RP1: Cloud Storage and Computing Report

Tom Hendrickx
System & Network Engineering
Universiteit van Amsterdam
tom.hendrickx@os3.nl

February 2, 2009

Abstract

This document contains the results of my research project on *cloud storage and cloud computing* for UNINETT[1], Norway. Cloud storage and cloud computing is an upcoming concept, which promises an unlimited amount of accessible resources over a network (Internet or local network). The objective of this research was to define the concept, to present an overview of its capabilities, and to describe its applicability for the higher education and research community in general and for an NREN, like UNINETT, in specific.

1 Acknowledgments

I would like to thank the following people for their time and help during my research project:

Jan Meijer, project manager at UNINETT - for welcoming me to Trondheim in Norway, for his guidance throughout the whole research project and his critical questions, which has broadened my insight into the matter and beyond.

Jaap van Ginkel, Professor at UvA - for his recommendations and help in receiving a research project at UNINETT, under the supervision of Jan Meijer.

Jacko Koster, project manager at UNINETT Sigma - for giving some insight in the NorStore[2] project and storage needs in general.

Anne C. Elster, HPC group leader and Professor at NTNU[3] - for sharing her view on cloud computing in the field of high performance computing.

2 Preface

This report is written as part of my first research project, for my study of System and Network Engineering at the University of Amsterdam. I analysed the applicability of the cloud concept for an NREN, which stands for National Research and Education Network. The timeframe of the project was one month, of which the first and fourth week were done at my home in Belgium, to read into the subject and finalize the report. Week two and three were done at the UNINETT office, Trondheim (Norway), under supervision of Jan Meijer. During my stay in Trondheim, I was invited to attend a meeting with Purity IT about "Green IT and the next generation of storage solutions", to a 4K demo-film and to a meeting with Jerry Sobieski of NORDUnet and representatives of the Norwegian University of Science and Technology about the future prospects of the network. They were very informative and useful to broaden my insight in storage and HPC.

Contents

3 Introduction

Ever since IT has made its way into companies, schools, and other socio economic structures more money and resources has been invested into their IT infrastructure, personnel and knowledge. But the last decade there has been a severe change. The tendency has grown to outsource more and more on different areas, which do not belong to the core business, going from IT personnel to even IT infrastructure. The latest evolution is the cloud concept, where through infrastructure as a service (IaaS), platform as a service (PaaS), software as a service (SaaS) and many more become available as an ever increasing business. The cloud concept is still rather new, but there is a growing interest in its promises and different enterprises have taken it up as the next level of IT.

For my project, UNINETT is interested in a broader definition of the cloud concept and in the possibility of how it can be used in the context of the higher education and research community in general and of an NREN in specific. When comparing the latter to a commercial vendor, an NREN has a special position which plays in favor of its constituency for a couple of reasons. First of all, they already have an excellent network at their disposal, with high-bandwidth connections. Second, an NREN is not a mere supplier, but it works together with its clients and is generally not in pursuit of profit. Aside from these advantages, which makes it a trusted third party for its target group, it has the distinguished feature of not easily going bankrupt. These characteristics are vital to how an NREN's cloud model could become unparalleled by those of other providers. Aside from defining the concept, I will use these characteristics to determine if cloud computing and cloud storage can be a fit solution for UNINETT.

3.1 Research questions

The main research question for my project is:

> *Define the cloud concept in context of the higher education and research community in general and of an NREN in specific.10*

To be able to answer this question to its fullest, the following *sub-questions* have been examined:

- How can the cloud concept be defined in general?5.3
- What is the difference between a cluster and a cloud filesystem?6.3.1
- How can a cloud be built provided that the following properties need to be guaranteed:6

  - Privacy;
  - Security;
  - Availability;
  - Reliability;
  - Scalability;
  - Quality of service?

- Is a cloud existing out of commodity computers feasible? 6.5.1
- Can it be used as an alternative for:
  - Regular resources in a regular datacenter?7.2
  - An addition to a national grid such like NorStore7.2 or NOTUR[4]7.3?
- What advantages are there in joining an NREN's cloud compared to:
  - Keeping everything local;5.1
  - Joining a commercial cloud.7.1
- Which are the possible legal concerns?8

To provide the reader with an answer to these questions I examined different existing clouds (like Amazon's EC2[5] and S3[6]), and different storage methods (like Google File System[7] and traditional cluster file systems[8]) in order to find a scalable storage solution. For high performance computing, I compared the possibilities of supercomputers, clusters and grids to those of cloud computing. With these results and the characteristics of an NREN I figured out where and how the cloud could be used to provide services to their constituency.

3.2 Document layout

The following section describes the context in which this research was performed. Section 5 defines concepts such as cloud computing, cloud storage, SaaS, IaaS, etc to get familiar with the subject. Section 6 starts with describing the common features of a cloud architecture. Afterwards the architecture of a storage cloud will be handled more specifically in section 6.2, by regarding traditional storage architectures such as a storage area network and a network attached storage. A clustered commodity architecture is then presented as a more fitting architecture for cloud storage in section 6.3. Thereupon will section 6.4 present the possible uses of cloud computing with regard to providing high performance computing and virtual machine servers. Both storage as computing architectures are based on commodity hardware and as such the feasibility and the manageability of commodity hardware will be considered in section 6.5. Finally section 6.6 discusses the current level of interoperability between different cloud architectures. Section 7 examines the cloud's concepts, with regard to the needs of an NREN and more specific to the needs of UNINETT. This section also discusses how storage and computing services can be brought together and what would be needed to obtain the most out of the cloud's advantages. Section 8 presents some legal questions which should be solved, before providing any cloud services. Section 9 discusses future work and section 10 will conclude this report.

4 Research context

4.1 UNINETT and related projects

> UNINETT[9]is a Norwegian government-owned company responsible for deploying and maintaining a national research based computer network, to which all of Norway's colleges and universities are connected. They also conduct network research and pilot projects related to high speed connectivity.
>
> The backbones of UNINETT are typically 1 or 2.5 Gbit/s fibre optic links. For institutions further away from the backbone, the maximum capacity is 155 Mbit/s. UNINETT is connected to other similar networks in Nordic countries via NORDUnet, which are both part of the Internet. The main links are now being upgraded to 10Gbit/s between the major cities for example: Oslo, Bergen, Troms and Trondheim.

UNINETT itself is interested in offering services such as storage and computing over the cloud as these would extend their current services. For the moment there are quite a few Norwegian projects who would benefit from such a cloud :

NORGRID - Norwegian GRID Initiative[10]

> The aim of the NorGrid project is to establish and maintain a national grid infrastructure in Norway. The project will provide grid-based services that enhance the utilization of the resources in the national infrastructure for computational science and facilitates transparent sharing of data between user groups.

NorStore - Norwegian Storage Infrastructure

> The objective of the project is to establish and operate a national infrastructure that provides non-trivial services to scientific disciplines with a variety of needs for storing digital data. The infrastructure will provide easy, secure and transparent access to distributed storage resources, provide large aggregate capacities for storage and data transfer, and optimize the utilization of the overall resource capacity. The project will be a broad and nationally coordinated effort.

NOTUR - The Norwegian Metacenter For Computational Science

> The Notur project provides the national infrastructure for computational science in Norway. The project provides computational resources and services to individuals or groups involved in

> education and research at Norwegian universities and colleges, operational forecasting and research at the Meteorological Institute (met.no) and to institutes and industry that collaborate with the project.

The following section will present the current and future needs of one of these projects: NorStore, as well as Norway's part in storing CERN-data, to give a first impression of the future storage needs in Norway.

4.2 Datagrowth in Norway: Some prospects

NorStore has for the moment a capacity of 600TB, which is a combination of tape and disk storage and of which 100TB is currently being used. While new proposals will be made to make use of its storage in the coming month, exact numbers are still unknown. Although, the storage demand is expected to increase enormously. Aside from its current storage, there is also 200TB of climate data stored in tape robots, which is considered to be included in NorStore.

CERN, the European Organisation for Nuclear Research, runs the biggest particle accelerator (LHC - Large Hadron Collidor[11]) in the world to study the smallest known particles, to improve our knowledge about the laws of nature. The data generated by this study will be so enormous, that it is almost impossible to store or process all at one place. Predictions are that it will generate about 15PB (petabyte) yearly, which will be divided over institutions situated in 33 different countries. Norway is part of this by means of the Nordic Data Grid Facility which is:[12]

> a collaboration between the Nordic contries (Denmark, Finland, Norway, Sweden).
>
> The motivation for NDGF is to ensure that researchers in the Nordic countries can create and participate in computational challenges of scope and size unreachable for the national research groups alone.
>
> NDGF is a production grid facility that leverages existing, national computational resources and grid infrastructures.
>
> To qualify for support research groups should form a virtual organization, a VO. The VO provides compute resources for sharing and NDGF operates a grid interface for the sharing of these resources.
>
> Currently, several Nordic resources are accessible with ARC and gLite grid-middleware, some sites with both.

> Today, the first operational user of the NDGF is the Nordic High Energy Physics community - the ALICE, ATLAS and CMS Virtual Organizations - through the operation of the Nordic Tier-1, which together with the Tier-0, CERN, and the other 10 Tier-1s collects, stores and processes the data produced by the Large Hadron Collider Experiment at CERN.

At this moment Norway's Tier-1 stores about 400-500TB of physics-data generated by CERN. This makes CERN one of Norway's biggest storage consumers of the moment. Predictions are even an increase to 1,5PB to 5PB in 2010 and this is only considering Norway's part. At least half of this will have to be directly accessible, which means it has to be stored on disk.

A point of interest for the higher education is storing video formats. For example, the highest-quality video has a resolution of 4K at the moment. This means that a one hour film has the size of 1TB when it is compressed, but uncompressed it has the size of 7TB. And when considering that:

- newer formats will keep getting bigger,
- schools have to hold on to lots of student material for many years to come,
  - Student films;
  - ...
- this material may have to be instantly accessible,

it is clear that huge amounts of accessible storage will be needed in the future.

5 The cloud concept

What is the cloud concept?
Today it is used as a popular marketing term which lots of people in sales departments tend to use for every kind of service which is provided over the internet. So there is not really a single straightforward definition to the term. In the continuation of this section most uses of the cloud and the ideas behind it will be discuss, whilst this section will conclude with a more general definition of the cloud's concept.

5.1 Benefits of outsourcing to the cloud

Until a couple of years ago, the internet was mainly used to provide a network connection from one side of the cloud to another. Now, as the cloud is emerging, we will be able to think of the internet as a source of not only bandwidth, but also a storage and computing facility. With only an internet connection, it will become possible to purchase services as storage and computing resources dynamically to one's needs. One will be able to counteract the changing needs of their organization, without acquiring the necessary infrastructure and this on a big scale. Whilst not having to provide the infrastructure oneself, a lot of other problems are also "outsourced" to the provider. The following problems also reflect the disadvantages of providing everything by oneself:

- Infrastructure:
  - Acquirements;
  - Timely expansion;
  - Architecting the design.
  - Possible under-utilization;
- Having in-house knowledge and specialists;
- Managing the whole;
- Cooling;
- Energy-provisions;
- Physical security.however

A consequence is that these problems will be handled with much more effiency caused by their specialization. A comparison could also be made between a cloud provider and an employment agency. Whilst the agency provides the manpower with the necessary skills, the cloud provider will provide the services with an enormous flexibility. It will not matter if one needs this service temporarily or indefinitely. This means that one will not have to wonder about what has to be done with the infrastructure, when he no longer needs it services.

The following is an example of a cloud's flexibility and scalability benefits. Imagine an emerging web-enterprise, which owns its own infrastructure and must be able to predict its growing process. But if the infrastructure is too small, the requested services can possibly not be fulfilled completely and clientele can be lost. When the infrastructure is oversized, it is possible that the enterprise cannot overcome its initial costs. If the enterprise relies on a cloud provider for delivering those services, it would not have to worry about these problems because the necessary infrastructure will be provided for. Like this, it could start out small and scale dynamically to its needs. The example was set in a commercial context, as most services delivered in the cloud today. UNINETT does not have a commercial interest, so in the continuation of this report will no longer be focussed on the commercial context.

5.2 Services over the cloud

Next, a brief overview of the different services which are offered in the cloud will be given:

SaaS or Software as a Service offers an application over the internet and nothing more. Examples are: the social networking service Facebook[13], Google Apps[14] which is a web application provider from Google, etc.

PaaS or Platform as a Service is a continuation of the former service. It provides a framework that provides applications and where applications can be built on. Most often, software design tools are incorporated to facilitate the programming and interactions of applications in the framework. An example is the Dutch company GravityZoo[15], but also many others.

IaaS or Infrastructure as a Service extends the latter even more. IaaS offers a complete virtual infrastructure where it is possible to set up a server to a network of multiple virtual images. Where some provide solely their own images, others give the possibility to set up one's own XEN, VMWare and/or other images in the Cloud. This infrastructure can be used for all kind of setups, ranging from server or desktop replacement to setting up a complete cluster. A popular well-known IaaS provider is Amazon with its EC2 services, but there also are quite a few interesting products which offer one's own IaaS software. This means that they are not a cloud providor themselves, but they offer instead a hardware or software product which allows one to set up its own cloud infrastructure. A few interesting examples to mention are Open Source products like OpenNebula[16], Nimbus[17] and Eucalyptus[18].

STaaS or Storage as a Service offers a scalable storage solution and guarantees that no data will be lost, etc. One of the major players is Amazon with its S3 service and some interesting products to set one up are ParaScale[19], Atmos[20] and the open source variant CloudStore[21].

5.3 The cloud concept: definition

The cloud might be a buzz word for all of these services, but it does have general aspects which define it as a"cloud":

- The service must be available 24/7, independently from the device or from the access location, in a secure fashion and as long as one's internet connection is working.
- The services must be reliable and robust to guarantee business continuity.
- The service is flexible to one's needs. Sudden requests for more resource utilization may not decrease the overall service performance.
- The resources used for the service may scale on an unprescribed rate, without increasing the cost per resource ratio.
- The costs of the resources in the cloud are calculated on one's use of it. For instance, if an VM (Virtual Machine) in the cloud is turned off, costs for computing time will not be charged.

How these aspects are provided depends on the provider. In the following section is the following described: The possible architectural constructions, my opinion on which architecture fits the cloud's concept best, and the challenges which are faced to come to an interoperability between different cloud architectures.

6 Architecture in the cloud

The architecture of the cloud is very important to be able to guarantee its main features which were discussed in the previous section. Depending on the cloud provider's choice of service, the architecture may differ. Although, the following features are common in every architecture in the cloud.

6.1 Common features of a cloud architecture

6.1.1 Device and location independent access

It is important to offer access to the cloud services with platform independant protocols, which are able to be used with the latest security protocols. For example: (S)NFS, WebDAV, HTTP(S) and FTP(S). This must provide accessability from computers to PDAs and whilst still providing the same secure connection.

6.1.2 Control of access

Clients are concerned about losing control over certain sensitive data. Providers typically log access and some providers let their clients view these logs. This should not be too difficult through the use of a few scripts, but it should be implemented in the providers policy.

6.1.3 Efficient cost per resource

This can be obtained through the following points:[22]

- Multi-tenancy enables sharing of resources and costs among a large pool of users, allowing for:
  - Peak-load capacity increases dynamically.
  - Use and efficiency improvements for systems that are often used for only 10-20%.
- Customers minimize capital expenditure; this lowers barriers to entry, as infrastructure is owned by the provider and does not need to be purchased for one-time or infrequent intensive computing tasks.

Aside from the sharing of equipment between multiple clients, providers can also lower their costs by setting up their infrastructure based on commodity hardware. The feasibility of this concept can be read in section 6.5.1. Another cost-saving aspect of commodity hardware is its manageability, which will be described further in sections 6.3 and 6.4.1.

6.1.4 Increased robustness/reliability

Reliability improves significantly through the use of multiple redundant sites. This guarantees business continuity, as well as disaster recovery. Whilst the cloud provider monitors performance to keep it consistent, there is always the danger of insufficient bandwidth or high network load. Commercial cloud providers already encountered these problems and have suffered outages in the past. The monitoring is a very important aspect of the cloud. If a node fails, this will easily be discovered and recovered with the self-healing capacity of the cloud. This capacity, in context of a computing cloud, means that all processes from a failing node will be restarted on another node; whilst in context of a storage cloud the data will be copied to another node, from one of the replica nodes.
When economies of scale are used to improve robustness instead of profit, reliability may even be easier to be upheld. Bandwidth could possibly be easier to guarantee when a dedicated network is in place. These goals are probably very hard to reach in a commercial context, but when thinking in the context of an NREN, it is not that difficult anymore. Although, if network availability cannot be guaranteed by the provider, this should be mentioned specifically in the SLA (Which is done in the Amazon SLA for example). This essential point could make the difference for an organization to put its resources on the cloud or not, or only the parts which are not essential for business continuity.

Other features are specific to which services are ment to be provided in the cloud. In the following sections are the possible architectures for providing cloud storage and cloud computing discussed, and how they can deal with features as manageability and scalability.

6.2 Storage: Traditional architectures

Before looking at the most common architecture for cloud storage, namely a clustered commodity architecture based on the Google File System, I will present the more traditional architectures, such as a storage area network (SAN)[23, 24] and a network attached storage (NAS)[25]. Their main features will be discussed next, whilst regarding the data storage they are fit for, and why it is less fit as a cloud storage architecture.

6.2.1 Network Attached Storage

These storage servers are actually preconfigured file servers. It has its own internal storage, CPU's and all the other major components of a typical computer, but with a stripped-down or special operating system optimized for file-sharing. Accessibility is reached by using familiar networkprotocols.

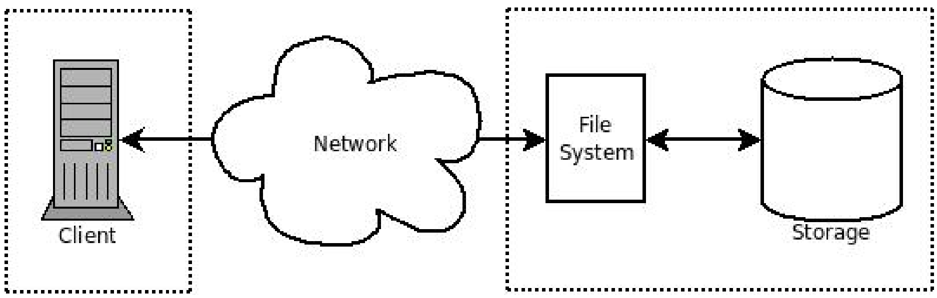

Figure 1: Network Attached Storage

Features:

Reliability is mainly in function of how well it is designed internally. As long as it does not have redundant components like data access paths, controllers and power supplies it is probably less reliable than a setup of redundant storage servers (like in a storage cluster for example). Adding redundant components in an NAS device can be costly however.

Performance depends mainly on its hardware setup. Whilst regular servers can be upgraded, an NAS has mostly a fixed hardware setup, which is also its drawback. If the NAS server gets too slow because of having too many users or too many I/O operations or CPU processing power which is too demanding, the whole box must be replaced and this is immediately a larger expense.
Besides the hardware configuration an NAS is also dependant of its network protocols. Overuse of the network can create a performance bottleneck when:

- computing disk usage of separate directories;
- rapidly indexing files;
- using I/O-intensive applications like:
  - Databases;
  - Video processing;
  - Batch processes;
  - Multimedia applications.

Scalability[26] and manageability is another problem. Plug&Play is implemented in most NAS systems which makes it easy to set up and maintain. But whilst a single NAS system is perfectly manageable, it has a certain limitation in its scalability. To scale up, it is posshoweverible to use multiple NAS systems together, but the management level increases exponentially. This, because it will add to the need of load balancing on each system and migrating data between systems. As a result, the costs per storage quantity will increase.

The network attached storage's specifics are actually very well for the following:

- Centralized storage:
  - NAS was developed especially for file-sharing. (It combines the storage aspect with a file system component and synchronizes accesses)
  - Sharing data is easy between different operating systems.
  - Can be used for centralized backup.
- Provide storage for load-balancing and faul-tolerant email- or web-servers, etc.;
- Storage for Home multimedia data.

The earlier mentioned problems about its limited scalability and its management problems when more storage is required, limits its possibilities. It might be a perfect solution when a limited storage space is needed, but it is unfit as a solution for the storage cloud.

6.2.2 Storage Area Network

SAN is an architecture which attaches different storage devices to servers in such a way that the devices appear as locally attached to the operating system. It provides block-based storage instead of a file-based storage and it uses faster protocols such as SCSI (Small Computer System Interface), Fibre Channel, iSCSI (SCSI over the over the Internet), ATA over Ethernet, or HyperSCSI (SCSI over ethernet). SAN equipment is relatively expensive, especially when used in combination with Fibre Channel. iSCSI is expected, however, to produce eventually cheaper SANs. Its higher costs and complexity make a SAN rather uncommon outside large enterprises.

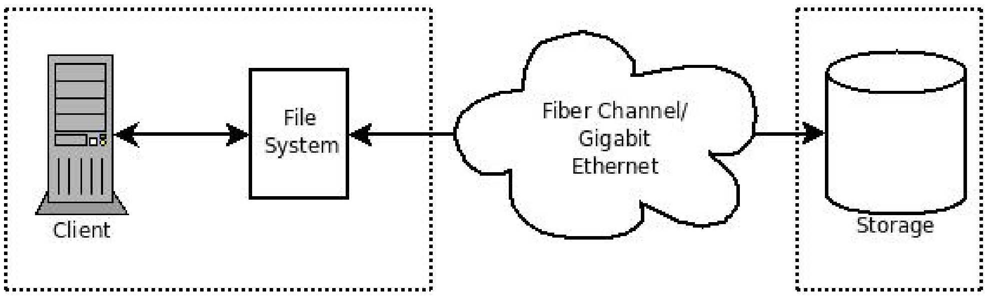

Figure 2: Storage Area Network

Features:

Reliability is ensured by use of multiple redundant components, which is another aspect that increases the costs of an SAN. Per-node bandwidth usage control can also be configured to ensure fair and prioritized bandwidth use across the network. This allows for insufficiently available bandwidth to be divided fairly.

Performance of storage area networks are mostly higher than those of network attached storage because:

- Faster protocols are used;
- Has possible I/O caching;
- Replaceability issues. The storage area exists of multiple components, which can be replaced or upgraded when it fits the needs. One is no longer locked to a fixed box of hardware. This advantage is lost of course, when using a SAN-in-a-box solution.
- Consists of block-based storage. It is easier to move a huge file, if it can be split up in different blocks and send block by block over the most efficient path through the network. The same is possible with

file-based storage, but the number of translation steps in between make it less performative.

Scalability and manageability might prove to be problematic. First of all, managing and setting up LUNs can be very laborious and the more it scales, the more LUNs are necessary and the harder it becomes to manage. Second, adding storage arrays are not as easy as plug and play.[27] Most of the time, it implies rigorous testing in a duplicate environment, before scheduling it carefully into the network.

Especially its higher performance makes it more useful for:

- Transactionally accessed data (high-speed block-level access to disk)
  - Email servers;
  - Databases;
  - High usage file servers.
- Media editing;
- Server-booting from SAN;
  - In case of failure: replacement server may use LUN of faulty server.

It is, however, not fit for sharing data between different operating systems. If more types of operating systems should write on the same LUN, this can cause compatibility issues and failures. It is possible to use LUN masking or define zones for different operating systems, but it is much easier having a storage solution with one file system like on a network attached storage device.
Whilst its purposes lie closer to high-performant storage and because of its higher costs, caused by the level of management and expensive equipment, I consider a storage area network not the best-fit for the storage cloud. Next, a special cluster of commodity hardware will be described, which is a better solution for a storage cloud.

6.3 Storage: Clustered commodity architecture

The Google File System is a proprietary cluster file system which is able to scale enormously, whilst staying very manageable. Its concepts will be presented, because most cloud storage architectures are based on its design principles. The following paragraphs describe: its features, its differences with a traditional cluster file system, and specific vendor implementations.

6.3.1 Google File System cluster

Its concepts are totally different then most current storage solutions. It is based on lots of commodity servers and as can be noticed in figure 3, its architecture exists of two main components: The master server and chunk servers.

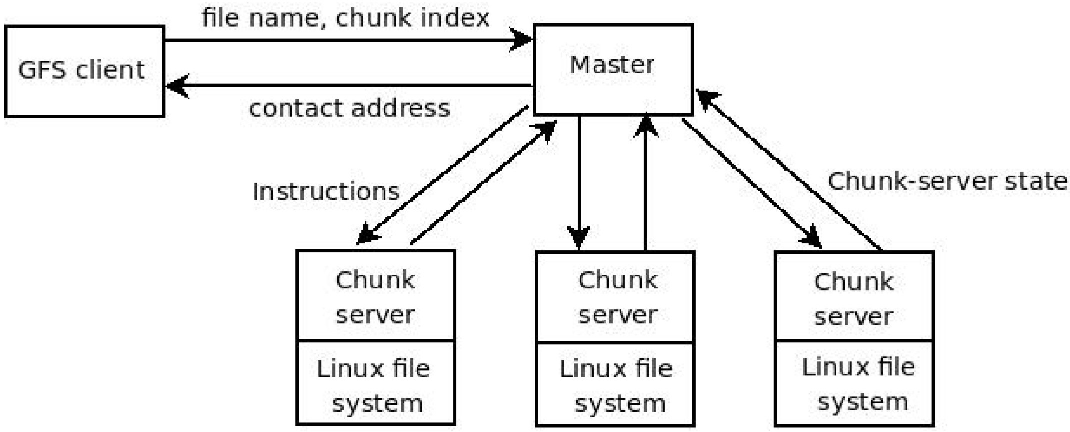

Figure 3: Google cluster architecture

Because of the scheme's simplicity, one master is able to control many hundred chunk servers. This size is unseen in common cluster-architectures and is caused mainly because of the following reasons:

1. The master's workload is kept to a minimum:
   - It maintains a namespace and mappings from file name to chunks;
   - It does not attempt to keep accurate account of chunk locations;
     =>For this reason a lower consistency level is set.
   - Contacts chunk servers occasionally to see which chunks they have stored;
     =>gets picture of where what data is stored.
   - Master is only contacted for metadata, by the client;
     - Client passes file name and chunck index to master and master replies with a chunk's contact addresses;
       =>Much chance at least one will be still working.
   - Master allocates chunks to chunk servers;
2. Bulk of the work is done by the chunk servers;
   - Client communicates with the chunk servers to fetch the data;
   - Server failures are the norm instead of the exception, so chunks are replicated to handle failures. (primary-backup scheme)

        - Client contacts the nearest chunk server holding the data and pushes updates to that server;
        - This server pushes the update to the next closest server holding the data, etc.;
        - Once all updates have been propagated:
            * Client contacts the primary chunk server;
            * Chunk server assign a sequence number to the update operation and passes it on to the backups.
              =>The master is kept out of the loop!

3. Hierarchical name space for files is implemented using a simple single-level table; path names are mapped to metadata.
    - Entire table is kept in main memory, along with the mapping of files to chunks. Updates on these data are logged to persistent storage;
    - If the log becomes too large, a checkpoint is made by which the main-memory is stored in such a way that it can be immediately mapped back into main memory;
    - =>I/O of GFS master is strongly reduced.

Specifics to how Google files are handled:

- They tend to be very large, ranging up to multiple gigabytes;
- These files each contain lots of smaller objects;
- Each file is divided into chunks of 64 megabytes and distributed among the chunk servers;
- Chunk servers keep a record of what they have stored;
- Updates take mostly place by appending to files rather than overwriting parts of it;

Consequences of the implementation

Because Google's services are organised into smaller services that are mapped onto these clusters, it is understandable that they can handle all their loads of data so well. However, the architecture is built adjusted to their needs and is therefore not appropriate for all goals.
It is especially fit for handling files which are:

- Large;
- Write once, read many;

These files are especially proficient for:

- Backups;
- Media-files (Possibly with streaming-capacity);
- Archiving;
- Large scientific data which is written once and read one to many times;
- Web-data.

Difference with traditional clustered filesystems
The main difference between the GFS and other cluster file systems is their architecture, which has a huge impact on the level of scalability. Whilst a GFS cluster limits the internode and master-chunkserver communication, a general cluster filesystem needs to distribute metadata across the system. This increases file-access latency and high-cache-coherent-network-traffic, which limits the scalability and performance considerately.

Performance gain offered
Aside from the performance gain which is described in the former paragraph, it has also a performance advantage above most big-system solutions. This is because the systems are made up of multiple different nodes, each having their own CPU. Those can be used for parallel reads and writes, but also to guarantee a better distribution of content, because files may be replicated over more nodes when the demand increases, making good use of the multiple CPUs. When the system runs on performant hardware (more RAM, SAS drives, multicore CPUs), it is even capable of streaming media and keeping up with the demands if they fluctuate.

Manageability
Whilst a SAN and NAS becomes harder to manage when it grows, the GFS architecture does not have this problem. All one has to do is install the specific operating system or software on the hardware of preferences, plug it into the LAN and the node will be discovered automatically. If the cloud exists out of one or fivehundred nodes, it makes no difference on a management-level, because of its self-monitoring and self-healing capacity. Self-monitoring means that the master node will automatically recognize a new node which is added to the cluster, but also when the number of copies of a data item droppes below the demanded number. This will cause the self-healing to take place, which will automatically replicate the data item to (an)other node(s).

6.3.2 Implementations by different vendors

The basic architecture has remained approximate the same for all vendors using this architecture, aside from some added features to provide a more specific solution/service. The following paragraphs will present some vendor specifics from Amazon, whom provides cloud storage, and ParaScale and EMC, whom provide the software(/hardware) to build one's storage cloud.

Handling the namespace

The namespace is handled mostly on the same way by each system, but they name the instances differently. For instance:

Amazon S3 uses buckets. This bucket can be seen as a single namespace, which has to be named uniquely. Inside this namespace, the bucket consists of unique objects. These objects, however, do not need to be unique to other bucket's objects.

ParaScale does almost the same. They layer an object file system on top of many Linux servers to provide a single namespace. Multiple namespaces can be used however and as such resembles to Amazon S3's bucket-system.

Atmos from EMC, has the same use of objects inside each namespace, but not much more information is available upon its workings or architecture.

Extra features

How Amazon implements their data-storage is more of a secret and little is known about their architecture. They are a cloud provider themselves, so although it might be interesting to know how they implement their service, the specifics are less interesting to their customers.

ParaScale and Atmos provide the software(/hardware) to build one's own cloud, so they provide more information on which features their products offer:

- Object-level deduplication;
- Extra compression methods;
- Replication policy:
  - Number of copies;
  - Geographic location;

  - ParaScale: Per virtual FS (per user or group(s)).

- Use of NFS;
  - ParaScale specifies its NFS version. NFSv3: file-locking, but last write wins.

ParaScale is for the moment in a beta-phase and as such many of the mentioned features still need to be implemented in their second release around August-September or in a later release.

What they offer

Amazon S3 offers its cloud storage solution over the internet, which one pays monthly per gigabyte. Other vendors such as ParaScale and Atmos are not providers themselves, but they offer the means to build one's own storage cloud.

ParaScale offers a software only solution, which can be used on the hardware of one's preferences. Their pricing is a based on a one-time license for a certain capacity of raw storage and on a support-contract which consists of a 25-35% list-price.

Atmos offers a software solution in combination with their hardware appliances. The choice of hardware depends on one's needs of performance versus the needed amount of capacity. This has the advantage of not having to worry about which hardware to buy in combination to the software, but it also increases one's dependency on a vendor's hardware.

6.3.3 Cloud storage: Conclusion

Clustered commodity servers, based on the GFS, proves to be the best architecture to provide cloud storage. Because a single master server can manage many hundreds of nodes, it is an extremely manageable and scalable architecture. One of its essential components is the use of commodity servers. The feasibility and management aspects of commodity hardware will be discussed in section 6.5.

6.4 Cloud computing architecture

6.4.1 High performance computing

In this section the different options of an organization to implement high performance computing (HPC) will be discussed. These options are the use of supercomputers, clusters or grids, and now also through subscribing to a cloud computing service. The following paragraphs will provide an overview of the different options and closure with how and when a cloud computing service may prove to be a better solution.

Supercomputers

In the late 1950s the first supercomputers were constructed for military use. This enabled processing of data with an astonishing speed, considering the period of time. In the following decades, the advancements of supercomputers increased exponentially, but so did the costs. With concepts such as Moore's law and economies of scale taking ground, the HPC capabilities boomed even more as before. Nowadays, a single desktop computer is more powerful than a ten-year old supercomputer and this has helped in redesigning the basic supercomputing architecture enormously. This architecture is now made up of a cluster of multiprocessors, which are highly interconnected by a high-speed network and share tasks using symmetric multiprocessing.

Cluster Computing

Now, HPC can also be done with commodity hardware to save expenses and this whilst withholding the need of buying an expensive supercomputer. These computers still have to be very densely coupled over a high-speed LAN. This is needed for the frequent communication between nodes, to be able to work coherently together. Due to this, a cluster of computers can be programmed and used as one large computer. Because of the high-use of bandwidth of inter-node communication, there is a limit to a cluster's scalability. Another point is that the cluster should be build up out of homogeneous machines making load balancing inside the cluster a lot easier, but this is not mandatory. Another cost-saving aspect on behalf of a cluster is the used operating system. When most supercomputers have their own specific, and as a result expensiver, OS, cluster computers are mostly outfitted with a UNIX or LINUX OS. Whilst it is a lot cheaper, it has also its disadvantages. For instance, a cluster is mostly set up by a scientist for a specific task and is afterwards broken down, to be set up again when the need arises. This can however be a very laborious task, which takes some time to complete.
In the 1980s-1990s attention turned to parallel processing. Instead of working together on one task, the task is split up into multiple parts, which can then be processed in parallel and concurrently.

Grid Computing

The next step in parallelization is grid computing. Whilst clusters need to be highly-coupled for internode-communication, the tasks for which grid computing is used consists mainly of independent jobs. This means that no internode-communication is necessary and each node can handle their own tasks. One of the main advantages of this setup, is that the grid may be spread over the internet and may exist out of heterogeneous machines. This has lead to some very nice implementations, using the spare CPU-power of home computers to help in finding cures for cancer or Alzheimer or even help in analysing radio telescope data.

Cloud Computing

One may subscribe to a cloud computing provider for HPC tasks, when one has:

- No dedicated hardware at his disposal;
- The need for more HPC resources as an addition to its own.

Important to consider is that HPC in a cloud is less qualified for some high-dense computing tasks, because everything in the cloud is virtualized. Examples of these tasks are:

- Weather calculations;
- Rendering;
- Data scrubbing;
- Image transformations.

It is, however, a perfect solution for parallel-computing tasks which were previously done on grids or clusters. The CERN project, where an enormous amount of physics-data needs to be processed in parallel, is a perfect use case for this purpose.
The virtualized cloud computing environment has more advantages though. It is constructed to be self-monitoring and self-healing, which takes away a lot of the management difficulties as seen with clusters. If one node should fail, it will automatically be repaired or replaced by another if needed. Another great advantage is the ease of setting a cluster up inside a cloud computing environment. Once an image has been configured, it is not more difficult to set up 100 of virtual nodes instead of a few. These images can also be re-used or shared, so the laborious task of setting it up can be a thing of the past. As a consequence, it saves a lot of time and as such a lot of costs. If an organization would be determined to hold on to their own infrastructure, the cloud could still be of an benefit. Imagine instead of having to acquire an infrastructure for huge loads of computations, whilst

50% might only be used for 5-10% of the time, acquiring HPC resources from the cloud for handling computation-peaks.

Concluding, cloud computing may be used to provide HPC as an addition or in substitution to an organization's own HPC resources, and its ability to set up a cluster or grid in a few clicks makes it a good facility for parallel-computing tasks.

6.4.2 Virtual machine servers

A second cloud computing service to discuss is providing virtual machine(VM) servers. This can actually be seen as a component of HPC, where one VM server is a single node. Likewise, if a node fails, its VMs will recover at another node.
Although, this section will concentrate on how these VMs should be provided. Should big server-boxes be acquired which can hold 1000s of VM's or would it be more beneficial to acquire multiple commodity servers to do the same? Normally those huge servers were said to be more reliable, but this seemes to have changed over the years. An example of this is how SalesForce[29], a SaaS provider, changed their infrastructure completely:[30]

> They used to run their services on big SUN-servers, that could allow a thousand clients to run their software. This gives one problem though: when one server went down, those same thousand clients' software stopped running for multiple hours.
> SalesForce has replaced now all its SUN servers with DELL commodity servers.

Of course, it has not officially been confirmed if it was as a direct result of the failure or not, but many cloud-service providers seem to be following the same course. Still, I can deduce the following:

- The cloud's architecture provides reliability by auto-monitoring and healing the VMs;
- Regarding the infrastructure, a higher price setting does not necessarily ensure the same increase in reliability.

Concluding, whilst expensive high-end servers provide in a certain degree more reliability than commodity servers, the cloud copes with this by an architecture which provides reliability by itself. Because of this, commodity servers may prove to be financially more interesting, without losing reliability in the offered services.

6.5 Commodity hardware in the cloud

6.5.1 Feasibility in the storage cloud

In this section is a rough outline made of the costs to estimate the feasibility of commodity hardware in a storage cloud. The costs of 100TB of data stored at Amazon's S3 for a period of three years are compared to the costs of commodity hardware and to the approximate costs of the same quantity stored at NorStore. For NorStore and commodity hardware, an estimate is used for the energy and cooling costs of 100TB from an online article[28]. With regards to the completeness of the calculations must be mention that the following matters were left out:

- Amazon's data transfer out costs;
  - Only transfer of data into the cloud is included in the costs;
- Cloud software costs (in case a commercial product is being used);
- Installation, placement and management costs (For commodity hardware as well as for NorStore);
- Economy of scale is not taken into consideration for buying lots of commodity hardware.

Calculations:

Amazon S3

```
price:  0,18 Dollar/GB/Month (first 50TB) /
        0,17 Dollar/GB/Month (next 50TB)
(0,18 Euro x 500 000 + 0,17 Dollar x 50 000) x 12 x 3
+ 0,1 Dollar x 100 000 (transfer in costs)
------------------------------------------
640 000 Dollar (for 100TB NET)
```

= +/- 500 000 Euro

Cloud consisting of commodity hardware:

```
NORCO DS-1500 15-bay 4U Hot-swap Rackmount eSATA RAID Hard Drive
Storage Subsystem
=~ 900 Euro
Seagate 1.5TB Barracuda 7200RPM SATA-300 32MB
=~ 140 Euro
x 15
= 2100 Euro + 900 Euro = 3000 Euro
6000 Euro (Approximate costs for 2 servers (CPU power + 2 HD of 1,5TB))
------------------------------------------
```

```
9000 Euro (for 28,5 TB GROSS)
x 4
+ ~(36400 Euro x 3 year) (Energy and cooling costs)
------------------------------------------
```

= +/- 145200 Euro (for 114 TB GROSS)

NorStore's infrastructure has for the moment around 600TB NET at the price of 200000 Euro.

```
=> 100TB NET = 33333 Euro
+ ~(36400 Euro x 3 year) (Energy and cooling costs)
------------------------------------------
```

= +/- 142533 Euro

The comparison cannot be left entirely like this. First of all, it is worth considering that NorStore's hardware price has been negotiated in two to three months, whilst the commodity hardware prices are based on internet prices. These do not have the huge discounts when bought in mass, considering the economy of scale factor. Second, the commodity hardware needs possibly more storage place to reach 100TB NET. However, cloud software vendors promise compression and deduplication with their products, so the estimats will doubtfully differentiate much.

From the previous reasonings and calculations, I deduce the following. First, providing a cloud oneself instead of outsourcing it to a commercial provider like Amazon proves to be much cheaper and especially if it has to be persistent storage for many years. Second, considering that the price of commodity hardware will be lower because of the mentioned reasons, commodity hardware will more than probably have a feasible price setting compared to the current NorStore's infrastructure.

6.5.2 Managing aspects

A few matters need to be taken in consideration when opting to implement, for instance, a few thousand of commodity nodes.

- How to handle this physically?
- What to do with broken nodes?

Because this is a huge mass to handle, the phyical management can not be handled on the traditional manner. This would waste too many man-hours, if every broken node should be replaced instantly. Instead, another way of managing such numbers of nodes has been introduced by companies as Google, etc. What they do is, leave broken nodes in the racks until a certain percentage is broken. At this point, a team can start to replace or repair nodes on the spot, whilst someone monitors the other nodes to make sure that no unbroken nodes are disconnected in the process. This does have the disadvantage of loosing place to store broken nodes. When handling such a huge mass of nodes and racks it is, however, more advantageous considering the time and money which would be lost when replacing each broken node instantly.

6.6 Interoperability between cloud architectures

This section discusses one of the major challenges, which still needs to be overcome before a cloud of clouds can be achieved. Namely how clouds of different providers can cooperate and buy resources from each other to server their own clientele. In cloud computing, there are already a lot of promising steps towards the goal, but cloud storage solutions still have the issue of vendor lock-ins. Because there are no open standards defined, interoperability between the different storage clouds do not exist yet. Whilst it would be possible to agree using one and the same cloud software among a group of providers, it is more interesting when each provider can chose their cloud software of preference and still be able to collaborate.
Although, a project is on its way about making an interface to interconnect multiple storage clouds. Its status is unknown to me, but I am familiar with one of their issues. This was where to place the interface, at the client's side or at the cloud's site. I would definitely say at the cloud's site, because the client's site should be unaware of the different cloud architectures. Instead, it should be handled by the cloud provider, but more of this will be discussed in section 7.5 on a storage exchange.

The interoperability for cloud computing is well expressed in the Reservoir project. The continuation of this section will describe the project, to give a general idea of how a cooperation of clouds can work out.
The Reservoir project is funded by the European Union and will deliver a definition of an architecture based on open standards that can serve as a cloud computing infrastructure without barriers. Without barriers because it aims for a federation of clouds. A federation where computing can dynamically be outsourced to other clouds, when one has insufficient resources by itself. Reservoir's core is based upon OpenNebula, which has interfaces to collaborate with Amazon's EC2, cloud technologies such as Eucalyptus and Globus Nimbus and probably even more towards the future if new cloud systems emerge. It also provides an integral management of one's virtual services, including networking and image management and it can be integrated with a lease-manager such as Haizea[31]. In figure 4, you can see Reservoir's vision on a Service Oriented Infrastructure, which is composed of[32]:

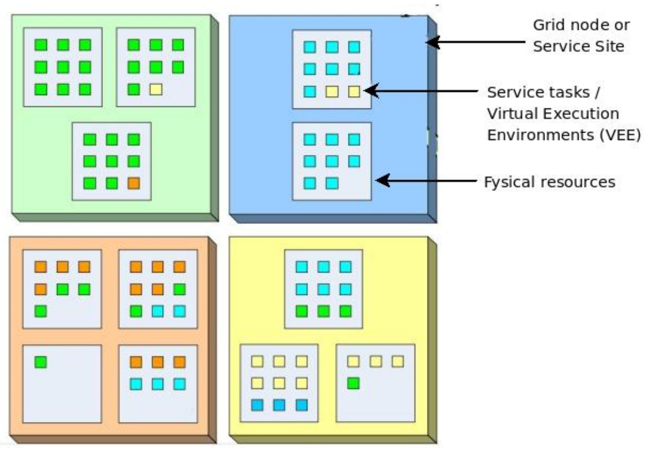

Figure 4: Service Oriented Infrastructure

Virtualization-Aware Grid - e.g., VM usage/size as the unit for metering and billing;

Grid-Aware Virtualization - e.g., live migration across administrative domains;

Business Service Management - e.g., policy-based management or service-level agreement.

and is based on the following policies:

- If possible keep Virtual Execution Environments (VEEs) from the same organization in the same physical box;
- Turn off underutilized physical boxes;
- If possible keep VEEs in “owning” organization;
- If possible keep VEEs in least number of external organizations;
- “Follow” the customer =>“Content Delivery System”

7 NREN as a cloud provider

Until now, I reflected on what the cloud concept actually is and which architecture fits best to provide its services. In this section, I will use these ideas to see where an NREN can use the cloud for and more specifically, where UNINETT can use the cloud for.

7.1 NREN: The pros as a cloud provider

By means of the following NREN characteristics is pointed out why an NREN is so well suited for a role as a cloud provider:

1. Its aim is not for profit, but to provide its services as best as possible to its constituency;
2. It does not have a traditional client/supplier structure, but it works together with its constituency;
3. It is a dedicated internet service provider, with the following network specifics:
    - Flat rate;
    - High bandwidth;
    - Low-latency;
    - Extremely reliable.
4. It has geographically spread locations.
5. Its network is connected to:
    - The PAN-European Network[33];
    - The GÉANT2 network[34];

Points 1 and 2 have as consequence that an NREN will not try to provide the least possible of services for as much money as possible, to their the constituency. Instead, they will know that what they pay will be spend to increase the cloud's robustness and quality. It also ensures one of the most important factors in the cloud concept, which is trust in their provider. This is an essential factor when trusting its resources to a provider's care.

Points 3 and 4 reflect the extreme favourable position of an NREN towards constructing a cloud. Because the network is already in place and considering its geographical locations, an NREN already has all the factors in place to build a nicely spread cloud. In contrast with other providers whom may only have two datacenters, which is then also called "geographically spread".

Point 5 reflects the ideal position of NRENs to form a federation of clouds. Because with all the same network specifics as mentioned in point 3, this federation could spread the cloud over even more geographical locations and be able to handle the fluxuating requests of storage more easily, as it becomes more robust. These interconnectability prospects are perfectly featured in figure 5, but will be handled more in section 7.5.

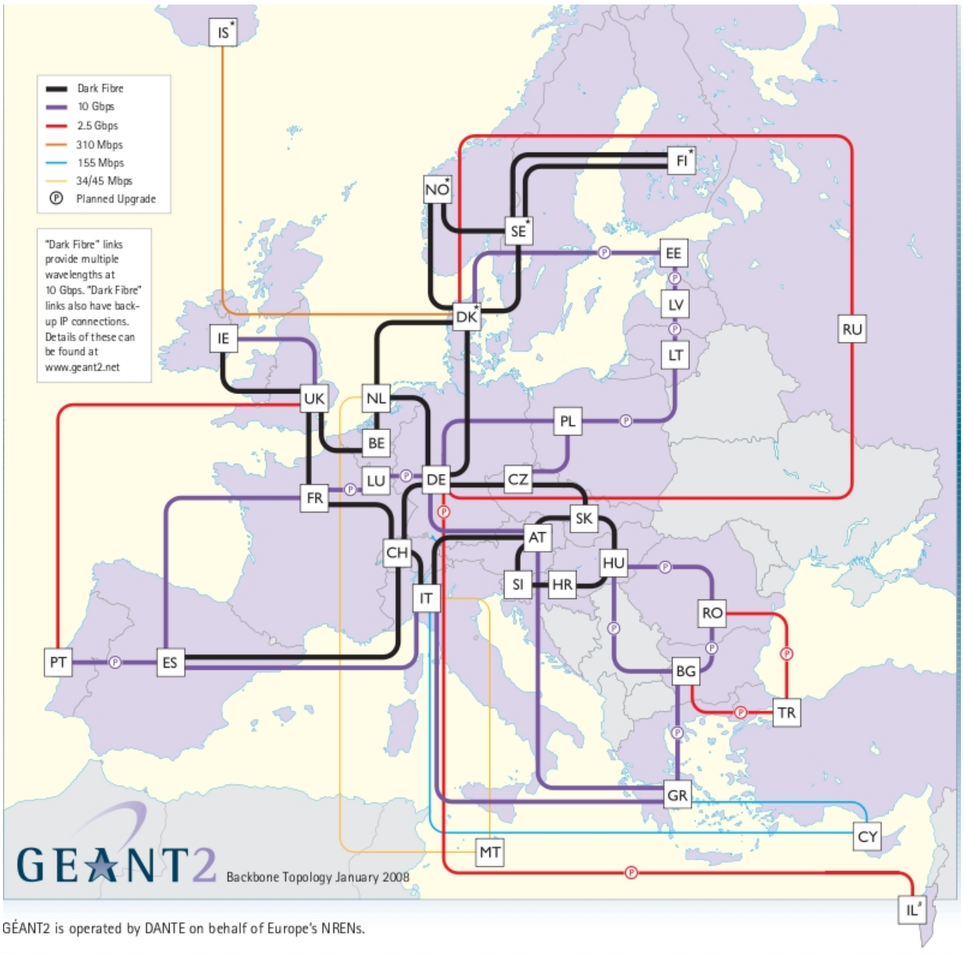

Figure 5: GÉANT2

Sections 6.3 and 6.4.1 elaborated on the possible implementations of cloud storage and cloud computing. The following two sections will elaborate on its applicability as Storage as a Service and Infrastructure as a Service for the higher education and research community.

7.2 Storage as a Service

The possible uses of STaaS for the higher education and research community can be looked at from two sides. First of all, if it can be used to store regular resources as is done in a datacenter and secondly if it can be used as an addition to a national grid such as NorStore.

Regular resources in a regular datacenter.

This can be perceived to be similar as a virtual (private) datacenter. Lots of organizations tend to these implementations themselves, but if storage keeps growing, they will be forced to continue doing this on the same price-level by preference. Working data should preferentially not be committed to the cloud, but there exists lots of data which is perfectly suited to be storaged on the cloud. These include:

- Back-ups;
- Data archiving;
- Peaks in a client's storage-needs.

Organizations will still be able to chose to which extend they would trust their data to the cloud. Do they wish to store their resources completely on the cloud, relying on the redundancy of the cloud and the speed of the connecting network, or do they wish to use it as an off-site backup and as a disaster recovery solution.
As mentioned in section 6.3, the cloud is perfect for storing and streaming, if necessary, film material. And why is this important? Because certain schools have a media department, which have a growing need of storage capacity. Whilst the storage cloud will not be fit for adapting the films on the cloud, it will be for the persistent storage of the films, once they are finished. Thereupon, they will have to be stored for multiple years, whilst still being accessible to be viewed.

As an addition to a national grid such as NorStore.

Research communities have the need to store their in-between and end-results in a persistent storage. When looking to Norway's research communities, they can submit an application to use a certain amount of storage on NorStore. This storage will then be funded by the Norwegian Research Council. As can be seen in section 4.2, there is still quite some storage-space left in NorStore, but future expansion will eventually be necessary. It is interesting to wonder if the cloud concept would be fit as part of this expansion.
NorStore's data exists out of many different types of data. Some must be able to be changed quite often, but there is also much of "write once, read

many" data and this is ideal for the cloud. Why would they prefer using an NREN's cloud then? Because they have credibility towards the Norwegian Research Council and when they have the infrastructure, they would be in the perfect place for NorStore's storage to be outsourced to.

Because it can be such a cheap solution on a very scalable manner, it can be a perfect solution for many of these previous examples. Once Norway's part of CERN data increases, which is also "write once, read many data", the cloud could even serve perfectly as its storage facility. If the cloud capacity would be temporary too low to handle all the data, this could be solved by having multiple clouds work together. But more on this in section 7.5. In the following paragraph will be discussed what must be considered when reflecting on how these services should be payed for.

When providing storage for a constituency, not part of NorStore, there are some parameters which should be taken in account:

- Which degree of replication is demanded;
  - e.g., possibly only one copy if it is combined with their own private cloud for redundancy and failure-protection.
- Performance demands;
  - e.g., To stream media material, higher-end equipment is needed.

When providing storage for NorStore, there is another point to mention. The project which requests storage, gets only a certain amound allocated on NorStore, which is then paid by the Norwegian Research Council. If NorStore outsources this storage to UNINETT's Storage Cloud, the payment would also be done by the same council. However, what if the projects funds for capacity has run out and they demand more? Should UNINETT wait for more funds to become available by the Research Council and put everything on hold, or should they allocate extra storage in advance? In a way, the pay-for-what-you-use concept is very clear on the matter. One needs to pay or he does not get more storage allocated. The extra payment can be done by the Research Council, or paid in advance by the project-group. A way of sorting things out however, is having it clearly formulated in the SLA. A certain amount of time for extra storage capacity could be added, which can be given in advance. At least an agreement is necessary between the project-group, the Research Council and UNINETT on how to handle extra storage allocation for every specific project. Even better would be, if the original agreement between the project-group and the Research Council provides extra funds when additional storage is needed. In the same line, should the following question be answered:

What must be done with resources of a client who does not pay?

Possible reactions are:

- Immediately remove it from the cloud;
- Data must be transferred to a client's site;
- Keep it on the cloud for a (un)limited period?

These different approaches should be well considered and as mentioned, well documented in the SLA.

7.3 Infrastructure as a Service

This section will describe the assets of cloud computing with regard to schools and research communities. Until now, when schools wanted to perform HPC, they used their own equipment. This has always been quite expensive, but the costs are getting lower. For the moment it is on a price/performance level that schools do not have to dig too deep into their budget to acquire the equipment, but I am certain that this is not the way to go for a certain number of reasons. First of all, HPC equipment is mostly used for maximum 10% of the time. If the equipment would be shared, it would be possible to do more for even less money.
One of the objections could be that one wants to be in total control of HPC environment. Whilst this is not possible in the cloud, for the hardware on itself, it is possible for what they run on it. Because virtualization permits this all. All they will be sharing is the equipment, not their work, nor their environment which they work on. Except when chosen otherwise of course, but the choice is theirs.
The ease of setting it up is also a big advantage. This counts for schools and possibly even more for researchers. When the latter have need of lots of computing resources, they mostly set up a temporary Cluster or something alike. Though this takes some time and when it would be possible to make only one image and build the cluster in a few clicks, this would improve the manageability and ease of use enormously! Without thinking of HPC, IaaS also gives the possibility to set up virtual networks, to run tests in environments which are otherwise not available and even to offer a webservice over the cloud. All one needs to do, is set up and configure the environment and the providor will ensure that everything keeps running.
Schools with a department of system and network administration, webdevelopment, or HPC would benefit from these possibilities.

Another possible asset would be for the NOTUR project. This project provides for instance computational resources and services to individuals or groups involved in education and research at Norwegian universities and colleges and is composed of multiple clusters, a distributed symmetric multiprocessing system and a massive parallel processing system. When the need would arise for more capacity, the cloud could offer the same benefits as stated previously to cluster or grid workings. This could be done by reserving a fixed part of the computing cloud or on a need-to-use basis.

7.4 Combining cloud-services

This section will discuss what needs to be considered when UNINETT opts for offering a combination of STaaS and IaaS. A scheme of how to combine both services would be necessary. If the data from the Storage Cloud is to be processed by the Computing Cloud, it must be known which resources have to be moved where. This might be moving the data to the computing cloud or the computing resources to the storage cloud. There are a few things to consider when making a choice on this matter:

- The data can consist of multiple megabytes to multiple terabytes;
- UNINETT has a superior WAN at its disposal for data-transfers;
- The storage cloud is not fit for making a lot of alterations to the data;
- To process the data in the computing cloud, it must be able to contain the working-data;
  - What if a working-node fails and its data is lost?
  - Where to store the in-between and end-results?

These considerations make me believe that more than one solution should be offered. One part should always be the same though; the computing and storage part should be as closely together as possible, especially if large parts of data should be processed.

Whether the data should be processed from on the storage cloud or first transferred to the computing cloud's storage, depends of how it must be processed. This is actually important for performance and consistency reasons. The storage cloud is not fit for multiple writes on the same data. When this is necessary, I would recommend moving the data to the computing cloud for processing and moving the end-results back to the storage cloud. If the computing nodes should access and process the data in parallel, it would be possible to move the data first as close as possible to the a computing facility within the cloud and process the data directly from on the storage cloud. See figure 6. This final example works approximately as a content distribution network (CDN). Where the content is moved as closely as possible to the requesting location.

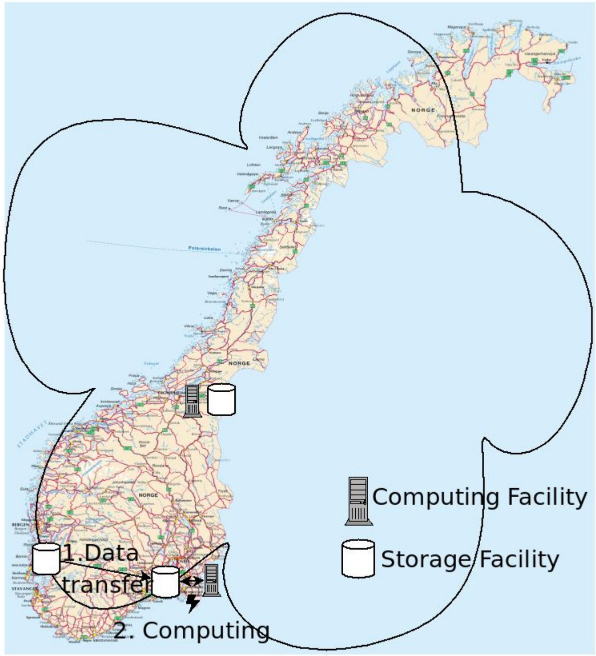

Figure 6: Storage movement in the cloud

7.5 Storage exchange

This section includes an invaluable concept of the cloud. A resource exchange in the cloud, is how I see that the future should be to obtain more out of the cloud. As is done now on the energy market, where electricity flows from one country to another to counteract deficiencies, should it be done in the cloud, but with resources. To simplify the concept, only storage resources will be taken into account. The general idea is that there would be a select group of trustworthy providers, which would interconnect their clouds to be able to handle the fluctuating requests of resources more easily. One provider, NorStore for instance, could then lend or sell a certain amount of storage to another provider, for credits or money (See figure 7), to counteract deficiencies in storage.

It would be possible to buy storage from another provider on a contract-basis, but my idea is a more dynamical approach. I suggest the creation of an international storage pool, which would:

- Keep account of each provider's storage provision;
- Keep account of how much storage each provider wants to keep in reserve for its own purposes;

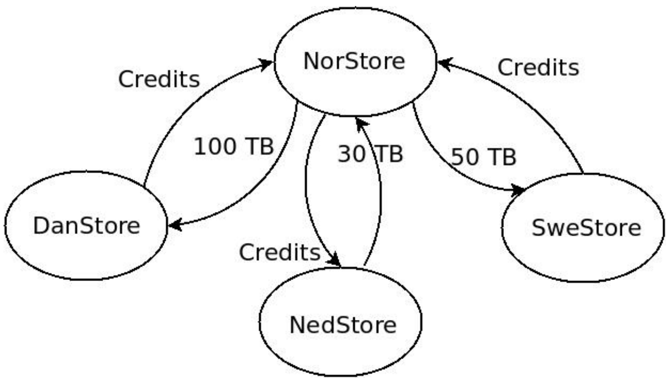

Figure 7: Storage exchange

- Redirect storage placements to other providers according to:
  - Costs;
  - Distance;
  - Bandwidth;
  - Storage availability;
  - SLA.
- Re-arrange storage placements, to keep storage with its own provider as much as possible;
- Could act as a CDN if data allows it; (Research content, . . .)
- Possible replication at multiple locations for disaster recovery;
- Arrange billing/payment among providers.

In the end, the client who requests storage does this with his provider and will not care (in certain limits) where it would be storaged. As long as his requirements are always fulfilled:

- Accessibility;
- Reliability;
- SLA requests (within country borders if requested, . . .)
- Security;
- . . .

In figure 8 the main lines are sketched of how the concept should work.

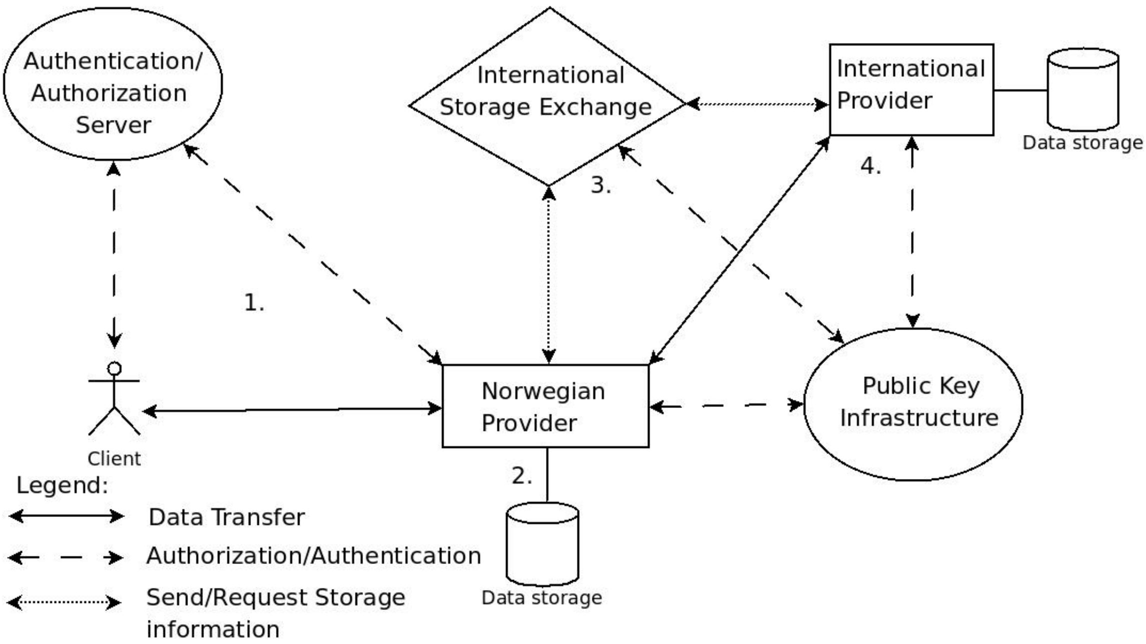

Figure 8: Storage exchange

The example represents a Norwegian client who contacts its provider to place data on the cloud.

1. Client contacts its provider, authenticates and asks a certain amount of storage;

2. Norwegian provider checks the available storage and assigns it if sufficient is available;

3. If not, the Norwegian provider contacts the storage pool and

    - Authenticates through, for example PKI;
    - States his requests (Storage needs, SLA-demands, ...);
    - Gets redirected to an appropriate provider.

4. Norwegian provider contacts the appointed provider and

    - Authenticates through, for example PKI;
    - Redirect its clients data;
    - Stores where to retrieve it.

This architecture would allow excellent scalability and ease of management. In the mean time a client will only contact its own provider, which keeps the information of his clients where their data is stored. The “foreign” providers won’t have access to this personal information and if demanded, the data can always be encrypted for extra safekeeping.

Ideally, the providers who join the storage pool should be a collection of NREN’s, providing their fast interconnected networks to transfer data over the cloud and to the constituency. Their ideally interconnections over the GÉANT2 network are expressed in figure 5. But it does not have to stop here. It could as well be a commercial provider, who joins the storage pool. According to the demands set for the data, the best provider can be chosen from the pool. Normally when dealing with a commercial provider, there are additional costs such as data transfer. With certain providers, such as Amazon, this could be dropped when direct peering is an option and the amount of data to be stored reaches a certain amount. One of Amazon’s datacenters is at a very close proximity to an NREN’s datacenter in Ireland, which is directly connected to UNINETT over the GÉANT2 network. If the distance between this NREN and Amazon’s datacenter could be crossed with a dedicated link, direct peering would be possible. The rest of the data transfers will cross the PAN-European network, with no extra costs involved. Before considering a cloud of such proportion, a more limited approach could be considered. For example between the Nordic countries, whom already have a high degree of cooperation and are also highly interconnected as can be seen on figure 9.

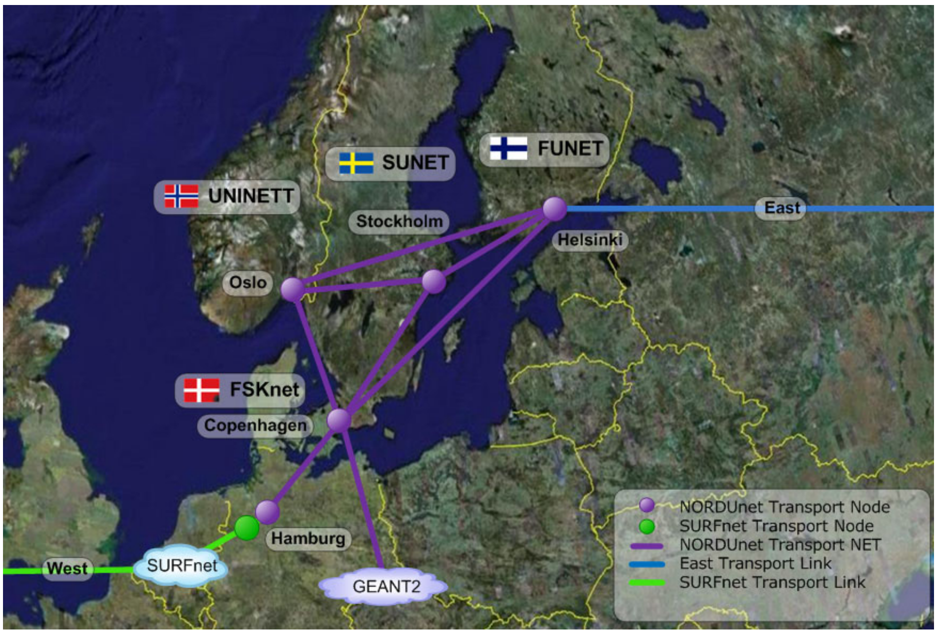

Figure 9: NOX

Although, before this future can become reality, the necessary interfaces still need to be developed, as was stated in section 6.6 and some legal questions need to be solved too. These questions will be reflected on in the following section.

8 Legal questions

Storage and infrastructure as a service are really nice concepts, but there are still quite a couple of legal questions which need te be answered. UNINETT or a legal counsellor should find an answer to these, before providing any cloud services.

- What is the provider required and allowed to do with the data?
- Would cloud storage comply to the law with regard to the privacy and protection of personal information?
  - May it be used to store personal information?
- May resources be store in any foreign country?
  - The US government could access stored information at will (US Patriot Act).
- Should the client be informed in which parts of the cloud his resource reside?

9 Future work

The cloud concept is not yet in its final stage, so there are still a few things on the to-do-list. One of them is getting some suspicion out of the way, for putting resources on the cloud.
As a starter, it would be best to set up a few test-cases. First of all, this will provide more in-house knowledge and secondly, it can be used to persuade the constituency in its ease of use. Because, no matter how well the prospects may be, a concept must be known and sound familiar to be trusted.
Then it is still the question: What to test? The Reservoir project (and related OpenNebula) is already one of the top 100 players in cloud computing and would certainly be good for a test case. According to me is its open source architecture and its promises towards interoperability exactly fit for the needs of an NREN's computing cloud. Such a test case could be used, for instance, to test the efficiency of virtual clusters and grids compared to their physical counterpart and more.
When looking for a test case of cloud storage, as well Atmos as ParaScale offer a good product, both included with an easy to manage interface. ParaScale, for instance, offers a free software demo for a maximum allocation of 4 TB data, but they are still in a beta-phase. Because of this beta-phase, it might be interesting to see if a demo version can be provided of the hardware-solution Atmos.
One of the other points of cloud storage is its interoperability, which is discussed at section 6.6. The problem mentioned comes down to the fact that there are no standard interfaces, which lets two storage clouds "of a different vendor" interconnect. This could be solved by creating an interface which would be placed as a middleware solution. Although, this problem could also be solved if different providers would implement similar cloud architectures.
As mentioned in section 7.5 about a storage exchange, discussions should be started with other NRENs, and this for two purposes. First, to see if they are interested in the cloud concept and if they are interested in being part of a federation of clouds. Secondly if they show interest, how this federation of clouds should be handled. Would they opt for a vendor specific storage cloud, which has no interoperability issues, or should an interconnection interface be constructed first.

10 Conclusion

UNINETT was interested in the cloud concept and how this could be used as a service for the higher education and the research community. I will first recapitulate the concept briefly.
The cloud is architectured such that its services are perceived by its constituency as:

- Available 24/7;
- Extremely reliable;
- Accessible over the internet, independently from the device or from the access location;
- Flexible in delivering resources to one's needs;
- Extremely scalable, without increasing the cost per resource ratio;
- Geographically dispersed.

This context is even more favourable for an NREN because it is already geographically spread, which provides the necessary locations, and already has an excellent network in place, which has:

- Flat rate;
- High bandwidth;
- Low-latency;
- Extreme reliability.

Now, what services can an NREN provide with the cloud concept? When looking at providing storage as a service, the architecture is especially fit for storing none-changeable content. This could be scientific data, web-content, films (from a media department for instance), etc. When commiting data to the cloud, this includes having it backed up, protected from local disasters and being duplicated for content delivery. Meaning with the latter, that data may be replicated over the network for a better distribution if its demand increases.
When looking to infrastructure as a service, the architecture can provide a single to many hundreds of virtual machines. These can be used for replacing servers in the organization, to constructing a huge grid of virtual machines. When thinking of the latter, this could save the client a lot of expenses because he will not have to pay for any underutilized resources.

As mentioned the cloud is also fit to provide a cloud of clouds, or a federation of cloud providers. An NREN is also extremely well suited for this because of its direct access, by the PAN-European and GÉANT2 network, to other NRENs. While it is not a necessity when providing cloud services, it does provide an extra safety measure for handling sudden peaks in resource-requests.

References

[1] UNINETT website
http://www.uninett.no/

[2] Norwegian storage infrastructure
http://www.norstore.no/

[3] Norwegian University of Science and Technology
http://www.ntnu.no/english/

[4] The Norwegian metacenter for computational science
http://www.notur.no/

[5] Amazon's Elastic Compute Cloud
http://aws.amazon.com/ec2/

[6] Amazon's Simple Storage Service
http://aws.amazon.com/s3/

[7] Google File System - Ghemawat et al.(2003)
http://labs.google.com/papers/gfs-sosp2003.pdf

[8] Wikipedia - List of Distributed file systems
http://en.wikipedia.org/wiki/List_of_file_systems#Distributed_file_systems

[9] About the UNINETT group
http://www.uninett.no/om.en.html
http://en.wikipedia.org/wiki/UNINETT

[10] Norwegian GRID Initiative
http://www.norgrid.no/

[11] The Large Hadron Collector
http://lhc.web.cern.ch/lhc/

[12] Nordic Data Grid Facility
http://www.ndgf.org/ndgfweb/home.html

[13] Facebook Website
www.facebook.com/

[14] Google Apps website
http://www.google.com/a

[15] Platform as a Service provider
http://www.gravityzoo.com/

[16] Open source IaaS software: core of the Reservoir project
`http://www.opennebula.org/`

[17] Open source IaaS software: special features for setting up clusters
`http://workspace.globus.org/`

[18] Open source IaaS software: developed at the University of California, Santa Barbara
`http://eucalyptus.cs.ucsb.edu/`

[19] ParaScale's homepage: STaaS software
`http://www.parascale.com/`

[20] STaaS software from EMC
`http://www.emc.com/products/detail/software/atmos.htm`

[21] Cloudstore STaaS software
`http://kosmosfs.sourceforge.net/`

[22] Wikipedia - Cloud computing
`http://en.wikipedia.org/wiki/Cloud_computing`

[23] Wikipedia on SAN
`http://en.wikipedia.org/wiki/Storage_area_network`

[24] G. Dierick and P. Geens (KHLeuven), *Selected topics in system administration*

[25] Wikipedia on NAS
`http://en.wikipedia.org/wiki/Network-attached_storage`

[26] Article on NAS' scalability
`http://searchstorage.techtarget.com.au/tips/28468-How-to-determine-a-NAS-system-s-scalability`

[27] Article on SAN management
`http://searchstorage.techtarget.com/magazineFeature/0,296894,sid5_gci1258035,00.html`

[28] POWER AND STORAGE: THE HIDDEN COST OF OWNERSHIP
`http://www.tgc.com/dsstar/03/0805/106372.html`

[29] Force.com website
`http://www.salesforce.com/platform/`

[30] Article on SalesForce infrastructure
`http://blogs.zdnet.com/SAAS/?p=105`

[31] Open-source VM-based lease manager
`http://haizea.cs.uchicago.edu/`

[32] The Reservoir project
http://www.reservoir-fp7.eu/presentations/

[33] The PAN-European network interconnects major European cities by means of a mesh of high-capacity optical fiber links.
http://ieeexplore.ieee.org/Xplore/login.jsp?url=/iel3/3845/11213/00540713.pdf?temp=x

[34] GÉANT2, the high-bandwidth, academic Internet serving Europes research and education community.
http://www.geant2.net/

[35] CERN Advanced STORage manager website
http://castor.web.cern.ch/castor/