy, and A. Apon, "Jummp: Job uninterrupted maneuverable mapreduce platform," in Cluster Computing (CLUSTER), 2013 IEEE International Conference on, 2013, pp. 1–8.

[5] O. Kulkarni, "Hadoop MapReduce over Lustre: high performance data division," http://www.opensfs.org/wp-content/uploads/2013/04/LUG2013. Hadoop-Lustre. Omkar Kulkarni.pdf, 2013.

[6] P. Zikopoulos et al., Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. McGraw-Hill Osborne Media.

[7] "MapReduce and Lustre: Running Hadoop in a High Performance Computing Environment," https://intel.activeevents.com/sf13/connect/ sessionDetail.ww?SESSION ID=1141, 2013.

[8] Y. Wang et al., "Hadoop acceleration through network levitated merge," in Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis, ser. SC ’11. New York, NY, USA: ACM, 2011, pp. 57:1–57:10.

[9] "Evaluation of HPC Applications on Cloud". Abhishek Gupta, Department of Computer Science University of Illinois at Urbana-Champaign. Dejan Milojicic, HP Labs Palo Alto.

[10] "The Who, What, Why and How of High Performance Computing Applications in the Cloud". Gupta A., V. Kale L., Gioachin F., et al. HP Laboratories, HPL-2013-49.

[11] A. Bhatele, S. Kumar, C. Mei, J. C. Phillips, G. Zheng, and L. V. Kale, "Overcoming scaling challenges in biomolecular simulations across multiple platforms," in Proceedings of IEEE International Parallel and Distributed Processing Symposium 2008, April 2008, pp. 1–12.

[12] "HPC Technology Evaluation and Application Strategy" http://www.nag.co.uk/hpc-technology-evaluation-and-application-strategy

[13] I. Foster, Y. Zhao, I. Raicu, and S. Lu. "Cloud computing and grid computing 360-degree compared". In Grid Computing Environments Workshop, 2008. Pages 1-10, Nov 2008.

[14] http://www.cloudscaling.com/blog/cloud-computing/grid-cloud-hpc-whats-the-diff/

[15] Rehr, J. J., Vila, F. D., Gardner, J. P., Svec, L., & Prange, M. (2010). « Scientific computing in the cloud". Computing in Science & Engineering, 12(3), 34-43.

[16] Khalidi, Y. "Building a cloud computing platform for new possibilities". Computer, 44(3), 29-34. 2011.

[17] Zaspel, P., & Griebel, M. (2011, November). „Massively Parallel Fluid Simulations on Amazon's HPC Cloud". In: 2011 First IEEE Intl. Symposium on Network Cloud Computing and Applications (NCCA), pp. 73-78.
[18] Juve G., Deelman E., VAHI K., et al. "Scientific workflow applications on Amazon EC2". In: Proc. 2009 5th IEEE International Conference on E-Science Workshops. p. 59-66.
[19] Matescu G., Gentzsch G., and Ribbens C.. 2011. "Hybrid Computing-Where HPC meets grid and Cloud Computing". Future Gener. Comput. Syst. 27, 5. May 2011.
[20] http://searchvirtualstorage.techtarget.com/definition/Raw-device-mapping-RDM
[21] http://www.admin-magazine.com/HPC/Articles/Moving-HPC-to-the-Cloud
[22] Jha S., Qiu J., Mantha P., Fox G. "A tale of two-data-intensive paradigms : applications, abstractions and architectures". In: Proc. IEEE Intl. Conf. on Big Data. Anchorage (AK). June 2014.
[23] Ahuja S., Mani S. "The state of high performance computing in the cloud". In: Journal of emerging trends in computing and information sciences. Vol. 3, No.2. February 2012.
[24] Deelman E., Juve G., Malawski M., Nabrzyski J. "Hosted science: managing computational workflows in the cloud". In: Parallel Processing Letters. Vol. 23, No. 2.June 2013.
[25] Qiu J., Jha S., Fox G. "HPC-ABDS : the case for integrating Apache Big Data Stack with HPC". In: NIST 1st JTC SGBD Meeting. San Diego Supercomputer Center. San Diego (CA). March 2014. http://jtc1bigdatasg.nist.gov/_workshop/15_NIST-HPC-ABDS.pdf