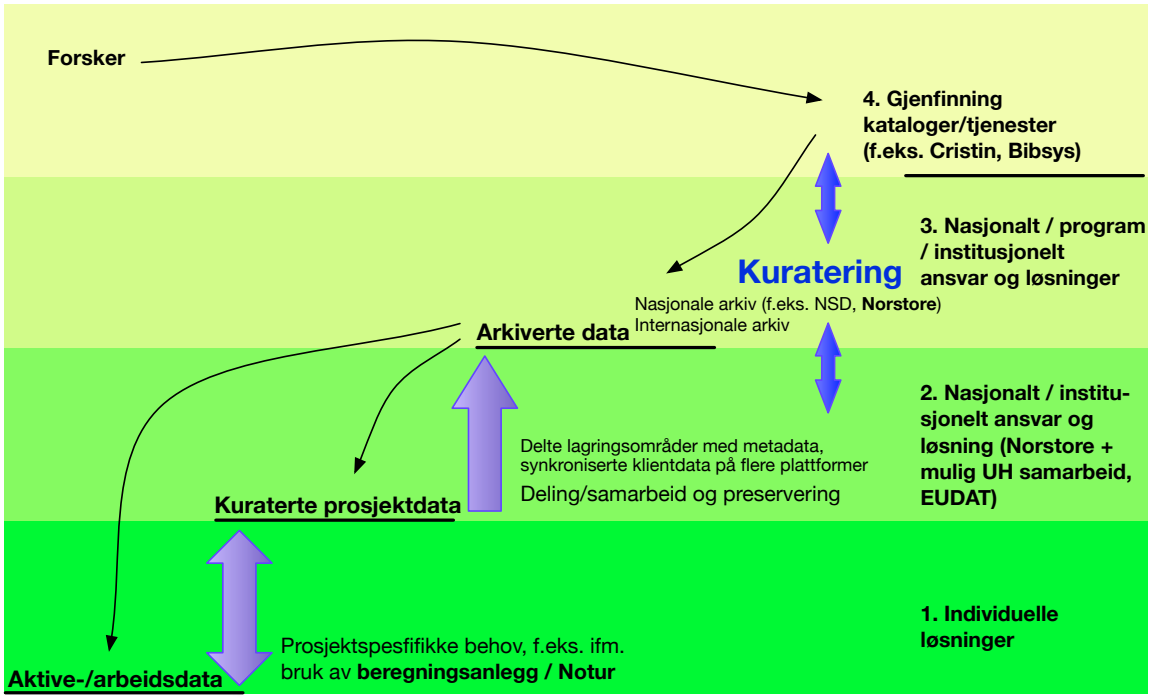
Notat

Til: Sigma2 / Sigma2 styre / universitetenes kunderepresentanter
Fra: Hans A. Eide (UiO), Bjørn Lindi (NTNU), Kjell Petersen (UiB), Alexander Oltu (UiB), Lars Ailo Bongo (UiT), Roy Dragseth (UiT)
Dato: 16.9.2015

Gjelder: **Avklaring, rolle og ansvarsområde for Norstore**

Vi mener at ansvarsområdet til Norstore må defineres til å dreie seg om forskningsdataproblematikken og *ikke* inkludere ansvar for midlertidig lagring i forbindelse med beregninger på HPC anlegg.

Figuren under kan tjene som illustrasjon. Kort beskrevet kan vi si at Noturs ansvar er data i forbindelse med egne tjenester (nivå 1), mens Norstores ansvar er fra nivå 2 og oppover.



Figur 1: Illustrasjon som viser ulike data- og metadatanivåer, dataflyten mellom disse og ansvar for løsninger

Dette fjerner ikke behovet for løsninger for transport av data til og fra regneanlegg og f.eks. Norstore, det hindrer heller ikke at noe Norstore lagring kan være i forbindelse med regneanlegg, men det eliminerer avhengigheter (f.eks. det at man må søke om hhv. Notur og Norstore prosjekt dersom man har ett beregningsprosjekt med stort datagrunnlag) og nødvendigheten av samlokalisering av alle Notur og Norstore ressurser og de føringene/begrensningene/kostnadene dette kan medføre. Norstore må ha brukerfokus på å legge til rette for at det blir så enkelt som mulig for forskerne å ta vare på verdifulle data, legge til metadata og gjøre data tilgjengelige.

Dagens løsning hvor infrastruktur i Notur er adskilt fra Norstore har blitt kritisert av brukere som har behov for å flytte data mellom Notur (beregningssystemer) og Norstore (lagringssystemer), f.eks. geo/klima miljøene. En tettere integrasjon mellom Notur og Norstore utover brukerinitiert manuell datakopiering er etterspurt. Men, det er ikke slik at grensen mellom de behov Notur dekker og de behov Norstore dekker nødvendigvis skal gå mellom CPUer og harddisker.

Aktive datasett, som er en del av datagrunnlaget til *pågående* Notur prosjekt og som benyttes i beregninger/analyse, må kunne lagres i forbindelse med det/de regneanlegg hvor de benyttes og utnytte de tekniske løsningene i regneanlegget. Slike data kan ved prosjektets oppstart komme fra et arkiv (kopi), men det er ikke gitt at alle data skal arkiveres når prosjektet er ferdig. Arbeidsdata i forbindelse med pågående prosjekt har ofte ikke metadata assosiert ved seg (men vil få det ved arkivering). Lagringssystemer for arbeidsdata kan gjøres rimeligere enn lagringssystemer som benyttes til faktisk kjørende beregninger (HPC/scratch lagring) og tillater derfor lagringsløsninger av anselig størrelse. Størrelsen på ulike typer lagring må tilpasses applikasjonsprofilen ved hver beregningsressurs, for å minimere behov for unødvendig overføring av data og for å effektivisere prosjektenes arbeid. Ved tildeling av regnetid på et regneanlegg må prosjektene kunne spesifisere behovet for lagring, og nødvendig lagring følge med CPU allokeringene.

Uavhengig av beregningsorientert forskning har vi et økende problem relatert til håndtering av forskningsdata i Norge. NFR og andre finansiører krever nå at forskningsdata behandles etter bestemte retningslinjer (f.eks. Open Access, og at det foreligger en data management plan), og det er et tiltakende problem at offentlig finansierte forskningsdata ikke stilles tilveie, kan gjenfinnes eller lar seg gjenbruke pga. manglende tjenester og metadata. Tilsvarende er etterprøvbarhet avhengig av at dataene er tilgjengelige og kuratert. Forskere trenger å kunne dele data seg imellom, innad i et prosjekt, eller med samarbeidspartnere i inn og utland, på en sikker måte (tilgangskontroll), med metadata og versjonskontroll. Krav om etterprøvbarhet og reproduserbarhet er økende og fordrer at data preserveres i sikre systemer med tilstrekkelige metadata. Metasenteret må ta ansvar for å lage tjenester og stille tilveie ressurser som gjør det mulig for forskerne å etterkomme finansiørenes og institusjonenes krav, som gjør det mulig å dele data på en sikker men åpen måte, og som medvirker til at verdifulle forskningsdata ikke går tapt.

Denne sistnevnte oppgaven alene er *formidabel* og *krevende* og bør være Norstore sitt hovedansvarsområde.

FOR-ANS: requirements and evaluation for a future national data infrastructure

Authors: A O Jaunsen, G S Dahiya, O W Saastad
Date: Oct 26, 2015

On behalf of FOR-ANS technical working group for data storage
(A O Jaunsen, G S Dahiya, A Paalsrud, M W Forsbring, N Budewitz, A D Fidjestøl)

Mandate:

“Mandat – utrede og anbefale krav til arkitektur, teknologi og struktur for nye nasjonale lagringsressurser”

Executive summary

The mandate of the working group was to a make a technical assessment of the requirements for a future national data infrastructure for research in Norway based on existing user requirements. In a meeting between representatives of the HPC and data storage technical working groups on Sep 17, 2015 it was requested that the report should (also) clarify any requirements or consequences of various location scenarios in relation to the localisation of HPC resources, e.g. the central components of the national e-infrastructure.

In the following report the FOR-ANS technical working group for a national data e-infrastructure for scientific research in Norway has investigated the needs and requirements from users and communities both from the existing user pool, selected communities from the national roadmap for research infrastructures, as well as strategic considerations.

The requirements suggest that data must be connected to resources that allow the users to process, analyse, share and publish the data. The data must be accessible in a simple and consistent manner that accommodates not only traditional compute-intensive users, but also new user communities that are data-driven. Infrastructures for research and larger laboratory facilities that generate large amount of data must be enabled to ingest their data via effective protocols that are supported by the data infrastructure. Data must be stored safely with a redundancy that will withstand failure of hardware on any level (disk, server, rack, site), while at the same time avoiding unnecessary duplication of data. The requirements for a given research

dataset may change during its lifetime and the infrastructure should enable seamless migration of data between the relevant storage devices based on data management policies and usage. The future national data infrastructure must be a scalable, reliable and flexible facility that can accommodate the vast majority of relevant users and communities.

The working group has evaluated the alternatives presented in the document with similar redundancy requirements and based on the criteria listed in the table of section “Evaluation and characteristics”. We have prioritized the recommended solutions (and localisations) in the following order:

1. **CDC (single site).** There are obvious benefits to be gained by co-locating the HPC and data resources. These include, but are not limited to, data availability, HPC integration, less network dependency, simpler operations, lower procurement and operational costs, better resource utilization and user friendliness. We estimate the procurement savings to account for approximately 16-27% of the storage capacity costs over the distributed data centre. Additionally there are operational cost savings due to the reduced network costs and fewer required staff (as there is only one site to operate).

2. **DDC (two sites).** In the distributed data centre (DDC) the storage resources are spread across two (or more) sites that are inherently integrated components of a single data infrastructure. It is assumed that each of the two HPC systems are located in the same facilities as the storage resources. It comes at a cost increase due to the added site and capacity, but with the benefit of dual-site security (each site can operate independently with full disk/server/rack and site redundancy).

3. **DDA (three or more sites).** The alternate dispersed data site (DDA) proposal is primarily intended to decouple the resources integrating with the core infrastructure (such as HPC installations) from the national data infrastructure itself. In this context such decoupling is not needed if the number of sites is less than three, because the configuration then is identical to the co-located (one site) and distributed model (two sites). With three or more sites, the challenges of data proximity to HPC resources (a disadvantage of the other dispersed models) is solved by dedicated storage resources on the HPC sites. More intricate rules for data replication is handled by introducing project-specific data management policies. The rating of this proposal is comparable to the DDC, but is unlikely to achieve the same level of data accessibility and resource utilisation. It will require a stronger dependence on network connectivity and likely somewhat higher procurement costs (due to the number of sites).

Introduction

This document describes possible solutions for the next national data infrastructure for research and will serve as the basis for a tender document on new storage facilities and its requirements

specifications, due to be released during the autumn/winter 2015. The document is based on user surveys, direct interview/feedback and more than five years of experience from operating the existing infrastructure for data services. The document is focused towards a data centric[1] infrastructure providing a variety of data services. A few conceptually different scenarios are discussed and compared.

- a co-located infrastructure
- a distributed infrastructure
- dispersed infrastructures

The *co-located infrastructure* consists of a single data centre in which data and services (e.g. HPC, data analytics, visualisation) are fully integrated and connected by high performance network or interconnects. It requires the storage resources to be placed in separate data rooms that are sufficiently secure and resistant to fire, flooding and other relevant disaster events. The infrastructure is considered to be located in a single geographical location and benefits from the obvious advantages of the hardware components being located in close proximity of each other.

In the *distributed infrastructure* the storage resources are located in separate geographical locations and thus consist of at least two data centres. This infrastructure relies on replicating data between the two data centres with a redundancy[2] that allows one of the data centres to be lost or become unavailable without loosing any data. This solution will thus rely on an adequate network bandwidth between the data centres. Services are provided on all data centres and each of the HPC systems are directly connected to a data centre to achieve a good integration among all infrastructure components.

The *dispersed infrastructures* are characterised by being largely autonomous and separate resources with a lower degree of integration. In these configurations one or more HPC system is not located in proximity of a data centre, thus requiring dedicated storage resources to be procured for such HPC systems. The dispersed alternatives provide a larger degree of autonomy, but result in service fragmentation, isolation of data and a modest infrastructure integration level.

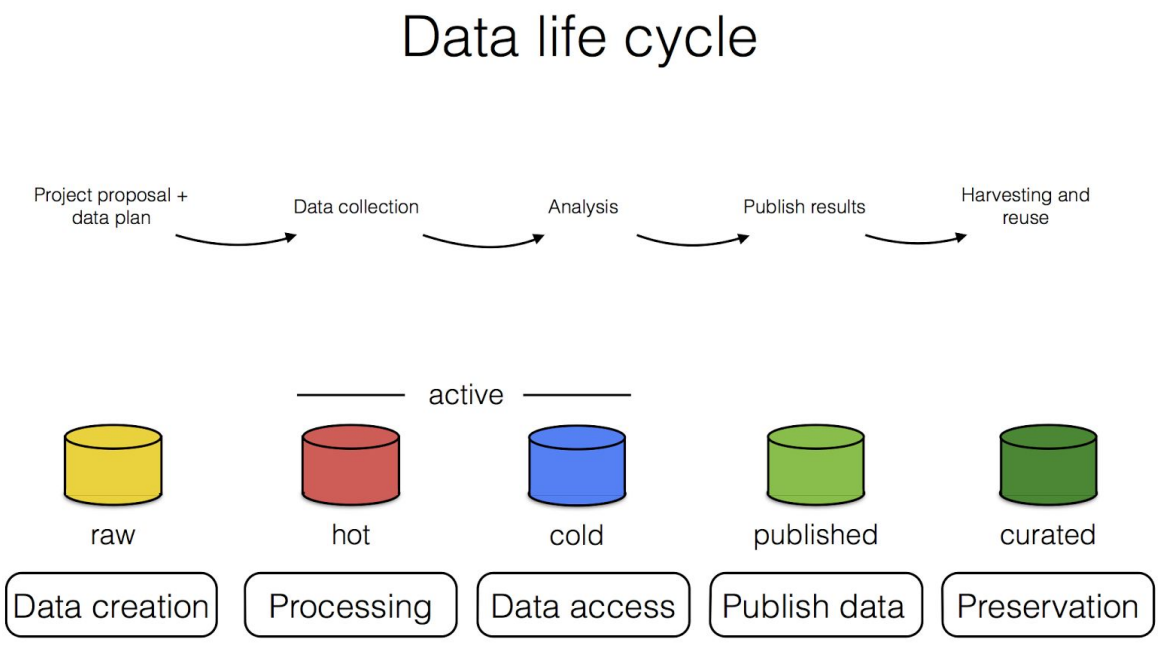
A major challenge for any future data infrastructure is how to tackle the growth of data. The data does not only need to be stored safely, but also need to be connected to resources to enable services such as e.g. compute, visualization, analyses and other miscellaneous services. Data accessibility is therefore of key importance. Transporting data is already today a growing challenge with the growth of data exceeding the network capacity increments. Dedicated high bandwidth network links come at a significant cost. A data-centric architecture is therefore an

[1] Def. data-centric – an infrastructure where the user interacts with stored data across services without data movement and duplication.
[2] "Redundancy - Wikipedia, the free encyclopedia." 2011. 19 Oct. 2015 <https://en.wikipedia.org/wiki/Redundancy>

appealing concept that is well suited to meet the challenges of a future national e-infrastructure for scientific research.

Data Life Cycle

Data is categorised in different classes that represent the relevant stages of the data life cycle. This cycle typically starts with the creation of data in lab/experiments or computing environments, encompasses an ‘active’ phase in which the data is processed and prepared for interpretation and analysis. When the data has been thoroughly analysed (hopefully resulting in scientific achievements) it is expected that the data is archived and possibly published. The various categories are shown and discussed below.



The illustration shows the typical phases of the scientific process from a data-centric perspective. The classification of data aims to highlight the use of the data, provenance and potential points of branching. In the table a description of the various data classes is given.

Class	**Description**
raw (/project)	Data that is not reproducible and in its pristine state is considered ‘raw’. Such data is typically the result of recorded measurements by an instrument or experiment, typically requiring further processing to become meaningful. Raw data is static by definition and may be required in the future to enable reproduction of previous results or to verify/dismiss

	claimed errors. This class of data is therefore valuable and best practices often recommend that such data must be secured for the future.
hot (/home, /project, /work)	Data that is in active use and accessed frequently is dubbed 'hot'. Hot data is typically accessed, processed and can serve as input to new calculations on an HPC system for instance. It is therefore necessary to maintain this data on a storage technology with high read and write performance while at the same time coping with multiple users and process. The performance is determined by the connectivity between the storage system and the resources (such as HPC). Latency should be on the order of a few ms. The storage resource performance for the filesystem may be improved by use of fast storage mediums (SSDs) for caching data.
cold (/project, /home)	Data that is still relevant, but accessed less frequently. Cold data should be accessible via various protocols (i.e. POSIX, S3, HTTP-REST), but can be stored on less expensive hardware (i.e. SATA drives). Typical data access times is about 100 milliseconds.
copy (/project/copy)*	*Data that serves as a (backup) copy only. This is a subclass of cold data, but with no redundancy policy and liberal access time requirements (hours to days). Data is only accessed in the event of the (external) data becoming corrupted. Data of this type must be stored with a reference checksum value.*
published (not mounted)	Data that is archived and published by issuing a DOI. Such data is typically archived for 10 years, but with no curation requirements. It is expected that the data is no longer useful after a decade and can, in principle, be deleted after such time.
curated (not mounted)	Published data that has a (documented) need for long-term preservation and 'permanent storage' requirement must be curated by a data librarian and set up with a preservation plan.

Requirements

User Requirements & Services

UNINETT Sigma2 completed a user survey in June, 2015 to get current and future requirements from scientific communities in Norway. The result from this survey shows usage of various NorStore data storage services provided currently. The service is used for active, backup and archival use cases. Considering the various use cases, users have requested the data to be kept for 3 months to multiple years. To access the services, users are currently using standard tools e.g. SSH/SFTP. However, there is a clear demand to have a user friendly and modern methods e.g. Desktop client using WebDAV/SMB or Dropbox like Sync n Share to interact with data services.

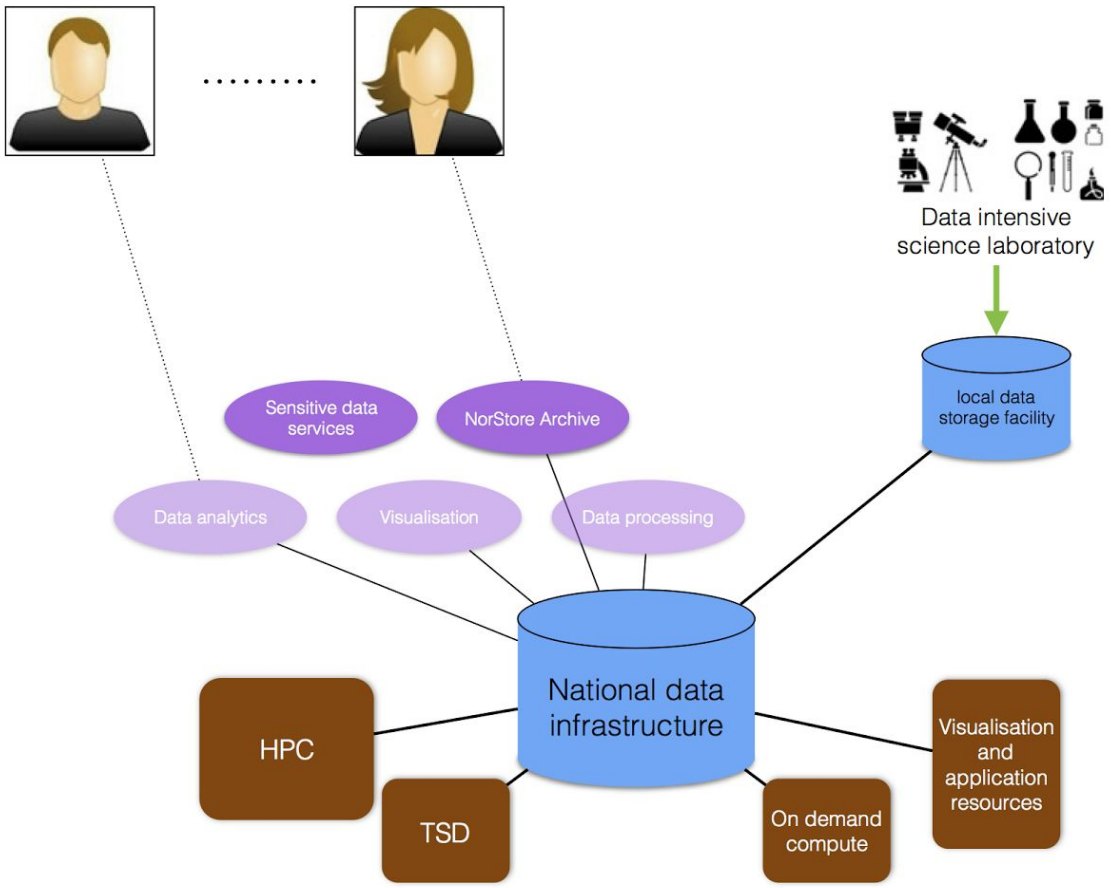
Currently when users need to process data stored in Norstore, they are required to copy this data manually to the HPC project area. This results in duplication[3] of data as well as limiting the users to the storage capacity available on the HPC facility. Moreover, it results in data getting out of sync due to manual copy and modification which further results in bad users experience. In the user survey, users have asked that they would like to have data directly accessible on the HPC system to process it. This will result in a better utilization of storage resources by avoiding duplication and offering a smoother user experience. Finally, there is an increasing need for storage and data management services for non-HPC users, and in particular shared project areas with fine-grained access control, metadata and publishing services.

In addition to the current data services, users have shown interest in dedicated resources, data analytics and visualisation services. Dedicated resource service enables users to have a certain amount of reserved compute and storage to run user's desired tools. The compute and storage resources are permanently reserved for the duration of resources allocated to a user. Data analytics service represents analysing big data using frameworks e.g. Apache Spark/Hadoop. Visualization service represents offering a service to remotely visualize large datasets using dedicated hardware e.g. GPUs.

A challenge presented by the Life Science community is the legislative requirement to protect the sensitive data originating from any human material. This poses new challenges in securing and restricting access and integrity of this information. Services for sensitive data are required by a sizable community and discussed later in this document.

In a data-centric infrastructure resources are attached to the data providing the necessary support for established data services. With time such resources and services will come (and go) but the data infrastructure remains independent of these securing data accessibility to users. Laboratories and research infrastructures generating large or steady streams of data that should be stored and used nationally may benefit from permanently connecting their resources to the national data infrastructure via suitable protocols such as i.e. S3, REST or Sync n share (Dropbox like). Examples of such facilities may be genome sequencers or other high data-volume research instruments.

[3] "Duplication - Wikipedia, the free encyclopedia." 2011. 19 Oct. 2015 <https://en.wikipedia.org/wiki/Duplication>



Sensitive data services is an important part of the service provisioning and a significant number of users within the fields of medicine and life science rely on this service today. The level of service requirements among various user groups and communities is not identical and may require providing different services in the future. Currently a national service is provided by UiO/USIT. UNINETT Sigma AS supported the development of TSD and co-funded the efforts in the period 2012-14. National resources are allocated through the RFK based on received applications in response to biannual calls. We consider this service to be tightly integrated with the existing infrastructure at UiO and moving this service to a new infrastructure will require a significant effort. It is beyond the scope of this document to explore the challenges this may incur and we recommend that the alternatives and possible solutions are considered separately in a dedicated project.

HPC Requirements for Data Access

The HPC systems need access to fast direct attached (same connection using the interconnect fabric, often InfiniBand) storage pool. Jobs need to access a high performance, high throughput and low latency storage system in order to store data during computation and to read input files that are accessed many times during a run and same for storing data and files during runs. This

storage system needs to be as fast as practically possible without excessive cost in comparison to the HPC system's budget.

This high performance system is generally not backed up and is not expected to be used as neither short term nor long term storage. It should be used as scratch storage space. Files stored in scratch space usually have retention time in the order of days, so files older than 30-45 days are deleted without user notification. This high performance storage system is an integral part of the HPC, hence co-located with the HPC system (distance of a few meters). The commonly used name for this storage are scratch or /work (/work is the normal mount point name).

A HPC facility also need access to a short and medium term storage. Generally referred to as home, since the user's personal data is stored here. Each user get allocated space on the home area and will use this space for daily computation. Output data and results are processed and transferred from the fast work area. The home area is generally backed up so files there are fairly safe and there is no retention limitation on the files. These files might live forever e.g scripts, source code. The performance of home area are generally somewhat less than that of the work area. Some HPC systems provide home using the same global parallel file system as that used for work while other accept lower performance (NFS). For jobs that have moderate I/O demands, it is common to run them directly from the home directory while jobs with more demanding I/O requirements use the work area. The fraction of jobs that run on which type of storage system will vary with the performance of the home area. The table below shows the performance characteristics of various storage areas.

File system (POSIX mount point)	Expected Throughput [GBytes/s]	Expected IO performance [IO operations / s]	Approx size [PetaBytes]
/work (scratch)	30 - 100	100 000 +	1-3
/home	10 - 30	10 000 - 50 000	1-2
/project	5 - 10	5000 - 10000	4-8

In addition to home/work data, there is often a need to access large data volumes with reference data such as databases of proteins, genomes etc or large amount historic data or output from earlier runs. This data may not be accessible with the same (high) performance as data in work/home area and this data is very often shared by a group of users in a specific project. The common name for this kind of storage area is /project. While performance is not of primary concern it is still accessed as files using a POSIX file system, e.g. mounted directories on the HPC system. The users typically want to access and process large amounts of data found in this area. Another feature of this project storage is the fact that it is shared by a project group during the lifetime of the project.

A data-centric data infrastructure with large storage capacity is well suited to host these data types. Access to the centralised project data should be provided to the HPC users using the POSIX file system access and/or HTTP REST based. This will avoid the need to move data around and enable data to be served directly to the HPC systems with moderate performance.

Data Processing and Visualisation Services

Access to data is also needed for data intensive services not only to compute intensive service as HPC. These are services where number of operations per byte is very small, like scaling all entries in a file, changing format on an image, a genome, or transcoding a video etc. These jobs are generally limited by the IO system's capacity to stream data and the processing elements to handle these large amount of data. Many such operations are requested to be done in real time.

Another example of data centric service is visualisation and animations. This often require huge amounts of data to be processed to make images that are either displayed or sequenced to make animations. In addition comes graphic interfaces that are used interactively to access and process large amounts of data.

With the recent increase in amount of unstructured/machine generated data, many open source frameworks has been developed to process large data sets in parallel running on commodity hardware. Such processing requires access to vast quantities of data collected from different sources like array of sensors, data from the internet, genomic data or other sources of unstructured data. There are many ways of doing such analytics, distributed or more centralised. The frameworks, e.g. Apache Spark/Hadoop, to process such data sets provide distributed fault tolerance abilities. These frameworks accept running analysis on commodity hardware due to their inbuilt fault tolerance against server/network failure. Considering the large data sets and abilities to run on commodity hardware, it is better suited to run them outside HPC system.

The infrastructure for these services must be in close proximity to the storage facility as the bandwidth and latency requirements are of an order that prohibits transport over long distances. Historically users have been using their workstation to perform such analysis. With the data growth, this is no longer a sustainable option.

Storage Requirements for Services

Taking the user requirements for current and future services into consideration, we require the storage solution to be **scalable**, **reliable** and **flexible**. The storage system should provide multiple methods to interact with e.g. POSIX file system, HTTP REST, Amazon S3 APIs. The focus of the storage system is to be an enabler of new services and allow users to interact with national services in an intuitive and flexible way. The storage system should provide a global name space and enabling users to access their data from any resource is a key feature. The users should be able to access data interchangeably from POSIX file system, HTTP APIs or any other supported protocols. The storage system should support the use of different storage

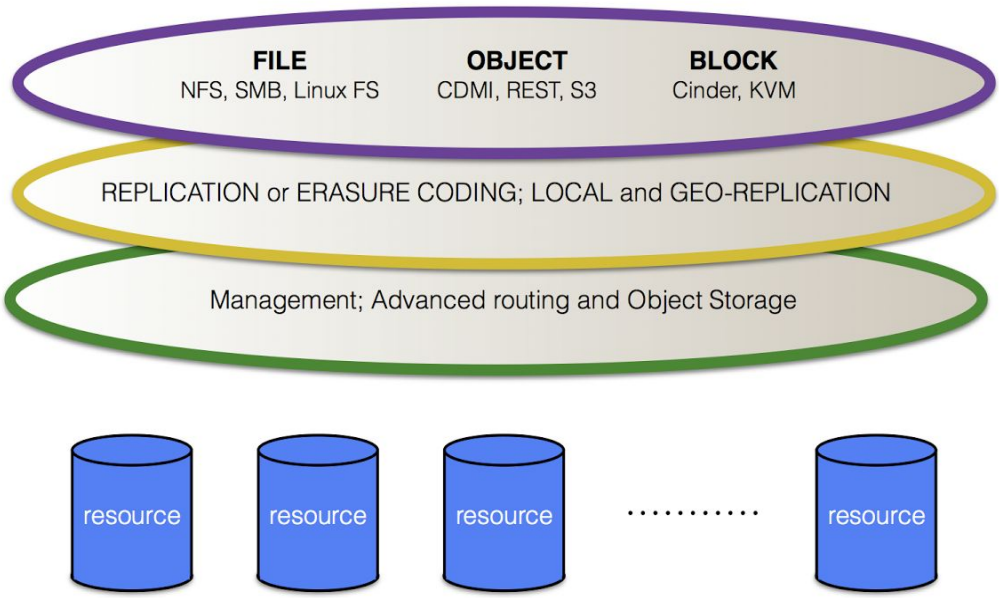
technologies to achieve high performance (read+write) and balance between low cost capacity storage. Depending on the access pattern of specific data objects the system should automatically move the data between the storage pools; e.g. moving data that was frequently modified / accessed from the performance storage pool to the more cost effective storage pool (e.g. erasure coded storage on SATA disks).

To avoid the data duplication between HPC and data centric services, data from the storage system should be accessible on all the national services. The users can deposit data in the storage system using different methods e.g. Desktop clients using WebDAV, Sync-n-Share (Dropbox like), SSH/SFTP, HTTP REST etc. Once data is received in the storage system, users can access this data from the HPC or any other data services offered by the national e-infrastructure. Thus avoiding data movement and duplication while enabling a smoother and more user-friendly experience without compromising the performance.

The storage system should be designed to be free from single point of failures (SPOFs), mainly by ensuring component redundancy. The system should be able to operate normally under disk/controller/server/rack/site failure given enough free capacity. The free capacity needed depends on the type of data. The additional capacity varies from 35-45% for active data resources. The archive data will be replicated on both sites, therefore does not require additional capacity in case of site failure. To provide redundancy in case of fire or natural disaster, the storage system needs to have a minimum of two sites[4]. Data must be replicated between these sites to achieve high availability in case one site becomes unavailable. The replication can be performed asynchronously to avoid the performance penalty and reducing the requirement for large network bandwidth between two sites.

The figure below shows how such a storage system can be achieved. At the bottom the pool of resources represent the hardware such as storage resources and servers that are needed to provide the services. These resources are controlled by layers consisting of (software) technologies and policies that enable the management and access to the resources. The bottom layer represents the way in which physical resources are managed and made available to the layer above as objects. The middle layer takes care of how objects are stored either being replicated or erasure coded and to which rack or site. This takes care of the failure domain to keep data available despite failing hardware on any level. The top layer represents protocols that enable clients/users to interact with the storage system. It enables a coherent view of stored data across different categories of protocols.

[4] Def. site - a separate location, facility or room



Protocols for Interacting with Data Storage Resources

The table below lists the expected relevant protocols for various key tasks.

User data deposit/access	Sync n Share (WebDAV or similar), SSH/SFTP, Object based HTTP REST API (S3)
HPC Data Access	Parallel POSIX compliant filesystem, Object based HTTP REST API (S3)
Data Analytics	Parallel POSIX/Hadoop compliant file system, Object based HTTP REST API (S3)
Dedicated Resources	Block storage for virtual machines, POSIX/SMB file system access
Visualisation	Parallel POSIX compliant file system

Infrastructure configuration alternatives

Configuration scenarios

To safeguard against loss of data in the event of catastrophic events such as a fire or flooding it is necessary to require data redundancy between two 'sites'. In this way one data centre site can be lost due to a fire or flooding while the other site will still provide access to the data and services. In the unlikely event of such a scenario the data redundancy must be restored within the remaining site (provided there is sufficient resources available) or to a third (backup) site. A national data infrastructure will therefore need a minimum of two sites for reliability and availability reasons. There are several ways to implement such configurations. In the table below we provide the various configuration alternatives taking into account the localisation and connectivity to the HPC system(s).

Configuration label	**Description**	**Site A**	**Site B**	**Site C**	**Site D**	**Data link**	**HPC/DS connectivity**
I	Colocated Data Centre (CDC)	HPC1+DR1 HPC2+DR2				IB/LAN	IB/LAN
II	Distributed Data Centre (DDC)	HPC1+ DS1		HPC2 +DS2		WAN	IB/LAN
IIIa	Partially Dispersed Sites (PDS)	HPC1+ DS1	DS2	HPC2		WAN	IB/LAN / WAN
IIIb	Fully Dispersed Sites (FDS)	HPC1	HPC2	DS1	DS2	WAN	WAN
IIIc	Alternate dispersed data sites (DDA)	HPC1+ storage	HPC2+ storage	DS1	DS2	WAN	IB/LAN

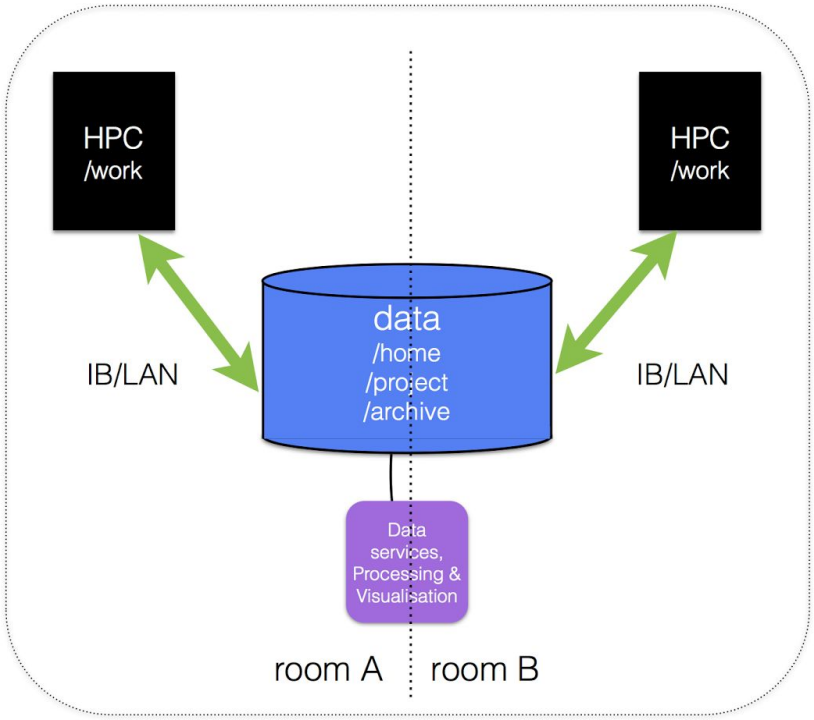
Colocated Data Centre

The CDC configuration (I) is fully co-located. The HPC is connected to the data resources by high capacity interconnect or LAN. The data centre[5] consists of two data rooms[6] (DR1 and DR2)

[5] "Data center - Wikipedia, the free encyclopedia." 2011. 8 Oct. 2015 <https://en.wikipedia.org/wiki/Data_center>

[6] "Data room - Wikipedia, the free encyclopedia." 2011. 7 Oct. 2015 <https://en.wikipedia.org/wiki/Data_room>

are isolated from each other, fire and water insulated, and close enough to establish a high capacity link. This link can be achieved by InfiniBand or Ethernet (several 100 GbE lines). If InfiniBand is chosen it is necessary that the two rooms are within a 100m of each other (see limits on EDR InfiniBand). Whereas in case of Ethernet, they can be separated by more than 100m and only limited by the cabling costs. The illustration below shows the configuration where HPC and storage sites are physically co-located.

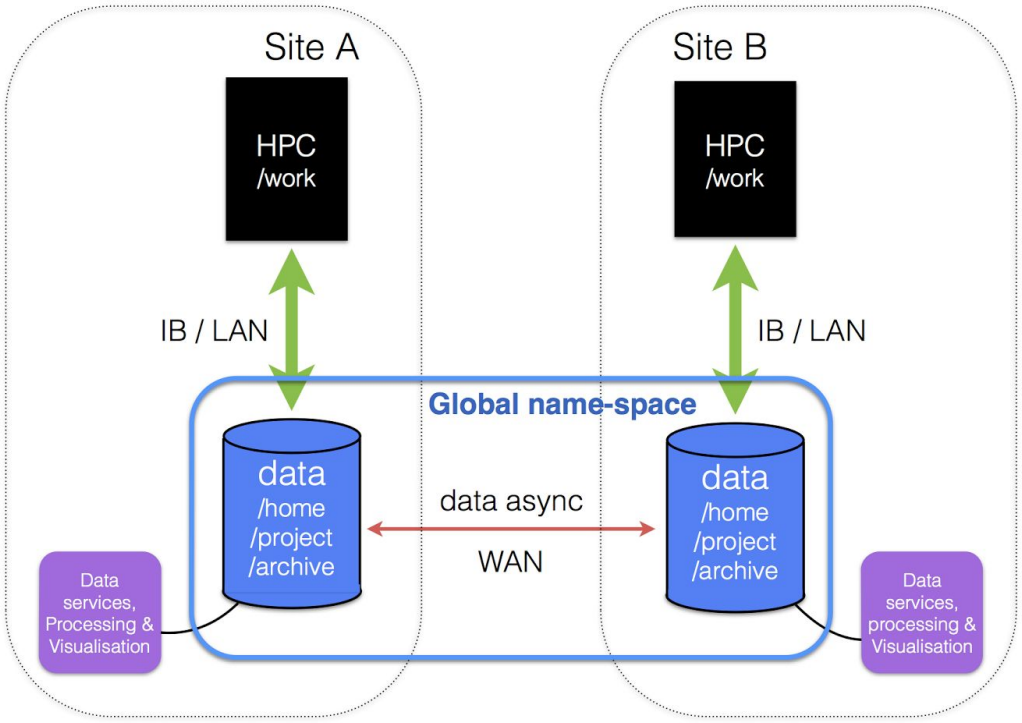


The high speed interconnect will enable very quick synchronisation between the data rooms, allowing a fully synchronised storage system. A high-speed connectivity between HPC and data enables an infrastructure with centralised national data services. The scratch storage (/work) is provided within the HPC system using its own interconnect fabric while /home and /project are provided as part of the data centric infrastructure with a parallel POSIX compliant file system. This design enables the two-site requirement for the storage system to provide safety in case of power/server/rack failures, but not in case of natural disaster.

Distributed Data Centre

In the DDC configuration (II), the two sites are geographically separated by distances of a few kilometers (typically hundreds of kilometers). This means that two data centres are required and the interconnect between the two centres will rely on Wide Area Network (WAN) (either connected via Forskningsnettet or a dedicated fiber). Latency limits the performance of data replication over such distances and it is necessary to rely on asynchronous replication between sites. Within a site this configuration retains the performance between the HPC and the data

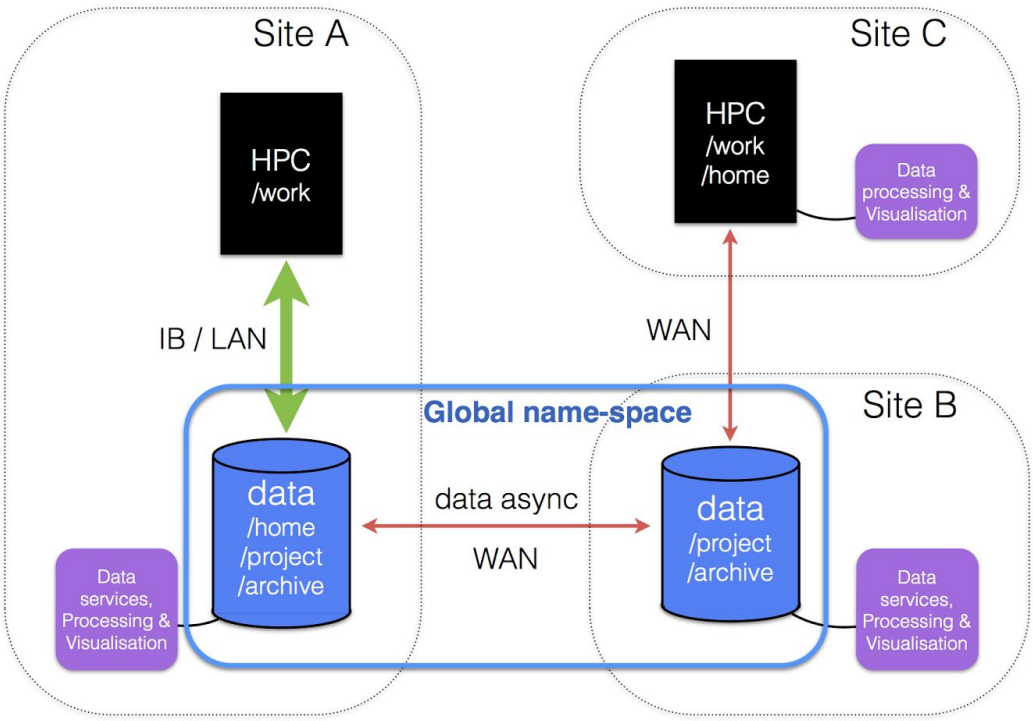
site, but the data replication between site A and B will be asynchronous (data sync is not achieved/guaranteed at all times). Services such as visualisation and data analytics must be offered on both sites and therefore will result in extra operational cost.



This configuration has the benefit that it is somewhat better equipped to resist catastrophic events that would take out one entire data centre, while at the same time maintaining all data intact on the remaining site. It does require a higher degree of storage redundancy (see table in section “Capacity and physical characteristics”) and will likely incur higher operational costs. We assume that service resources such as e.g. data processing and visualisation can be divided between the two sites.

Dispersed Sites

In this section we discuss the dispersed configurations (IIIa-b), where at least one HPC system is not co-located with a NorStore data centre. The illustration below shows the implications on connectivity between an HPC system and the data infrastructure for one such configuration (IIIa), due to these being operated at separate physical locations.



In the PDS alternative, we have one site where the HPC is located in a data centre, while the second HPC is not (see illustration above). This will result in the separate operation and maintenance routine for both sites and incurring unnecessary cost. At one site we will have better service experience whereas on the other site it will be sub optimal. In the FDS alternative, all the HPC and data sites are located on separate physical locations. This will require accessing data from storage services via a comparatively low performing WAN link. Ideally, one would hope for a high capacity link between an HPC and a data site. Based on the expected throughput mentioned in table (see section “HPC requirements for data access”) a dedicated link with bandwidth of 40Gbps between site B and site C is needed. The user experience will nevertheless be degraded due to the unavoidable effects of latency. The typical cost of a 10Gbps link between two sites (500km apart) is approximately 400 000 kr/year.

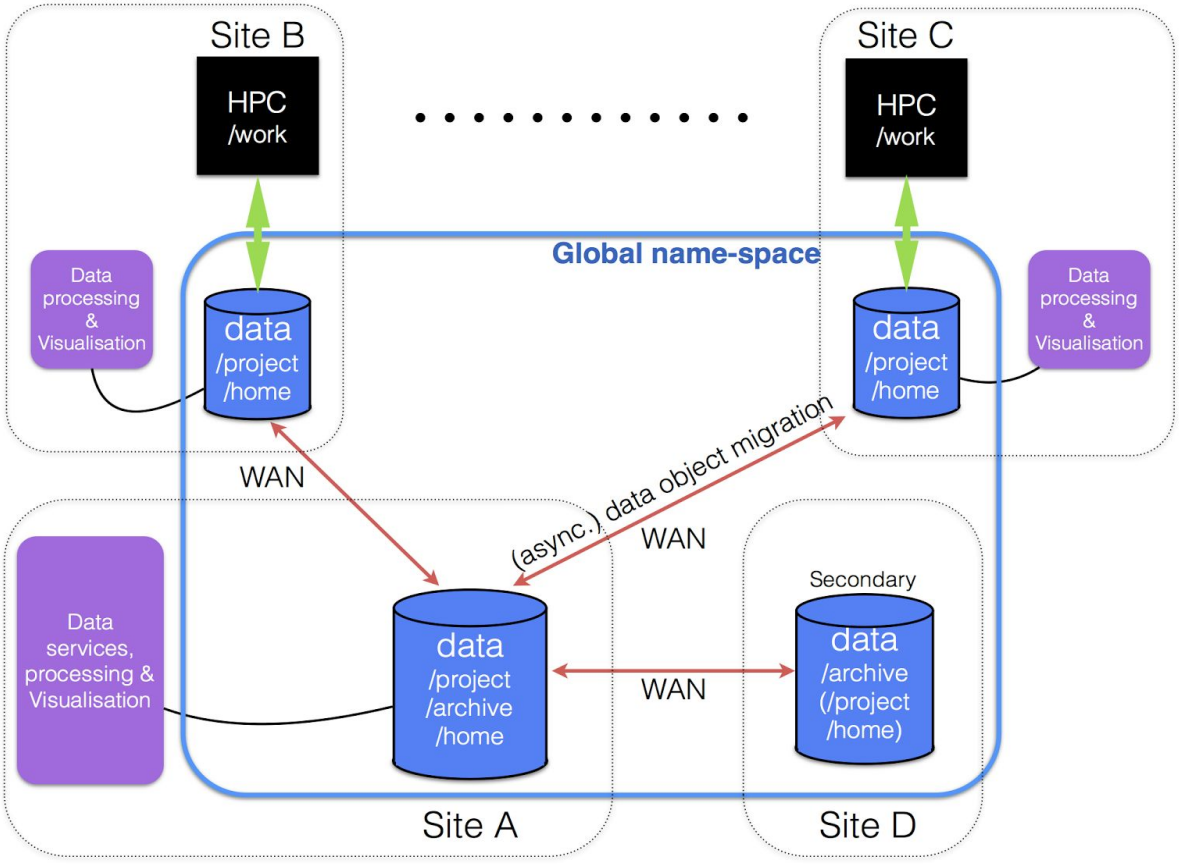
With a distributed setup where the HPC facilities and data sites are not co-located the /work and /home must be provided locally within the HPC system, as users need /home to be relatively fast and providing it over WAN is not feasible. To access /project, it needs to be available over WAN with a link from the HPC facility to the data site. This transport layer must also accommodate backup, as this is impractical to provide locally for each HPC facility.

This is a more complex solution relying on the availability of high performance links for a large amount of data access. The cost of transporting ever-increasing volumes of data also rises dramatically as network bandwidth does not grow with the same rate as data volume. For this reason the gap between network bandwidth and data volume is expected to be increasing faster in the years to come [Cisco report: “The Zettabyte Era”]. With the dispersed setup, it is not

realistic to provide the HPC users with POSIX access to a remote /project due to latency and low bandwidth. Such a solution cannot be regarded as ‘data centric‘. Moreover, it does not meet the user requirements for accessibility and flexibility and results in data duplication between HPC and data sites.

On Sep 16, 2015 the FOR-ANS working groups for data and computing received a memo from the site managers, suggesting to decouple the national research data infrastructure from the HPC resources. The alternative is suggested to be data-centric as data-producing or data-consuming resources and services (such as HPC) integrate with the core data architecture as they come and go. The memo points out that the core services, such as the archive, project data, publishing/sharing, data management tools, search and indexing functions are independent of HPC, and that this architecture and corresponding infrastructure has a different life-cycle than any given HPC resource, which typically have a life-span of four years. According to the proposers this is not to mean that the data infrastructure wouldn’t benefit from co-location with HPC systems or other data sources/consumers.

The alternative (IIIc) is illustrated in the figure below. It assumes a global name-space for research data available on all enabled resources, but with the data objects might not be “physically” located where data is needed. Thus data needs to be moved to make it available on or near the resource that needs it.



The illustration of the alternative (IIIc) above has two HPC sites (sites B and C), and the core data services at a separate dedicated site (site A), but *the core data services might be co-located with any or both the HPC sites*. Similarly, the secondary data site D might also be co-located with either of the HPC sites (as long as the core data services site A is not at the exact same location). Green arrows indicate high performance IB/LAN connections over short distances. Site A (core data services) should be geographically separated from the backup/secondary data site (site D). Note

Backup service

It is necessary to offer a backup service that can secure the history (changes and deletions) in /home area. It may also be relevant to provide a form of backup service for the /project data by snapshots locally on the storage system. The backup service should be cost effective to be feasible for multi-petabyte storage. This service is provided at all four Metacenter partners hence competence is available when needed.

Evaluation and characteristics

Evaluation criteria and scores

The evaluation criteria have been selected on the basis of service quality, user friendliness and cost-effectiveness and constitute the main categories of selection criteria for the infrastructure. In the following we compare the explored configurations and apply scores (High=+1, Medium=0, Low=-1) to attempt a comparison of the proposed alternatives.

Criteria	I (CDC)	II (DDC)	IIIa-b (PDC/FDS)	IIIc (DDA)
Min. number of data centres	one	two	three	three
Service quality				
a. Inter data-centre network independence	High	Medium	Low	Medium[8]
b. Reliability in the event of site facilities failure	Low[7]	High	High	High
c. Resource utilization (no data duplication)	High	Medium	Medium	Medium[8]
User friendliness				
a. HPC integration	High	High	Low	High
b. Data accessibility (access to any data)	High	Medium	Low	Medium[8]
Cost-effectiveness				
a. purchase/setup	High	Medium	Low	Medium
b. operations	High	Medium	Low	Low
c. network	High	Medium	Low	Low
SCORE	6	2	-5	0

[7] depends largely on the facility provisions in terms of protection from fire, flood, power failures, network redundancy etc

[8] at best, in the case where data is processed at only one HPC site. When data needs to be processed at more than one site, the rating goes to low.

Capacity and physical characteristics

Estimates of required capacities are based on existing allocations, requested resources for the coming years and historic evolution of capacity needs. We consider models IIIa-b to be considerably less attractive due to its fragmentation and poor scores in the table above. The three remaining scenarios are considered;

- I) a co-located data centre (CDC) where site A and site B are fused together
- II) a distributed data centre (DDC) where site A and site B are separated
- IIIc) alternate dispersed data sites (DDA) where the number of sites are at least three

The proposed storage system is divided into two types of storage racks; one optimized for performance and the other for capacity. The *performance storage* racks will provide good read and write performance and are intended to serve hot active data which is being processed, updated or modified within a policy defined period. These racks will consist of 6-8TB disks with medium disk density (10-12 disks per U) with a few SSDs to aid performance. Some of the performance rack shelves will be reserved for cpu nodes to provide data intensive services. The *capacity storage* racks are intended to serve the majority of the data, providing comparable read performance (as the performance storage) but lower write performance. These servers will consist of 8-12 TB disks with somewhat higher disk density (12-16 disks per U).

Scenario	Type	Effective capacity 2017 [PiB]	Number of copies	Raw capacity needed [PiB]	Raw capacity per rack [PiB]	# racks
(I) COLO	performance storage	3	4x	12	2.4	5
	capacity storage	7	2.4x	17	3.5	5
(II) NO COLO	performance storage	3	5x	15	2.4	7
	replicated capacity storage	7	2.8x	20	3.5	6
(IIIc) DDA	performance storage	3	5.4-7.8x[9]	16-24	2.4	7-10
	replicated capacity storage	7	2.8x	20	3.5	6

[9] These numbers assume protection against server/rack failures (as in other models). If, on an HPC site, protection is only provided against disk failure the required capacity will be 4.2-5.4x.

The storage capacity numbers are based on a resilience model to protect against a full site failure without loss in data or data availability. Some categories of data may not require this level of redundancy. It is possible to increase the effective capacity of the storage system by introducing a more refined (detailed) data replication policy for the various type of data, projects or services.

The dispersed data architecture with global name-space (DDA) solution puts the responsibility of (HPC/instrument) colocated data resources outside of NorStore. Yet, NorStore would need to provide for the ability to accept and accommodate this data into the central data storage, but the number of copies for active data NorStore would need to provide is reduced. This alternative requires only 13-16 racks for the estimated storage capacity requirements in 2017, but the cost savings will in many cases amount to a cost increases for the HPC centers (or other types of sites integrating with the data architecture), which may ultimately be funded by the same funding source.

The co-located data centers model (COLO) has the advantage in that it only requires 10 racks for the estimated storage capacity requirements in 2017. The advantage comes from the fact that the infrastructure is housed in a single data centre, but divided among two data rooms (“sites”). It benefits from the low latency and high bandwidth connection between the two data rooms, enabling a more integrated infrastructure with higher resource utilization. This solution is still robust against events that would take out one of the two data rooms, leaving the alternate room operational with all services.

The geographically distributed data centre model (II, NO COLO) requires the data to be replicated across two data centres in order to fulfill the high availability requirements in the event of a data centre failure. This requires an addition of six racks, bringing the total number of racks to 13.

An alternative model in NO COLO scenario is a 3-sites model. This model can reduce the amount of raw capacity required for capacity storage and still provide reliability against site failure. This will result in a cost of 1.7 - 1.8x of raw capacity (approx. 4 racks) in comparison to 2.4 - 2.8x to protect against multiple server/disk/rack and a single site failure. In 3 sites model, it would be required to have a large network capacity (multiples of 10 Gbps) and possibly low latency between sites. As data will be spread across 3 sites and any read operation has to go to minimum of 2 sites and write to all 3 sites. Moreover the erasure coded data pools require large amount of data transfer to recover from failures.

For scenarios IIIa-b it is more challenging to estimate the required number of racks per site due to lack of information (the distribution of data). These numbers are therefore not included here.

Rack Dimensions and Power Requirements

Open Rack form factor (W: 600mm, H: 2200mm, D: 1066mm) and with footprint of 0.64 m^2. The power capacity is upto 25 kW per rack and the power requirement per rack will be slightly different for archive data and for active data. For archive data the requirement will be around 8-10 kW, whereas for active data it will be around 12-14 kW. Active data racks will also contain compute resources to provide data centric services. The actual power consumption is expected to be 5-7 kW for capacity data racks and 7-9 kW for the performance data racks.

Contingency

The infrastructure is architected in such a way that the two-site (or two-room) redundancy ensures data accessibility in case of a catastrophic event at one data site (where that site is lost entirely). To restore an acceptable data redundancy in case of site failure it will be necessary for the remaining data site to have sufficient available capacity to rebuild its site redundancy from 1-2x to 3x or similar. This can be achieved in a matter of hours or days.

To recover the site redundancy it will be necessary to swiftly establish a third data site (to replace the lost site). It is assumed that the additional costs of maintaining a third operational site for redundancy is too high. An alternative is to have a smaller third site that will be actively used to provide resiliency for archive data from a site failure. This solution requires almost 0.5 - 1x less storage to protect against site failure because erasure coded pools will be distributed among 3 sites rather than being replicated. This can be a space in an existing IT centre or data facility. The space and power capacity for this site can be such that we can extend this site on short notice in case of a complete site failure of one of the main active data site.

Notat

Til: Sigma2 / FOR-ANS prosjektleder
Fra: FOR-ANS referansegruppe
Dato: 27.10.2015
Gjelder: Rapport fra FOR-ANS teknisk arbeidsgruppe for nasjonal datainfrastruktur

Bakgrunn

FOR-ANS referansegruppe skal sikre kvaliteten av det arbeidet som gjøres i prosjektets arbeidsgrupper, og i tillegg ivareta universitetenes strategiske interesser slik som beskrevet i samarbeidsavtalen. Jmf. referansegruppens mandat.

Referansegruppe har vurdert rapport fra FOR-ANS arbeidsgruppe for nasjonal datainfrastruktur datert 19.10.2015. Det ble i tillegg avholdt møte mellom referansegruppen, FOR-ANS prosjektets ledelse og leder for arbeidsgruppen for datainfrastruktur på Gardermoen 21.10.2015 hvor man gjennomgikk rapporten. Referat fra dette møtet foreligger.

Dette notatet er skrevet på bakgrunn av tilsendt rapport, møtet på Gardermoen og diskusjoner i referansegruppen.

Referansegruppens vurdering av rapporten

Rapporten fra FOR-ANS arbeidsgruppe for nasjonal datainfrastruktur gjør rede for livsfaser av vitenskapelige data, og krav til nasjonal infrastruktur mtp. ivaretaking av viktige kriterier som sikkerhet, datatilgang, integrering med nasjonal HPC infrastruktur, og hvordan dette vil gi økt opplevd brukervennlighet av de nasjonale infrastrukturtjenestene samlet sett. En slik vinkling har resultert i et tungt fokus på nettverkshastighet i forhold til evaluering av forskjellige alternativ opp mot hverandre.

De kostnadsbesparelser som er oppgitt er basert på hardware kostnader alene, utifra hvor mye økt duplisering av data som er nødvendig i de forskjellige alternativene for å gi tilsvarende liten sannsynlighet for tap av data. Det tallfestede estimatet på 16-27% i kostnadsbesparelser er spesifikt beregnet ut fra en sammenligning av alternativ I og II.

Arbeidsgruppen konkluderer med å prioritere alternativ I med full samlokalisering av to HPC-anlegg og NorStore (med tilstrekkelig fysisk skille ifht sikring av påkrevd dataredundans). I tillegg til argumentene om nettverkshastighet og potensiale for kostnadsbesparelser, er antatt lavere kompleksitet mtp. installasjon og drift trukket fram som en fordel med dette alternativet.

Referansegruppen finner anbefalt prioriteringsrekkefølge fornuftig ifht de oppgitte parametere og faktorer som sammenligningen av alternativer er gjort ut ifra. Vi vil avslutningsvis likevel peke på at den gode analysen av livsfaser for vitenskapelige data ikke er tatt særlig med i analysen og vektlegging av alternativene. Innebygget funksjonalitet for synkronisering av mindre deler av prosjektdata til forskjellige siter (inkludert siter i den nasjonale infrastrukturen) ville gjort lagringstjenestene ytterligere tilgjengelig i et nasjonalt perspektiv. Alternativ I forutsetter også at visualisering og data-analytics dedikerte regneressurser blir samlokalisert med resten av den nasjonale infrastrukturen. Kravene for kort ventetid og elastisitet er ikke diskutert i rapporten.

Øvrige moment fra diskusjon

Alternativ IIIc inneholder en eksplisitt funksjonalitet for å synkronisere utvalgte prosjekt og/eller datasett til fysiske lokasjoner, som det ikke er tatt høyde for i øvrige alternativ (i opprinnelig tilsendt dokument). Dette vil være mest aktuelt for datasett i den såkalte varme fasen i prosjektdata livssyklus oversikten (ref “Data life cycle” figur i rapporten). En slik funksjonalitet vil gjøre det relativt enklere å skalere til flere fysisk distribuerte infrastrukturer som benytter seg av data fra de nasjonale tjenestene. Det vil også være mer fleksibelt enn alternativ I mht plassering av HPC maskinene og lagringssytemene (dvs den kan samlet sett være en bedre løsning).

I rapporten tillegges Sigma2 mye av ansvaret for å gjøre de nasjonale regne- og lagringsinfrastruktur ressursene tilgjengelig også i andre nasjonale infrastrukturer.

FOR-ANS

Housing-gruppe II

Anbefaling, plassering av
HPC-anlegg A1 & B1

Postadresse/Postal address
UNINETT Sigma AS
NO-7465 Trondheim

Besøksadresse/Visiting address
Abels gate 5
Trondheim, Norway

Foretaksnr./Company reg. no.: 887 625 352
Bankgiro/Bank account: 4200 03 53994

Epost/E-mail: info@uninett.no
Tel.: +47 73 55 79 00
Fax: +47 73 55 79 01

Innholdsfortegnelse

Arbeidsgruppens mandat

Arbeidsgruppen skal med utgangspunkt i resultat fra tidligere arbeid i FOR-ANS prosjektet, inkludert referansegruppens vurderinger, utarbeide en anbefaling for fysisk plassering av utstyr. Med utstyr menes Notur maskiner og lagringsutstyr for NorStore tjenestene. Med begrepet *maskin* menes et sett av mulig heterogene noder i ett eller flere rack, men i samme datasenter og med et felles, hurtig, internkommunikasjonsnettverk (Interconnect) og lagringssystem.

Alternativer som skal utredes er:

- Delt plassering av HPC-anlegg mellom to av de fire universitetene i partnerskapet
- Delt plassering av HPC-anlegg mellom ett av universitetene og en ekstern leverandør
- Alt HPC utstyr plassert hos en ekstern leverandør

Installasjonstider for HPC-anleggene A1 og B1

A1 installasjonsstart	1. november 2016
B1 installasjonsstart	1. november 2017
Norstore samlokalisering	1. november 2016 (initielt kun 7-8 rack)

Habilitet

I og med at UiT er aktuelt sted for plassering av A1, ble det enighet om at Roy Dragseth trakk seg fra gruppen. Kjetil Otter Olsen ble fortsatt med i gruppen, siden UiO ikke er en kandidat for A1.

Info fra universitetene / Markedsundersøkelse til eksterne

Vi sendte ut et svarskjema til alle de fire universitetene i partnerskapet, der vi ba om klargjørende informasjon rundt deres fysiske infrastruktur med tanke på å kunne huse A1 eller B1 i enten 2016 eller 2017. Her ba vi om forpliktende svar, der vi ba dem om å svare hva de allerede hadde på plass, eventuelt hvilke investeringer de måtte ta for å kunne huse anlegg. I tillegg ba vi om en anslått leiepris som Sigma2 må betale pr. mnd. (Det være seg areal, leie av aggregat, UPS-er, kjøleanlegg med mer)

Vi sendte også ut en markedsundersøkelse til følgende seks eksterne leverandører:

- Basefarm
- Green Mountain
- Bluefjord
- DigiPlex
- Lefdal Mine
- Atea (Dora-hallen i Trondheim)

Atea svarte at de ikke hadde kapasitet til å huse en så krevende installasjon. De meldte tilbake at de gjerne kunne huse Norstore isolert, men det ga vi tilbakemelding om at er uaktuelt.

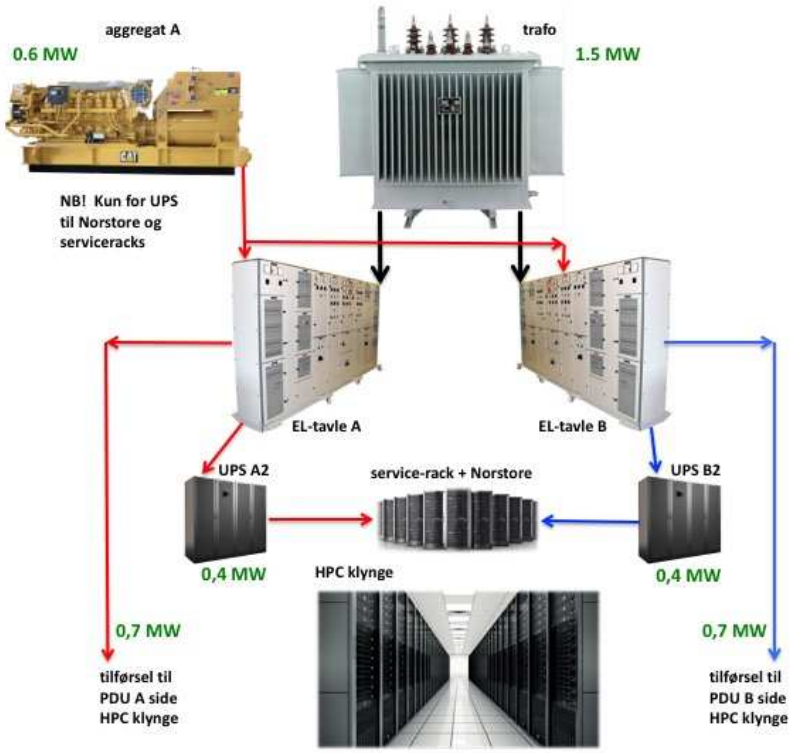
De eksterne tilbydernes besvarelser er mer omtalt i et vedlegg sist i rapporten.

Minstekrav til fysisk infrastruktur, A1 / B1 + Norstore

Vi ba også universitetene om tilbakemeldinger mht. hvor vi bør legge oss på når det gjelder krav til oppetid og redundans. Her var de fleste enige i at kost/nytte tilsier at man kan kjøre selve regneklyngen kun på bystrøm, uten beskyttelse av UPS og aggregat. Ett og annet strømbrudd tåler man i løpet av året, og det er pr. i dag ikke kritiske beregninger som ikke kan startes opp på nytt etter at strømmen er tilbake.
Når det gjelder serviceracks og Norstore er det bred enighet om at disse helt klart bør beskyttes av UPS og aggregat. Her er store disksystemer som kan ta skade av utilsiktede strømutfall.

Figuren under viser hva man trenger som et minimum til HPC-anleggene A1 og B1. Det kan variere med at man har 2 trafoer og en kombinert el-tavle for A og B –siden, men det er detaljer vi ikke skiller på i denne omgang.

Figur 1.



Besvarelsene fra universitetene

UiO:

- A1 ikke aktuelt, B1 kan potensielt være aktuelt
- Besvarelsen mangler investeringssummer
- Besvarelsen mangler leiepriser

NTNU

- B1 som hovedønske
- Kan ha mulighet for A1 + Norstore
- A1-klynge vil kjøres på bystrøm. Nye kabler fra hovedtavle må legges. Kostnadssituasjonen er uavklart og man vil mens Vilje fortsatt er i drift være nært opptil maksimalt effektuttak fra trafoen. Dette er en risiko.
- For B1 må det gjøres mindre endringer på strømtilførsel, kostnader ikke oppgitt
- Det kan leveres 12-20 grader kjølevann uten investeringer. Eventuelt kjølevann på 35 og 40 grader krever investeringer fra NTNUs side om de skal utnytte 60-65 grader utvann. NTNU må i så fall ta hele kostnaden selv uten at det har konsekvens for leiepris, siden dette er en gevinst for NTNU.
- Leie pr. m2 er veldig lav sammenlignet med UiT og UiB. Besvarelsen mangler mange kostnadselementer knyttet til fysisk infrastruktur (f.eks brannvern, serviceavtaler). Fremlegg av vanntilførsel er heller ikke priset. Det blir derfor veldig vanskelig å sammenligne deres besvarelse med andre universiteter som har en mer grundig besvarelse.
- Besvarelsen mangler underskrift fra den/de som er ansvarlig for økonomiske forpliktelser

UiB

- Ønsker enten A1 eller B1
- Detaljert besvarelse mht. investeringer og leiepriser
- God ledelsesforankring
- Må investere over 10 MNOK for å kunne ta imot A1 eller B1 (+Norstore). Investering nedskrives over 8 år. Investering med nedskriving over 4 vil gi høyere leipris.
- Bruker sjøvann til isvannsproduksjon. Det er kostnadseffektivt og mer driftsikkert.

UiT

- Ønsker enten A1 eller B1
- Best strømkapasitet av alle universitetene
- Har driftsklar løsning for varmegjenvinning hvis man kan bruke 40 graders kjølevann. Dette vil kreve ekstra luftkjøling som gir ekstra strømkostnad.
- Det er liten erfaring med å bruke 40 graders vann og det gir mindre sikkerhetsmarginer
- Tunge investeringer om man må ha lavere vanntemp 12 / 20 grader (10 MNOK/ 5 MNOK). Dette vil gi høyere leiepriser enn det som er oppgitt i tabellen.

Tabelloversikt – besvarelser

Tabellen samler informasjon fra både universitetene og de eksterne tilbyderne. Det ideelle ville være å kunne sammenligne ”epler mot epler”. Det er imidlertid ikke enkelt på alle punkter. Spesielt vanskelig er det når kostnadsbildet mellom universitetene skal sammenlignes mot de eksterne. Som eksempel er areal-leieprisene mellom NTNU og UiT vidt forskjellig. For enkelte universiteter har det også vært vanskelig å fremskaffe tallmateriale, og det at det ikke er definert leiepriser gjør det vanskelig for Sigma2 å ha kontroll på kostnadsbildet på lengre sikt.

	NTNU	UiO	UiB	UiT	Basefarm	Green M.	Bluefjord	DigiPlex	Lefdal Mine
Vi er klare til å huse anlegg i 2016 (A1) (Uten større investeringer)	X		X (inv)	X	X	X	X	X	X
Vi er klare til å huse anlegg i 2017 (B1) (Uten større investeringer)	X		X (inv)	X	X	X	X	X	X
EL-forsyning uten Norstore									
1.3 MW trafo	X			X	X	X		X	X
1.8 MW trafo				X	X	X		X	X
300 KW for serviceracks	X	X	X	X	X	X		X	X
1 MW UPS	X					X		X	X
2 x 100 KW UPS for serviceracks	X	X	X	X	X	X		X	X
EL-forsyning med Norstore									
1.7 MW trafo				X	X	X		X	X
2.1 MW trafo				X	X	X		X	X
600 KW aggregat for Norstore og serviceracks	X	X		X	X	X		X	X
2.1 MW aggregat (for alt)					X	X		X	X
1 MW UPS	X					X		X	X
2 x 400 KW UPS for Norstore og serviceracks	X	X		X	X	X		X	X
Areal og brannvern									
areal for 25 HPC rack	84000		83000	308333				239000	
areal for Norstore (50 m2)	50000		51250	154166					
SUM areal-leie pr.mnd	**134000**	**0**	**134250**	**462499**	**0**	**0**	**0**	**239000**	**0**
Kjøling									
12 grader	X		X		X	X	X	X	X
20 grader	X				X			X	
35 grader				X	X			X	
40 grader				X	X			X	
Norstore utstyr	X	X		X	X			X	
Brannsikring									
automatisk brannslukker / brannhemmende anlegg		X	X	X		X	X	X	X
Adkomst									
2.2m	X	X	X	X		X		X	X
2.4m			X	X (ikke DS1)		X		X	X
Leiepriser									
areal (fra SUM areal-leie pr.mnd i rad 26)	**134 000,00**	**0,00**	**325 250,00**	**462 499,00**	**0,00**	**0,00**	**0,00**		**0,00**
Full redundant løsning m/Norstore			110 000,00				328 530,00		
kun bystrøm på klynge, redundant, full støtte på Norstore og serviceracks					600 000,00	595 502,00			
Container-løsning, HPC									368 333,00
Container-løsning, serviceracks									77 500,00
Container-løsning, Norstore									63 333,00
Investeringer som er nødvendig for å tilfredsstille minstekravene			12 100 000,00	5 000 000,00					
hva det må investeres i			EL og areal	20 gr.vann					
SUM leiepriser (min. bystrøm på klynge + full redundans på SR+Norstore)	**134 000,00**	**uaktuell**	**435 250,00**	**462 499,00**	**600 000,00**	**595 502,00**	**328 530,00**		**509 166,00**

Spesielle utfordringer med fysisk infrastruktur ved intern plassering

Den største usikkerhetsfaktorer med housing på universitetene er knyttet til fleksibilitet på areal, strøm, kjøling og bruk av 40 graders innvann.

Dersom man benytter 40 grader, så gir det bedre muligheter for varmegjenvinning. 40 graders løsninger er imidlertid lite prøvd i markedet og har konsekvens for utstyret hvis man i full drift mister kontrollen på 40 graders vanntemperatur og den raskt stiger til 70-80 grader. For å redusere risiko er vår anbefaling at man i denne fasen ikke kan bruke 40 graders vann. Der bør tilrettelegges for maks 20 graders vann. Det betyr en investering for UiT på ca. 5 millioner.

Erfaringsmessig har man utvidet HPC-anleggene (F.eks. Abel, Stallo på ca. 25%) i løpet av deres levetid, og det er sannsynlig at det kommer til å fortsette. Man bør derfor ha fleksibilitet for fremtidige utvidelser med hensyn på areal, strøm og kjøling.

Ved fremtidig bytte av maskiner vil det i installasjonsperioden kreves ekstra areal, strøm og kjøling. Dette er en utfordring på samtlige interne alternativ.

Generelt er driftssikkerheten høyere hos eksterne tilbydere fordi de har dublering på fysisk infrastruktur.

Vurderinger A1 + Norstore

”Ekstern” i tabellen er en felleskolonne for de antatt tre mest relevante eksterne tilbyderne.

Krav	NTNU	UiB	UiT	Ekstern
Strøm	Hvis A1-klynge kan kjøre på uprioritert strøm så kan Norstore og serviceracks kjøre på eksisterende UPS sammen med Vilje inntil utfasing Trafo er på 1.6MW. Kostnader ikke oppgitt	Krever betydelige investeringer for å kunne klare A1 og Norstore (>10 MNOK)	Klar for A1 og Norstore	Uproblematisk
Kjøling	12-20 grader innvann er OK.	Har sjøvann til isvanns-produksjon.	Må investere ca. 5 MNOK for å klare vanntemp. på 20 grader.	Uproblematisk
Areal	Har plass til HPC og Norstore. Dog må kabling og rør legges. Kostnader ikke oppgitt	En del tilpasninger av tilstøtende areal trengs. Usikkerhet knyttet til kostnad.	Har plass til HPC og Norstore	Uproblematisk

Brannvern	Må installeres, mangler i dag. Kostnader ikke oppgitt.	Finnes i eksisterende areal, men må utvides	Finnes i aktuelle lokaler	Finnes

Vurderinger B1 + Norstore

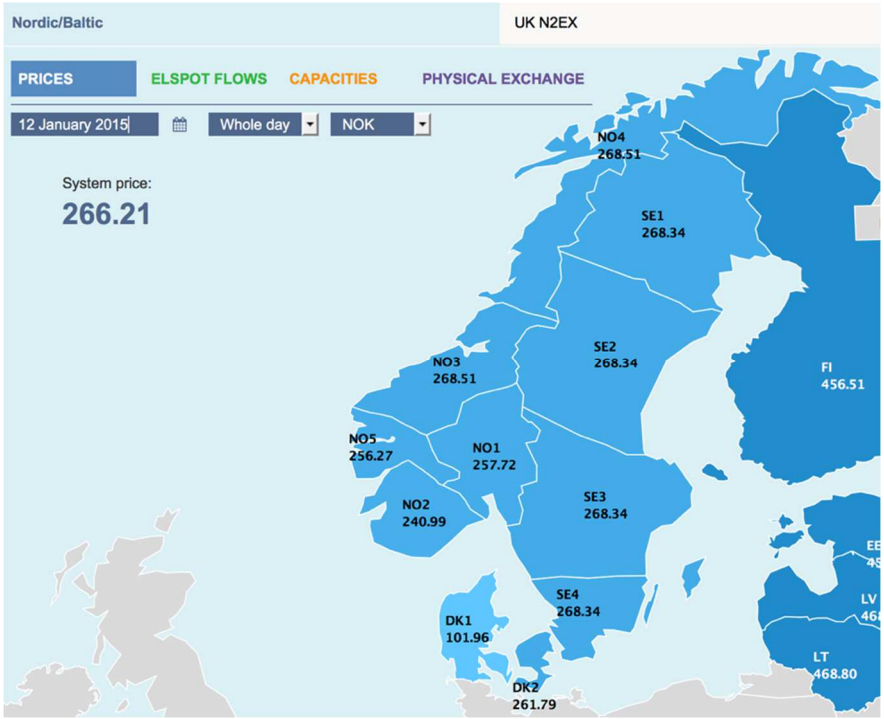
Krav	NTNU	UiB	UiO	UiT	Ekstern
Strøm	Tilstrekkelig Tilpasninger nødvendig, kostnader ikke oppgitt	Krever betydelige investeringer (>10 MNOK)	Mangelfull avklaring pr. i dag.	Tilstrekkelig	Uproblematisk
Kjøling	12-20 grader innvann er OK.	Har sjøvann til isvanns-produksjon.	Mangelfulle avklaringer pr. i dag.	Må investere ca. 5 MNOK for å klare vanntemp. på 20 grader.	Uproblematisk
Areal	Areal tilgjengelig, men kabling og rør må tilpasses. Kostnader ikke oppgitt	En del tilpasninger av tilstøtende areal trengs.	Ja, men leie-priser mangler	Areal tilgjengelig	Uproblematisk
Brannvern	Må installeres, mangler i dag. Kostnader ikke oppgitt	Finnes i eksisterende areal, men må utvides. Kostnader ikke oppgitt	Finnes i aktuelle lokaler	Finnes i aktuelle lokaler	Finnes

Kriterier for valget av lokasjoner for A1 + Norstore og B1

Arbeidsgruppen har vektlagt følgende:

- Forutsigbart kostnadsbilde slik at Sigma2 ikke får uforutsette utgifter underveis i HPC-anleggets levetid
- Risiko (tid, kostnad) ved utbygginger og endringer av fysisk infrastruktur
- Fysiske infrastruktur-løsninger som lett kan tilrettelegges/oppgraderes for HPC-anlegg i endring
- Skalerbarhet, parallellkjøring mht. prosessering ved senere utskifting av anlegg slik at 2 anlegg kan kjøres samtidig i perioder
- Unngå bindinger utover fire år, fordi man ikke vet hva som da blir kravene når nytt anlegg skal på plass om fire år.
- Solid, redundant fiberadgang til Uninetts forskningsnett, som sikrer nåværende og fremtidige behov
- Brannvern, fysisk sikring, sikring generelt
- Strømprisene (spotmarkedet) varierer mye avhengig av hvor i landet man befinner seg
  Samtidig vil de største eksterne tilbyderne nærme seg 5 MW-grensen for å få lavere el-avgift.
  Disse tilbyderne har bekreftet at den reduksjonen fullt og helt går til kundene.

Fra Nord Pool Spot kan man studere spotpris-markedets endringer gjennom året.
Eksemplet er fra den 12 januar, og flytter vi komma en posisjon til venstre så ser vi en systempris på 26.6 øre per kWh. Man ser også at region N05, N02 og N01 ligger noe lavere enn Midt-Norge og Nord-Norge (N03/N04). Dette varierer en del gjennom året, men trenden er at Midt-Norge ligger høyere i snitt enn de sydligere spotpris-sonene.
Besparelser på årlige strømutgifter ved å ha HPC-anlegg på størrelse 1 MW i rett sone (variasjon på ca. 10%) vil gi besparelser på ca. 260 KNOK/år. Det forsterkes ytterligere om man har anlegget hos en leverandør med lavere el-avgift. Det gir besparelser på ca. 1 MNOK/år.



Gevinst ved samlokalisering av Norstore

Det er kun en ekstern plassering som gir mulighet for samlokalisering av Norstore. Vi anslår at dette kan gi innkjøpsbesparelser på omlag 16-27% av det lagringskapasiten koster i forhold til en delt plassering. Ved en investering på 30 MNOK kan det gi besparelse på ca. 6 MNOK.

Av operasjonelle kostnadsbesparelser nevnes reduserte nettverkskostnader og noe enklere drift.

Oppsummering av risikofaktorer

Nr	* Risikofaktor	*S	* Konsekvens	*K	Kritikalitet	* Tiltak
1	Mangler tall for leiepriser fra NTNU (A1)	4	Uforutsette kostnader	3	12	Lite aktuelt
2	NTNU nært opptil maksimalt effektuttak fra trafoen (A1)	4	Mulige driftsproblemer	3	12	Lite aktuelt
3	B1 på NTNU krever tilpasninger, kostnader ikke oppgitt. Driftskostnader på fysisk infrastruktur mangler	2	Uforutsette kostnader	2	4	Krever detaljert kostnad -og tidsplan
4	Store investeringer UIB, mest for strøm men også på arealer for A1	4	Uforutsette kostnader og mulig forsinkelse pga liten tid	3	12	Lite aktuelt
5	Store investeringer UIB, mest for strøm men også på arealer for B1	3	Uforutsette kostnader	2	6	Krever tidsplan og avklaring rundt kortere nedskrivingstid
6	UiT: Liten erfaring med 40 graders vann	4	Små sikkerhetsmarginer, mulige driftsproblem	3	12	Blir OK om det investeres for 20 graders vann
7	Mangel på fleksibilitet internt ved framtidige utvidelser	3	Ekstra kostnader og/eller begrensninger på utvidelser	3	9	Lite aktuelt
8	Mindre erfaring med ekstern leverandører (Kun Gardar). Mulige samarbeidsproblemer mellom driftsgruppe og housingleverandør	2	Mulige driftsforstyrrelser	3	6	Ta ekstra hensyn i kontrakt og i driftsforberedelser
9	Begrenset housing erfaring for HPC hos ekstrene leverandører	4	Mulige driftsforstyrrelser	2	8	Ta ekstra hensyn i kontrakt og i driftsforberedelser

Anbefaling

De fleste universitetene er enig i at krav til fysisk infrastrukturkrav er i henhold til figur 1 tidligere i rapporten. Det er to universiteter som dermed skiller seg ut ved å ha minst behov for investeringer før de kan ta imot A1 + Norstore eller B1 + Norstore, er NTNU og UiT. Sett fra Sigma2 er det knyttet for stor usikkerhet knyttet til investeringsbehov ved NTNU pga. manglende informasjon.

Når det gjelder UiT må de gjøre betydelige investeringer for å endre temperaturen på innvann fra 40 til 20 grader, noe vi mener det er behov for.

UiB har et klart kostnadsbilde, men investeringer er svært store (10 MNOK). Er det riktig å gjøre en så stor investering for en kortere periode?

Arbeidsgruppen registrerer også at når universitetene får i oppgave å definere investeringer samt leiepriser på areal og utstyr er det mindre avstand mellom tilbudene fra universitetene og de eksterne tilbyderne. Universitetene har ikke synliggjort om de har reserver til å dekke utskifting av batteribanker, eventuelle større feil på aggregater eller UPS-er i avtaleperioden. Serviceavtaler med leverandører er heller ikke synliggjort i kostnadsbildet. Alt dette inngår i prisen fra de eksterne leverandørene.

Alternativer som skulle utredes var:

- Delt plassering av HPC anlegg mellom to av de fire universitetene i partnerskapet
- Delt plassering av HPC anlegg mellom ett av universitetene og en ekstern leverandør
- Alt HPC utstyr plassert hos en ekstern leverandør

Oppsummering av økonomi

Denne oppsummeringen gjelder for A1 + Norstore. Prisnivået på intern plassering ligger i størrelsesorden 435-462 KNOK/mnd. (NTNU pris er lavere men har stor usikkerhet). Dette er for lavt med de utfordringer vi har påpekt i dette notatet (kjøling ved UiT, (avskrivningstid ved UiB)). Årlig fast kostnad blir ca. 5.2-5.5 MNOK/år.

Ekstern plassering har et prisnivå på ca. 500-600 KNOK/mnd. (eks. MVA). Årlig kostnad blir ca. 7.5- 9 MNOK/år (inkl MVA). Det er forventet at pris kan bli lavere ved en anbudskonkurranse. Investering i datanett er ca. 2.5 MNOK. Mulige andre innsparelser ved ekstern plassering er redusert el-avgift på 12 øre/kWh, noe som utgjør ca. 1 MNOK/år for 1MW anlegg, dvs ca. 2 MNOK/år for samlokalisert anlegg.

Samlokalisering vil potensielt gi investeringsbesparelser på ca. 6 MNOK for NorStore i Fase1 og tilsvarende i Fase2.

Vår vurdering er at pris på løpende utgifter er nesten de samme mellom intern og ekstern plassering. Besparelser knyttet til investeringer vil favorisere ekstern plassering.

Oppsummering av årlige kostnader er gjort i tabellen nedenfor:

Plassering	Leiekostnad MNOK/år	Innsparing strøm MNOK/år	Innsparing data lagring MNOK/år	Ekstra kostnad datakomm. MNOK/år	Årlig kostnad MNOK/år
Intern	2 x 6 = 12	0	0	0	12
Ekstern	2 x 8.5 = 17	2 (12 øre/kWh)	6 / 4 = 1.5	2.5 / 4 = 0.6	14.1

Tilleggsvurderinger:

- Usikkerhet knyttet til leiepris fra universitetene, sannsynligvis er den høyere enn oppgitt
- Intern plassering mangler fleksibilitet for utvidelser, noe som kan bety økte kostnader
- Ved ekstern plassering er det sannsynlig at vi kan oppnå lavere pris for 2 anlegg

Arbeidsgruppens innstilling er, i foreslått rekkefølge:

1. **Alt utstyr plasseres hos ekstern leverandør (A1 + Norstore + B1)**

2. **A1 plasseres ved UiT, forutsatt at det investeres 5 MNOK i kjøling. Når det gjelder B1 må det gjøres en ny vurdering om beste plassering er ved UiB eller NTNU fordi NTNU ikke har levert tilstrekkelig underlag, men synes å ha minst behov for investeringer.**

3. **Delt plassering mellom ekstern leverandør og UiT.**
   Ekstern leverandør får A1, B1 plasseres ved UIT

Kommentarer til innstillingen

Førstevalget er altså alt samlet hos en ekstern leverandør. Arbeidsgruppen påpeker følgende fordeler med nettopp det:

- Forutsigbarhet, både teknologisk og økonomisk.
- Forenklet kommunikasjon med Norstore.
- Det er anslått at man sparer 6 MNOK ved å ha Norstore maskinromsnært til både A1 og B1.
- Skalerbarhet på strøm, kjøling og areal som langt overgår universitetene
- Om kravene til oppetid senere blir strengere kan aggregat og UPS-kraft lett implementeres også for selve regneklyngen
- Utfasing og innfasing av nye HPC-anlegg er betydelig enklere, parallellkjøring inkludert.
- Om man velger en av de største eksterne leverandørene kan redusert el-avgift bety lavere strømutgifter.
- Flere av de eksterne leverandørene er i spot-soner som over tid har lavere strømpriser enn for eks. Midt-Norge. Dette kan beløpe seg til flere MNOK ved de effektbehovene vi snakker om her.

Arbeidsgruppen har bestått av:

Helge Stranden, UNINETT
Kjetil Otter Olsen, UiO

Vedlegg 1

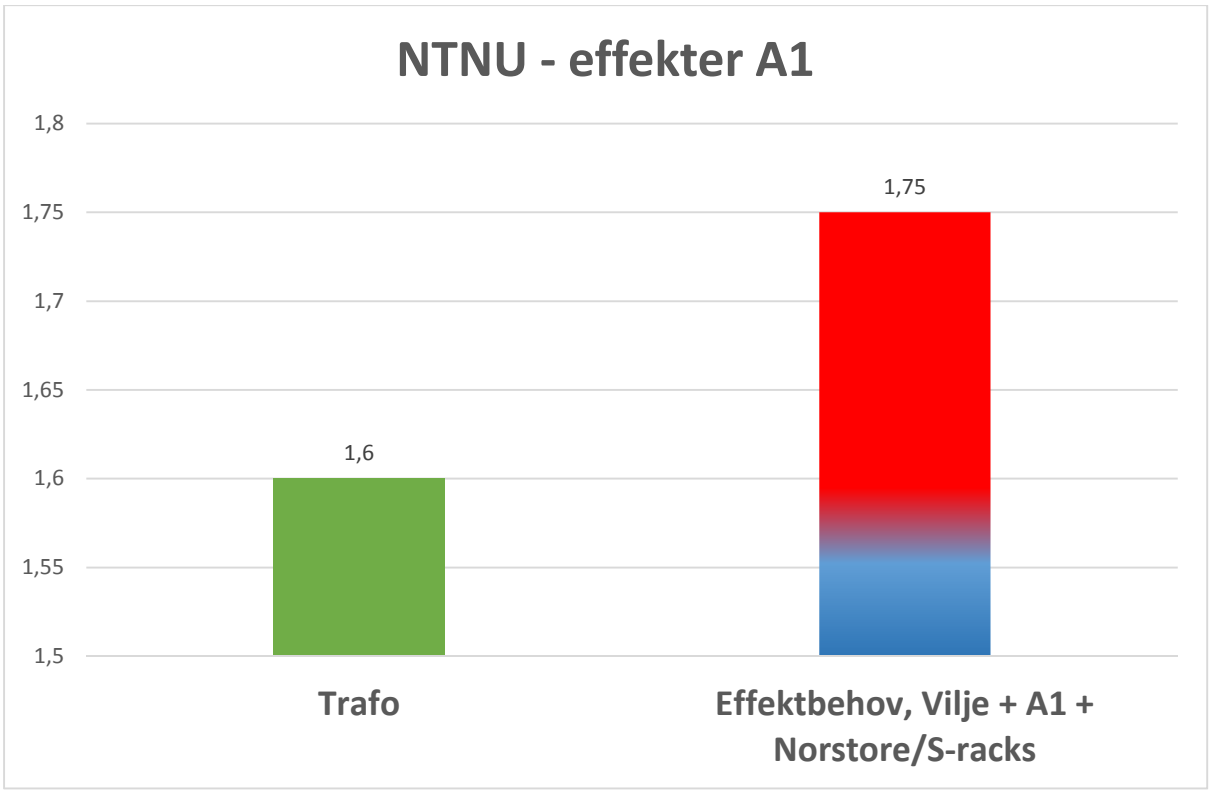
Effekt-scenario med NTNU som eksempel

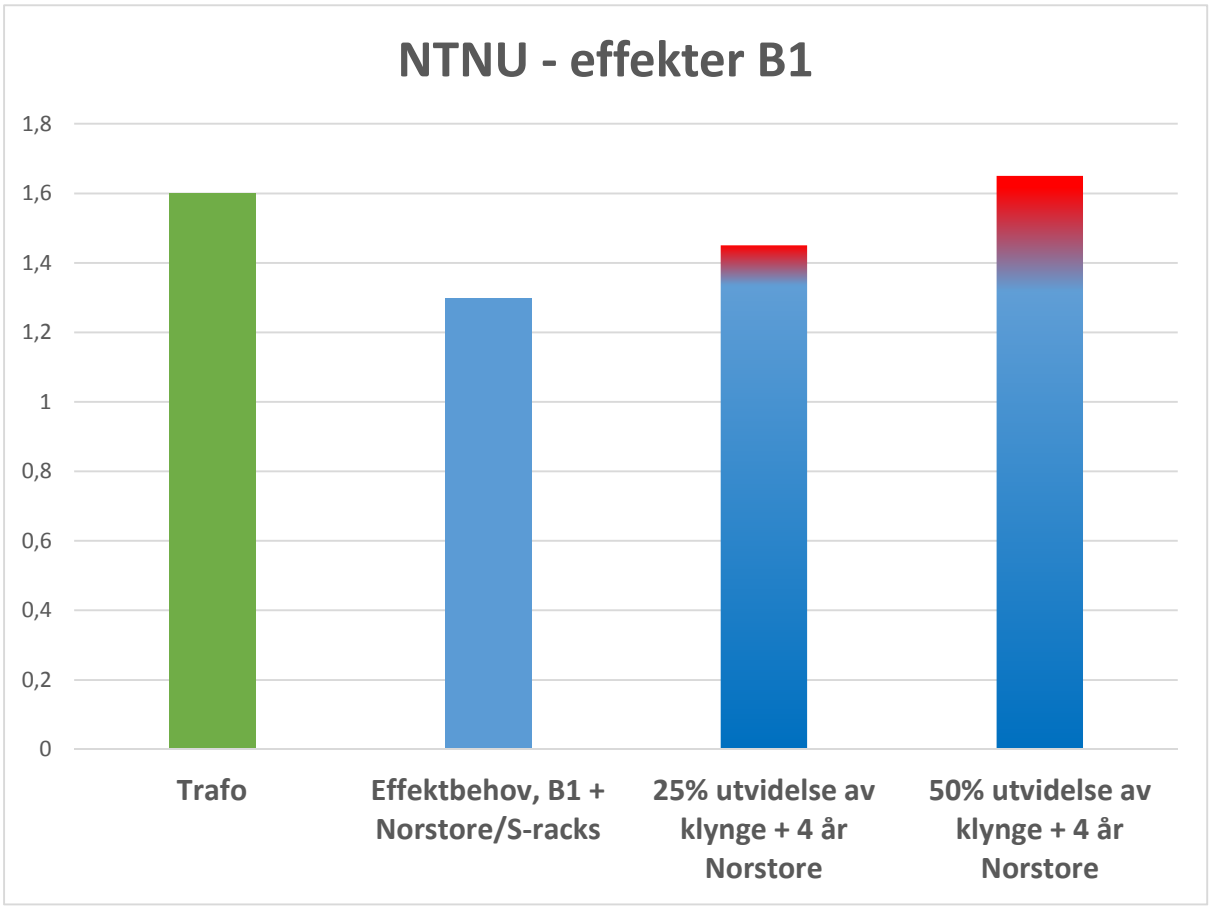
I figurene nedenfor ser man hvorfor NTNU er en mer naturlig B1 kandidat enn A1 kandidat. Man ser (figur 1) at man har en utilstrekkelig trafo-kapasitet om man kjører Vilje samtidig med A1 + Norstore & serviceracks.

Som B1 kandidat (figur 2), er det akseptable marginer så lenge man ikke har behov for utvidelser. Her har vi tatt med en økning av Norstore, som naturlig nok kommer etter hvert som diskbehovet øker med årene. Ved en potensiell økning på 25% eller 50% av klyngekapasiteten i 4 års perioden er man oppe i effekter som kan bli kritiske.

Det som det ikke er tatt høyde for her, er hvis det kommer krav/behov fra brukergrupper om at oppetiden må økes pga økende antall strømbrudd. I for eks. Tromsø har man sett en klar tendens de siste årene der antall strømbrudd pr. år har økt. Skulle slike krav bli tatt til følge i avtaleperioden på 4 år, så må UPS for regneklyngen også vurderes, og da snakker vi betydelige investeringer både på UPS, trafo og evt. EL-tavler samt nedetid i omleggingsfasen. I tillegg kan det komme behov for ytterligere kjølekapasitet som en følge av en utvidelse.

Sigma2 ønsker i minimere alle slike risikoer ved å plassere fremtidige anlegg i en lokasjon der det alltid vil være betydelig overkapasitet tilgjengelig slik at man kan gi forskningsmiljøene all den prosesseringskraft de til enhver tid har behov for.


NTNU - effekter A1
1,8
1,75
1,7
1,65
1,6
1,55
1,5
1,6
1,75
Trafo
Effektbehov, Vilje + A1 + Norstore/S-racks


NTNU - effekter B1
1,8
1,6
1,4
1,2
1
0,8
0,6
0,4
0,2
0
Trafo
Effektbehov, B1 + Norstore/S-racks
25% utvidelse av klynge + 4 år Norstore
50% utvidelse av klynge + 4 år Norstore

Vedlegg 2

Besvarelser fra eksterne tilbydere

Vi sendte også ut en ”Markedsundersøkelse” til de eksterne leverandørene. Målet var å få et bilde som viste om de var reelle alternativer til universitetene når det kom til plassering av våre HPC-anlegg og Norstore. Den forrige arbeidsgruppen ga utrykk for nettopp det, og besvarelsen vi nå har mottatt fra de fleste av de eksterne leverandørene har styrket det inntrykket.

Om det blir aktuelt med ekstern leverandør for noen av HPC anleggene vil man kjøre en grundig anbudsrunde. Da vil leverandørene få anledning til å spisse sine priser ytterligere utover de anslagene de har kommet med her. De mest solide besvarelsene i denne omgang har kommet fra Green Mountain, Basefarm og Lefdal Mine.

For alle de eksterne vil det komme en investering på ca. 2.5 MNOK i nettverksutstyr.

Basefarm

- Full fleksibilitet, kan levere alle nivåer av redundans
- Prismessig litt over gjennomsnittet blant de eksterne
- Alle store fiberleverandører er representert i deres lokaler
- Varmegjenvinning via Akershus Energi
- Kan enkelt sette nye anlegg i drift mens eksisterende tas ut

Green Mountain

- Full fleksibilitet, kan levere alle nivåer av redundans
- Prismessig gjennomsnittlig blant de eksterne
- Alle store fiberleverandører er representert i deres lokaler
- Ingen varmegjenvinning (pga. deres isolerte plassering)
- Kan enkelt sette nye anlegg i drift mens eksisterende tas ut
- Har definert etableringskost, men dette bør være mulig å få redusert eller forhandlet bort.

Bluefjord

- Besvarelsen er tynn, og det tyder på at de kun har noen få generelle housing-kunder. Blir etter all sannsynlighet ”for lett” til å bli oppfattet som en seriøs partner for oss.

DigiPlex

- De 800 m2 som vi evt. vil må få deler av er ennå ikke bygget. De må ha 6 måneders byggetid etter underskrevet kontrakt. Det kan bli utfordrende for Norstore, som må på plass 1. juni 2016.
- Generelt sett har de bra fleksibilitet, men vil dog kun tilby N+1, altså UPS/aggregat da deres standard er slik.
- Prismodellen de har skissert har 1500w/m2 og er nok ikke aktuell for oss. De varsler at ved et anbud blir denne tilpasset våre effekter. Vi har derfor ikke avklarende priser fra DigiPlex i denne omgang.
- Alle store fiberleverandører er representert i deres lokaler
- Har mulighet for varmegjenvinning

Lefdal Mine

- Container-basert konsept i Lefdal gruve. (tre store containere til HPC)
- Prismessig litt lavere enn gjennomsnittet blant de eksterne
- Egne containere for henholdsvis serviceracks og Norstore. Bør kunne kombineres
- Stor grad av fleksibilitet mht. fremtidige fornyelser av anlegg
- Interconnect-mulighetene (ettersom tre HPC-containere) må avklares
- Ingen varmegjenvinning er realistisk

Notat

Til: Sigma2 / FOR-ANS prosjektleder
Fra: FOR-ANS referansegruppe
Dato: 28.10.2015

Gjelder: Rapport fra FOR-ANS Housing-gruppe II (Anbefaling, plassering av HPC-anlegg A1 & B1)

Bakgrunn

FOR-ANS referansegruppe skal sikre kvaliteten av det arbeidet som gjøres i prosjektets arbeidsgrupper, og i tillegg ivareta universitetenes strategiske interesser slik som beskrevet i samarbeidsavtalen. Jmf. referansegruppens mandat.

Referansegruppe har vurdert rapport fra Housing-gruppe II tilsendt 19.10.2015. Det ble i tillegg avholdt møte mellom referansegruppen og FOR-ANS prosjektets ledelse på Gardermoen 21.10.2015 hvor man gjennomgikk rapporten. Referat fra dette møtet foreligger.

Dette notatet er skrevet på bakgrunn av tilsendt rapport, møtet på Gardermoen og diskusjoner i referansegruppen.

Referansegruppens vurdering av rapporten

Rapporten fra FOR-ANS Housing-gruppe II vurderer og anbefaler plassering av HPC-anleggene A1, B1 og NorStore. Arbeidet bygger på tidligere leveranser fra Housing-gruppe I, Notur teknisk arbeidsgruppe og arbeidsgruppen for datainfrastruktur, sammen med innsendt materiale fra universitetene og eksterne tilbydere av datasentertjenester.

Den samlede vurdering og prioritering av alternativer til arbeidsgruppen hviler tungt på følgende observasjon: kostnadsbildet for intern og ekstern housing ser ut til å være tilnærmet likt når universitetene synliggjør alle sine reelle kostnader. Vi vet at universitetene fikk svært liten tid på denne oppgaven, så det er vanskelig å sammenligne kostnadskomponenter på detaljnivå i de fire besvarelsene, og dette gjør at det er en risiko for at grundigere utredninger vil avdekke flere eller endre kostnadselementer.

Priser innhentet fra eksterne leverandører er kun veiledende (uforpliktende “listepris”) og gjelder kun oppdrag for ett anlegg (A1). Man kan forvente en bedring i kostnadsbildet ved å gjennomføre en anbudsprosess med flere tilbydere om housing av A1+NorStore+B1 (2MW+). Referansegruppen anbefaler derfor at det snarest vurderes iverksatt en anbudsprosess hvor man forbeholder rett til å forkaste alle tilbudene, forutsatt at installasjon av A1 ikke vil forsinkes av dette arbeidet. Dette vil gi et enda bedre sammenligningsgrunnlag for å kvalitetssikre det underliggende premisset om at intern og ekstern housing er ekvivalente kostnadsmessig. Hvis det er større forskjeller kostnadsmessig i favør universitetene, vil alternativet prioritert som nr. 2 av arbeidsgruppen stille sterkere.

Rapporten peker på klare fordeler, redusert risiko og fleksibilitet ved lokalisering i eksternt datasenter. Gitt at kostnadsbildet er omtrent det samme mellom eksterne tilbydere og universitetene, gir disse momentene føringer som resulterer i den endelige prioriteringsrekkefølgen med all infrastruktur samlet ved ett eksternt datasenter som prioritet 1.

Referansegruppen ser ikke hvordan alternativ med prioritet 3 (delt plassering mellom ett universitet og eksternt datasenter) skal kunne konkurrere kostnadsmessig eller hente ut alle fordeler tillagt alternativ 1 og 2. Referansegruppen anbefaler at dette alternativet legges tilside i det videre arbeid.

Det må også påpekes at rapporten fra arbeidsgruppen for data-infrastruktur gir høyest evaluering til en løsning som forutsetter alternativ 1.

Kommentarer og presisering til enkeltdeler av rapporten

Til tross for at vi anbefaler en kvalitetssikring av kostnadsnivået ved ekstern housing i vår overordnede vurdering, ser vi at det presenterte kostnadsbildet pr i dag representerer en vesentlig forskjell, som forpliktende tilbud fra anbudsrunde vil måtte eliminere før ekstern og intern housing kan sies å være ekvivalente kostnadsmessig.

Investeringer ved universitetene skal avskrives over en gitt periode, og ut fra dette beregnes leiekostnader. Når levetiden for enkelte HPC- og lagrings-anlegg er i størrelsesorden 4-5 år, finansiering fra NFR uten langsiktig horisont osv., reiser det spørsmål om hvilken avskrivingperiode som skal benyttes for investering i datasenterinfrastruktur. Sistnevnte avskrivningsperiode bør uansett settes etter samme prinsipp slik at kostnadene for universitetene kan sammenlignes. Dette aspektet favoriserer nå eksterne tilbydere, hvor leiekontrakter kan utformes med vilkårlig lengde. Eksempelvis viser NTNU sin leiekostnad med avskrivning over 20 år et kostnadsbilde på ca ⅓ av ekstern housing, mens UiB og UiT med avskrivning over 8 år lander på ca ¾ av ekstern housing i leiepris. For sammenligning mellom intern eller ekstern plassering er det viktig at prinsippene harmoniseres.

Generelt virker utvalget og vektingen av flere parametere lite gjennomarbeidet i forskjellige sammenhenger, noe som fører til et manglende oversiktsbilde totalt. Dette blir ikke avhjulpet av at betingelsene for hva universitetene skal gi tilbakemelding på har endret seg underveis i prosessen.

Et viktig aspekt som ikke er behandlet tilstrekkelig i rapporten er mulighetene for varmegjenvinning ut i fra et miljøperspektiv. Alle de tre universitetene som har sagt seg interessert i å bidra med datasenterkapasitet har synliggjort at de utnytter eller har ambisjoner om å utnytte restvarmen som kommer fra sine datasentre. Dette vil være være viktig for flere fagmiljø hvor infrastrukturens miljøprofil er et vesentlig moment.

Under diskusjon i referansegruppen rundt utfordringer med varmtvannskjøling (40 grader), var risiko for overoppheting og få tilbydere av kompatible HPC-anlegg de to sentrale temaene. Vi registrerer at erfaringene fra ett av Europas og verdens største regneanlegg, SuperMUC (https://en.wikipedia.org/wiki/SuperMUC), som har vært i operasjonell drift siden 2012 og benytter en tilsvarende teknologi for kjølingen er gode. Men det er vanskelig å vurderer risiko ut i fra dette enkelttilfellet. I tillegg er dette en referanseinstallasjon for én av de begrensede antall HPC-tilbyderene med slik teknologi i dag.

En risiko som referansegruppen mener ikke er tilstrekkelig vurdert i rapporten er rundt anbudsprosess nødvendig for ekstern housing. Vi har ikke i dag faktisk erfaring med denne typen anbud, og en slik prosess må gjøres riktig for at vi skal unngå forsinkelser som kan oppstå i tilfelle en tapende part bestrider en kontraktbeslutning.

Betraktninger ifht universitetenes egne strategier

Universitetet i Bergen

I UiB sine nedfelte strategier er klimaforskning, marin klynge i Bergen og bioinformatikk fagfeltene særlig prioritert. UiB har samarbeidet tett med næringslivet og det offentlige for å etablere infrastruktur for disse miljøene, med betydelige investeringer og et framtidsrettet fokus på grønn-IT. Dette gjenspeiles i det unike anlegget for bruk av sjøvann til kjøling av data-infrastruktur inkl HPC, hvor restvarme brukes direkte til oppvarming av annen forskningsinfrastruktur innen marin forskning. Dette gir både en miljø- og forskningsgevinst som er vel så betydningsfull som den rent økonomiske gevinsten man oppnår ved eventuell gjenbruk av varme til generell oppvarming . Alle tre fagfeltene er også tunge brukere innen tungregning og kapasitetslagring idag, slik at vekselvirkning mellom infrastruktur og forskning er reell på mange plan. For UiB er det særlig viktig at nasjonal infrastruktur benyttet til klimaforskning har et miljøfokus både på kjøling og øvrig el-forsyning, og gjenspeiles i eksisterende etablert infrastruktur lokalt. Dette er en uttrykt prioritering fra det internasjonalt anerkjente klimamiljøet i Bergen, og en betydelig faktor for valg av regneanlegg til klimaforskning internasjonalt.

Utover forankring i UiB sin strategi via faglige miljøer, er HPC og datalagring en del av helheten det pågående arbeidet for UiB sin digitaliseringsstrategi tar for seg. UiB har synliggjort i sitt innspill til arbeidsgruppen en tydelig lederforankring mtp vilje til å gjennomføre ytterligere investeringer, som er relativt mindre enn allerede investerte midler, for å holde fast ved eksisterende strategisk prioritering av et HPC-anlegg lokalisert i Bergen.

I arbeidsgruppens rapport er det oppgitt some en hovedrisikofaktor ved UiB som vertsinstitusjon for et HPC-anlegg, at det må investeres opp mot 12 mill NOK (ved A1, vesentlig mindre ved B1) og at dette må medføre en betydelig ombygning med en viss sannsynlighet for forsinkelser. Det er ikke tatt tilstrekkelig høyde for at nødvendige arealombygninger er av mindre omfang, dette er eksisterende datahaller som har vært midlertidig brukt til annet formål. For å underbygge vår påstand om lav sannsynlighet for forsinkelse ifht installering av A1, har UiB iverksatt en egen prosjektering av dette hos eiendomsavdelingen for arealombygningsarbeidet samt opprustning av el-forsyningssiden berammet til 5 mnd total ombygningstid, som viser hvilket arbeid som må utføres og en vurdering av gjennomførbarhet innenfor denne tidsrammen. For B1 er tidshorisonten lenger fra beslutning til installasjon, og sannsynligheten for forsinkelse vil være kraftig redusert.

Gjennom kvalitetssikring av plasseringsrapportens bruk av underlagsmaterialet UiB har sendt inn, har UiB sett det nødvendig å presisere flere momenter direkte med Sigma2 og arbeidsgruppen i egen kommunikasjon. Tidspress i FOR-ANS prosessen bør også vurderes som en risikofaktor ifht til om beslutningsgrunnlaget er godt nok kvalitetssikret.

Universitetet i Tromsø

UiT har over år bygd en sterk kompetanse for effektiv datasenterdrift og har med sitt nye datasenter for HPC ambisjoner om å gjenvinne opptil 80% av restvarmen som genereres til oppvarming av undervisningsbygg på campus. Dette vil gi store besparelser i strømutgiftene til drift av regneanlegg, med et strømforbruk på 1 MW og en strømpris på 70 øre per KWh vil 80% varmegjenvinning gi en besparelse på 4.9 MNOK per år. Dette vil gi UiT og norsk tungregning internasjonal oppmerksomhet som verdens mest miljøvennlige regnetjeneste. Denne gevinsten kan bare oppnås hvis UiT blir tildelt vertskapet for ett av de nasjonale regneanleggene slik at volumet av restvarme blir stort nok til å gjøre en kostnadseffektiv varmegjenvinning. UiT har i sitt tilbud om å være vertskap for det neste regneanlegget tydeliggjort at en er villig til å la en eventuell kostnadsgevinst komme hele konsortiet til gode, men at det er vanskelig å gi konkrete tall på dette før man vet mer om hvilken type regneanlegg som blir installert.

Siden markedsundersøkelsen ikke ser ut til å ha synliggjort betydelige besparelser ved å leie datasenterkapasitet i det kommersielle markedet er UiT bekymret for at størrelsen på en leiekontrakt dikterer at man må ta en full anbudsutlysning etter reglene for offentlig innkjøp og at det vil forsinke anbudsprosessen av et nytt regneanlegg med minst 3 måneder da man ikke kan

anskaffe regneanlegg av denne størrelsen uten å vite detaljene om de fysiske parameterne i datasenteret det skal bli plassert i. UiT har allerede ledige arealer og kapasitet i dag og ved å benytte denne kapasiteten vil man fjerne en stor del av risikoen for forsinkelsen i etableringen av et nytt regneanlegg som må være på plass innen utgangen av 2016.

UiT er også bekymret for at krav om fysisk samlokalisering fjerner fleksibiliteten for fremtidig eInfrastruktur, inkludert store data produsenter. Det ser heller ikke ut som det er tenkt en exit strategi fra eksterne tilbydere. Lagring og regneinfrastruktur har forskjellig levetid, så det vil være vanskelig å bytte ekstern leverandør hvis man har utstyr som skal være i produksjon over flere år og som må være fysisk samlokalisert. Vi tror også at det er mulig å flytte store datasett over nettverket i Norge, og at dette vil være nødvendig for data fra nye store instrumenter som EISCAT 3D eller moderne forskningsskip.

Ved å utnytte eksisterende datasenterkapasitet ved universitetene så får man beholdt kompetanse og ressurser i sektoren. UiT ser det også som vanskelig å kunne tilby housing for fremtidige anlegg hvis vi ikke får A1, siden vi ikke kan ha areal og reservestrøm stående uten å vite at vi i fremtiden vil få en maskin. Hvis det derfor ikke er store forskjeller i prisbildet mellom en kommersiell leverandør og å benytte universitetenes eksisterende infrastruktur (uten at varmegjenvinning er tatt hensyn til) bør man gå videre med alternativ 2.

Universitetet i Oslo

Rapporten peker på viktige fordeler med plassering i eksternt datasenter, blant annet et helt tydelig kostnadsbilde, redusert risiko og plassering av ansvar. Da behovene for regnekraft og lagring antagelig bare vil øke i fremtiden er det en stor fordel med fleksibiliteten (elastisiteten) en plassering i eksternt datasenter gir. Det er tilsvarende en fordel å kunne gjøre oppgraderinger og utskiftinger sømløst. En plassering av infrastrukturen i eksternt datasenter vil medføre bedre balanse (likeverdighet) mellom partene i driftssenteret og kunne medføre bedre slutttjenester for brukerne. Samtidig er det å påpeke at vi ikke har erfaring med eksterne datasentre, og at dette derfor må anses som en risiko.

Å få mest mulig forskning over totalkostnader må være det overordnede målet. Det er i UiOs syn ikke tilstrekkelig dokumentert at en løsning med to (geografisk adskilte) HPC anlegg og NorStore samlokalisert med disse er den som vil gi mest forskning over totalkostnader. En slik løsning vil kreve tildels store investeringer og/eller leieutgifter og være forbundet med risiko, og det kan tenkes at det er løsninger med flere og mer applikasjonstilpassede ressurser som utnytter allerede eksisterende dataromkapasitet som vil kunne gi like god eller bedre støtte til forskning og samtidig være mer kostnadssvarende. I alle tilfeller etterspørs detaljerte kostnadsanalyser for forskjellige alternativ som kan veies mot hverandre.

NTNU

NTNU mener at universitetene er best tjent med å huse de forestående e-infrastruktur anskaffelser. Den nye driftsmodellen for to HPC-anlegg samt Norstore er ikke utarbeidet og er ei heller operativ. Utviklingen og operasjonalisering av driftsmodellen vil kreve tid og modning, og bør stå sin prøve før man velger ekstern housing.

Lykkes man ikke med å etablere en driftsmodell som inkluderer alle universitetene, vil konsekvensene i tilfellet med ekstern housing bli at Sigma2 må ta driftsansvar mm. Et slikt scenario åpner veien opp for at Sigma2 vil utvikles til en CSC-modell, noe universitetene har uttrykt ikke er ønskelig. Sigma2 sitt styre vil i en slik situasjon pga aksjeloven og selskapets forpliktelser være tvunget til å la organisasjonen utvikle drift- og applikasjonskompetanse, selv om dette er i strid med samarbeidsavtalen.

I en situasjon hvor dataanleggene huses på et eller to universitet og driftsmodellen ikke viser seg effektiv, vil det være universitetene som huser infrastrukturen som vil måtte ta ta driftsansvar. Dette tvinger fram en reorganisering mellom partene, eventuelle et steg tilbake til dagens modell. Det reduserer imidlertid ikke universitetenes handlingsrom.

NTNU mener derfor at den nye driftsmodellen må utvikles og operasjonaliseres, slik at man har erfaring og samarbeidskontekst mellom alle parter som pr i dag utgjør NOTUR-prosjektet. Først da er man klar til å vurdere ekstern housing.

Forøvrig har NTNU et operativt varmegjennvinningsanlegg.