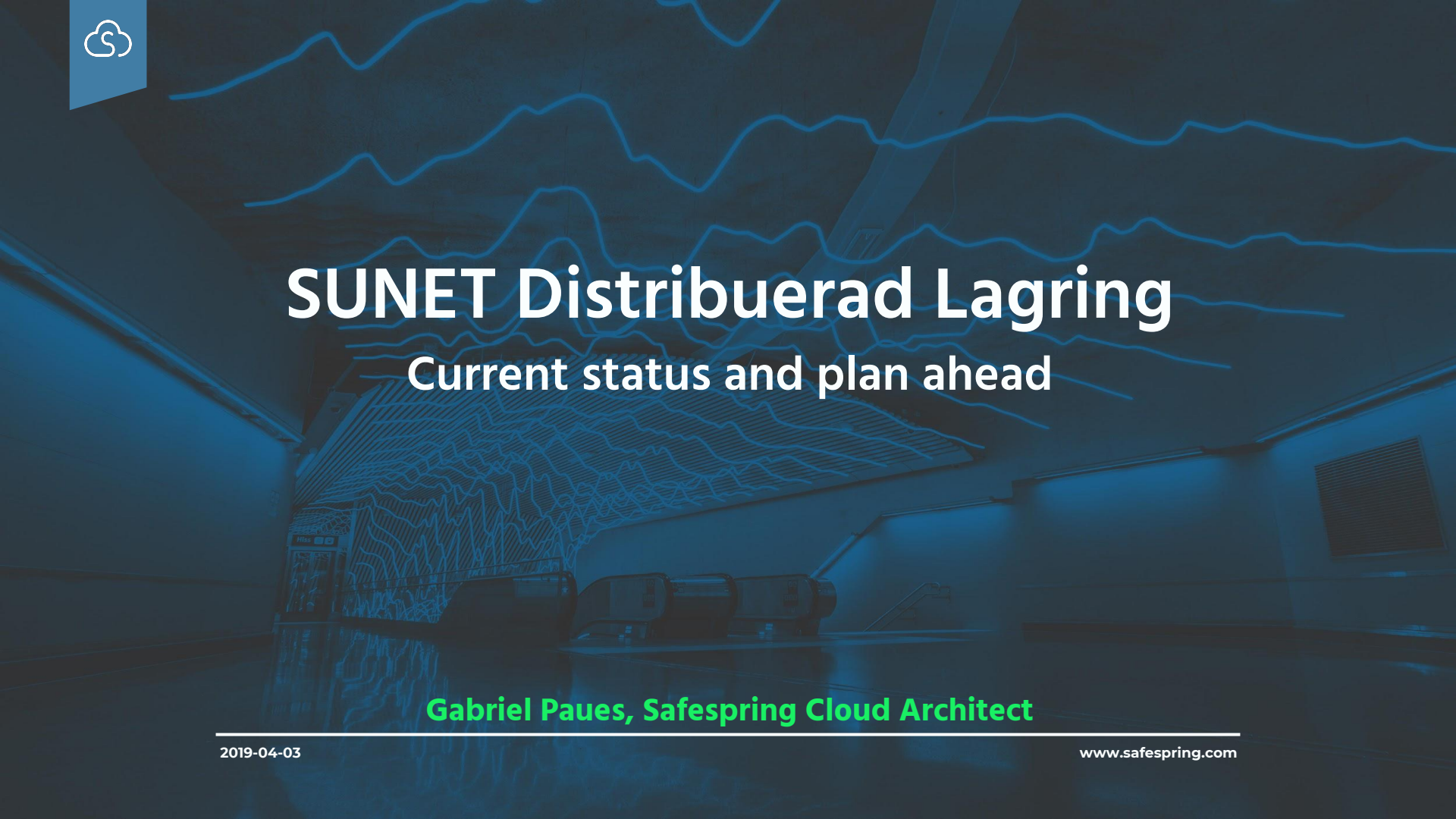
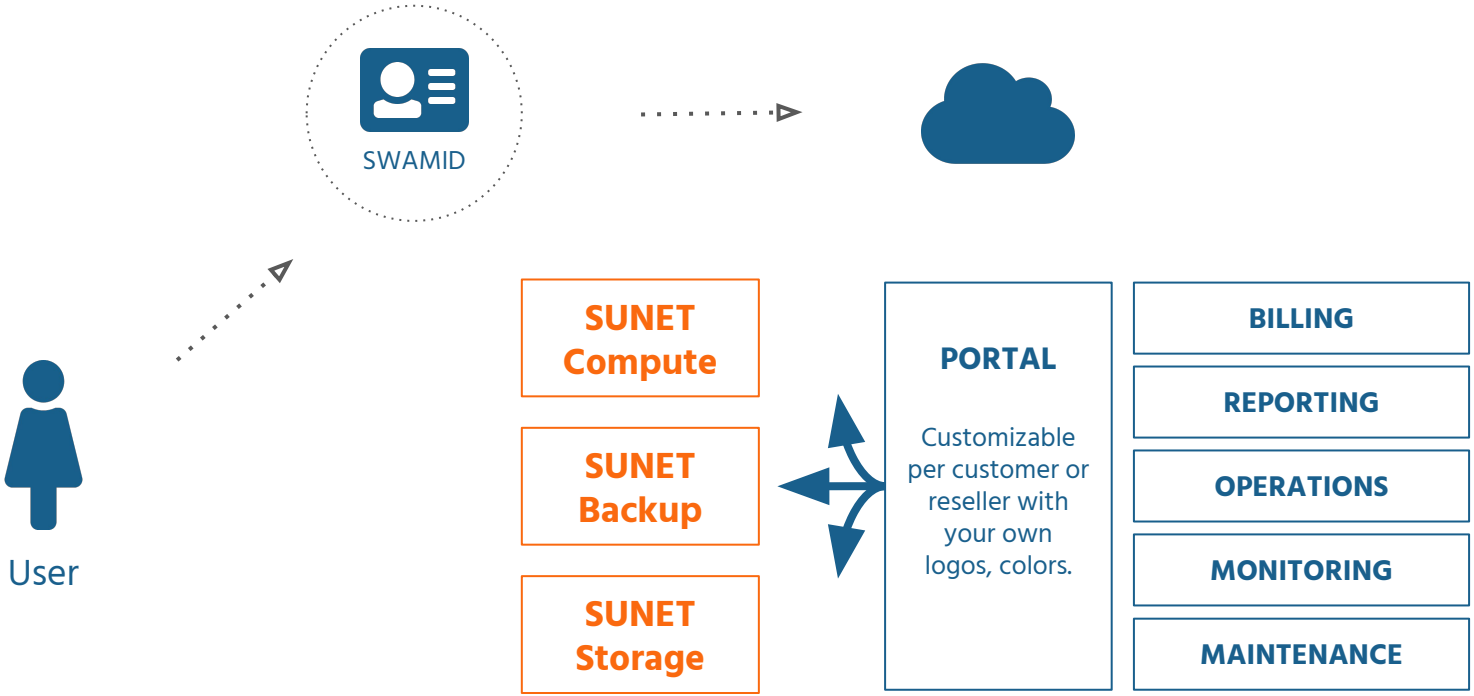


SUNET Distribuerad Lagring

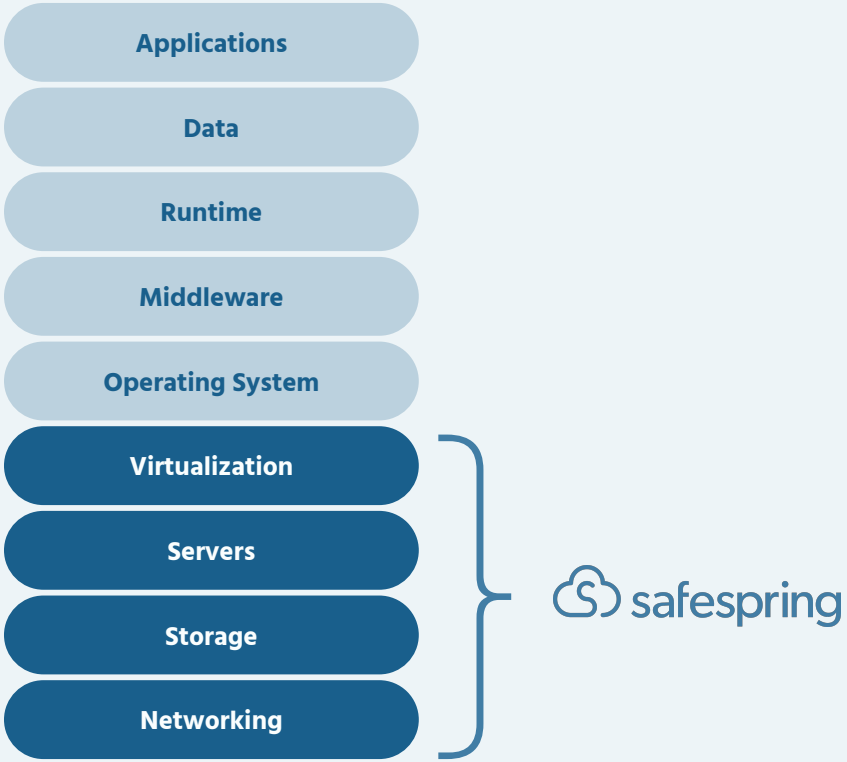
Current status and plan ahead

Gabriel Paues, Safespring Cloud Architect



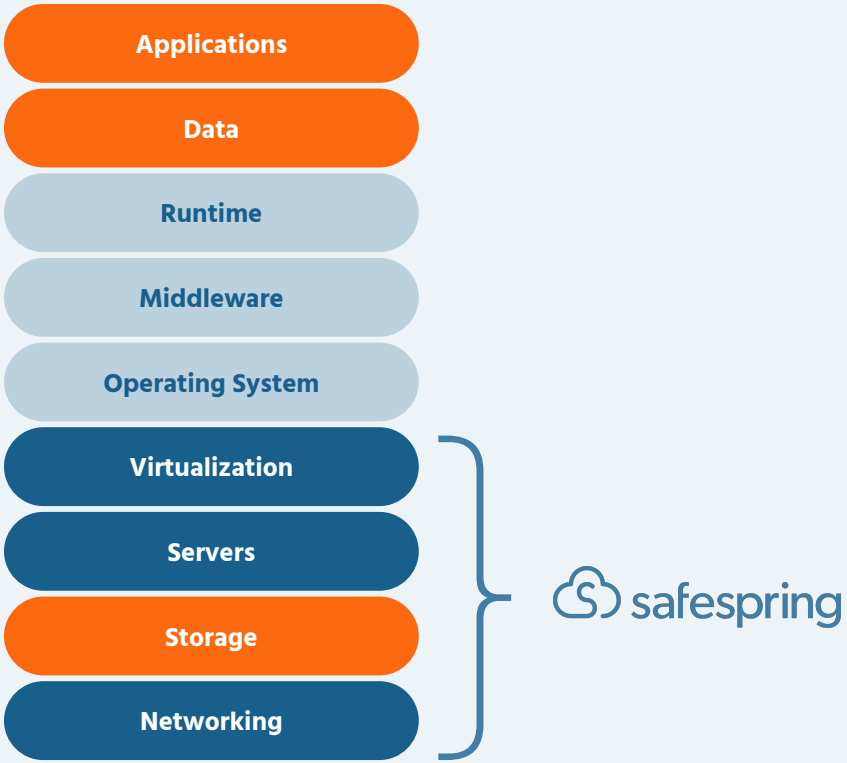
SUNET service delivery

Standardized building blocks



Safespring core competency

Identifying new building blocks



SUNET Object Storage

What is object storage?

Object storage is a computer storage architecture that manages data as objects, as opposed to other storage architectures like file systems which manage data as a file hierarchy.

Typically this is implemented by storing a binary *object* in a container together with *metadata* describing it.

https://en.wikipedia.org/wiki/Object_storage

What is object storage?

Typical data storage user requirements

- Availability
- Consistency
- Resilience
- Cost

It is significantly cheaper to provide guarantees *per object* instead of across all operations and/or objects as a filesystem.

What is object storage?

S3 API

A *de facto* standard based on representational state transfer (REST) over HTTP - the Amazon Simple Storage Service application programming interface.

Well-written specification has been available for over 7 years.

What is **SUNET** object storage?

A solution to *data gravity*
On premises
Data stay close to its users

Fast
Reliable
Cheap

FAST

RELIABLE

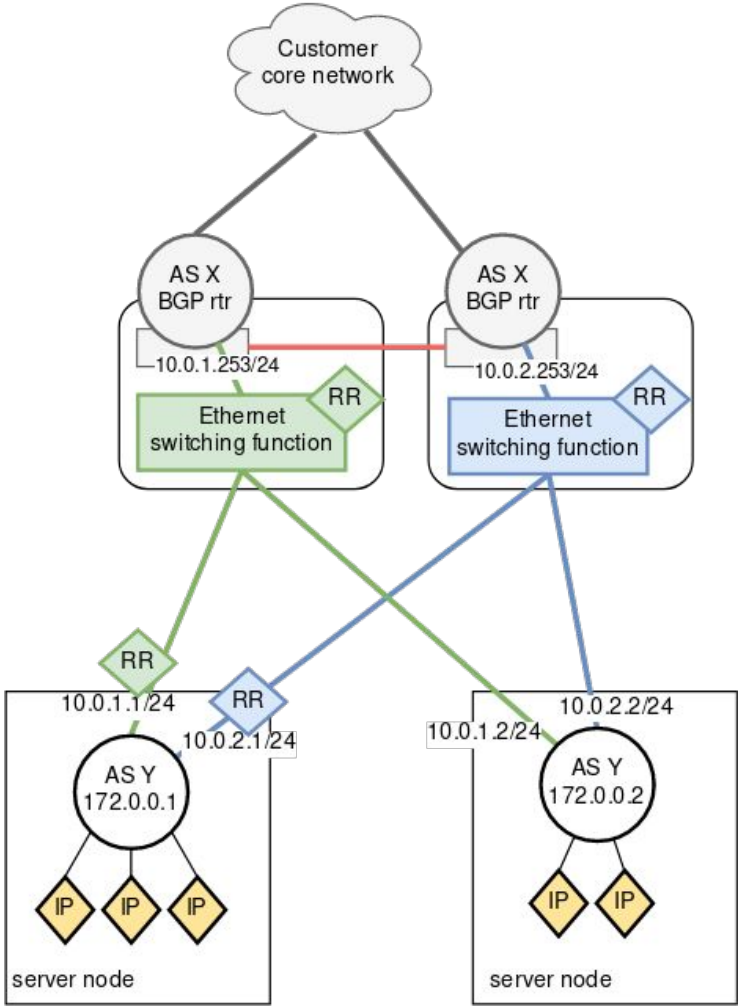
CHEAP

SUNET Object Storage

Proven BGP only network design

Internet as a *design pattern* for the datacenter cluster network ensure *scalability* and *predictable performance*. No surprises now, no surprises in the future.

Technically, this means *BGP is used everywhere*, even for the last hop to each server node.

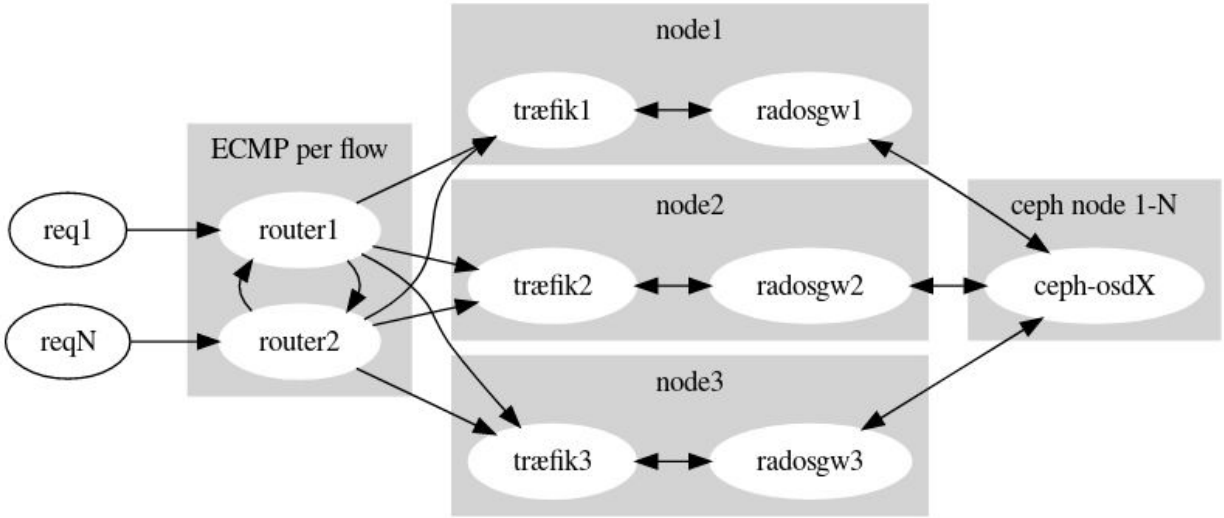

Customer
core network
AS X
BGP rtr
AS X
BGP rtr
10.0.1.253/24
10.0.2.253/24
RR
Ethernet
switching function
RR
Ethernet
switching function
RR
RR
10.0.1.1/24
10.0.2.1/24
10.0.1.2/24
10.0.2.2/24
AS Y
172.0.0.1
AS Y
172.0.0.2
IP
IP
IP
IP
IP
server node
server node

Load balancing as a primary concern

Safespring service operational knowledge combined with SUNET networking chops let us use technology patterns *normally not accessible* outside of hyperscale DCs.

SUNET and Safespring

S3 fast data path design



FAST

RELIABLE

CHEAP

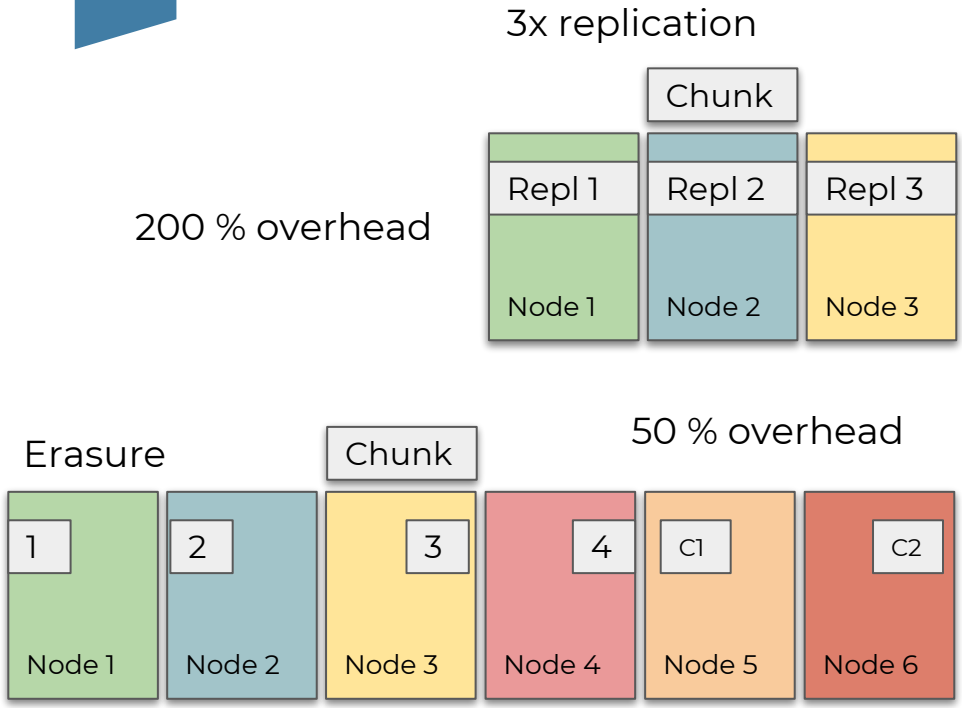
SUNET Object Storage

Well-known open source solution

The *Ceph Object Gateway* is an open source implementation of the S3 storage API. It is in use worldwide at large academic institutions to help solve petabyte-scale storage needs.

Multiple commercial software vendors (Redhat, SUSE) and service providers has backed the solution for several years.

SUNET Object Storage builds on top this openly available knowledge pool.



Software defined resilience levels

As a user you can decide what **storage resilience policy** you want to implement.

3x replication of each object is the standard, but **erasure encoding** or forward error makes it possible to further reduce costs at the expense of computational complexity.

FAST
RELIABLE
CHEAP

SUNET Object Storage

End user pricing

Goal:

Pricing *on par or better* than comparable products from public cloud vendors Google, Microsoft, and Amazon.

- Currently 0,06 SEK per GB/month
- SLA designed for volume storage

Efficient development and operations model

The Safespring devops model specifically targets *highly skilled senior engineers* with already proven field experience.

Operational costs are kept low due to best practice knowledge sharing and every responsibility being a team effort.

We build it, we operate it.

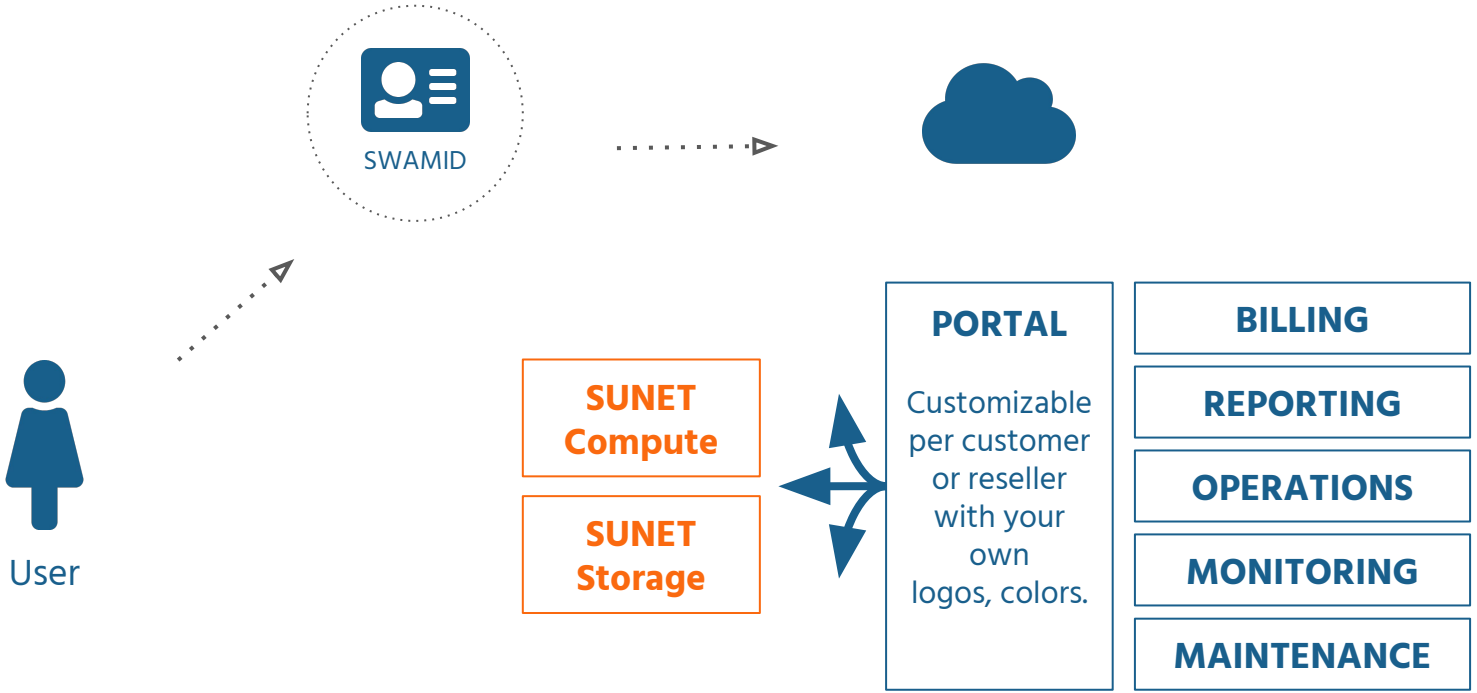
Hardware knowledge

Hardware is *really boring*. But Safespring know all the boring stuff, so you don't have to.

For example, this Supermicro box takes 24x12TB drives. It also holds *enough RAM and CPU* for all the drives to reliably work with it as a Ceph cluster node.

This is the current sweet spot for our *operational parameters*. Entry level site will have 12 nodes like this for ~2PB capacity.

Managed Compute

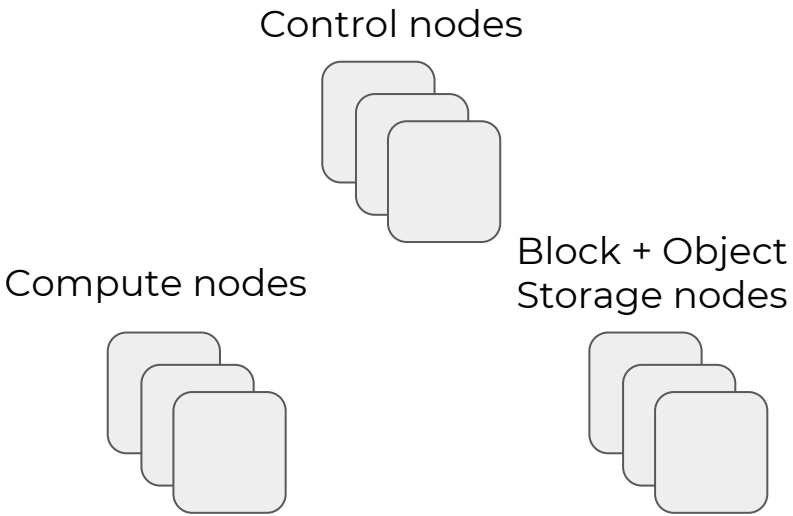


SUNET Private Cloud service delivery

Differences from Public Compute

Public Compute:

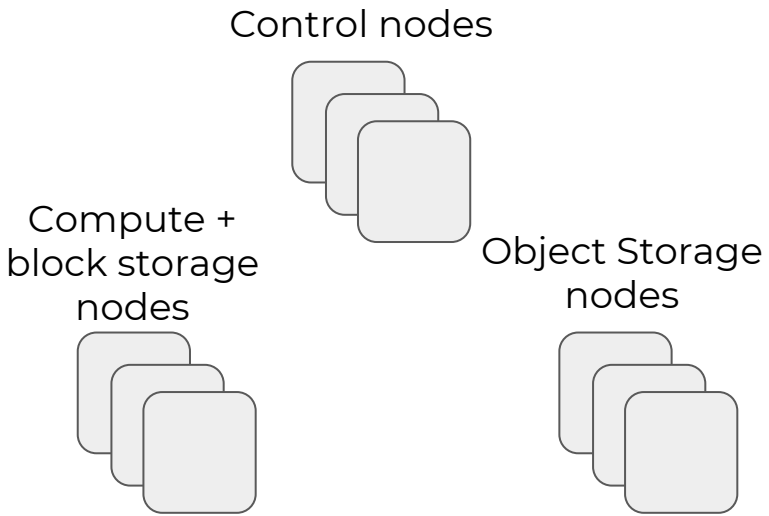
- Catering for different needs
- Windows and Linux
- Large boot volumes
- Reliability before performance (storage separated from compute hosts)



Differences from Public Compute

Managed Compute (adapted for science)

- Focus on performance
- Local block storage on compute nodes
- Mostly Linux
- Block storage used for temporary storage during analysis
- Object storage for long term storage



Hardware - Compute

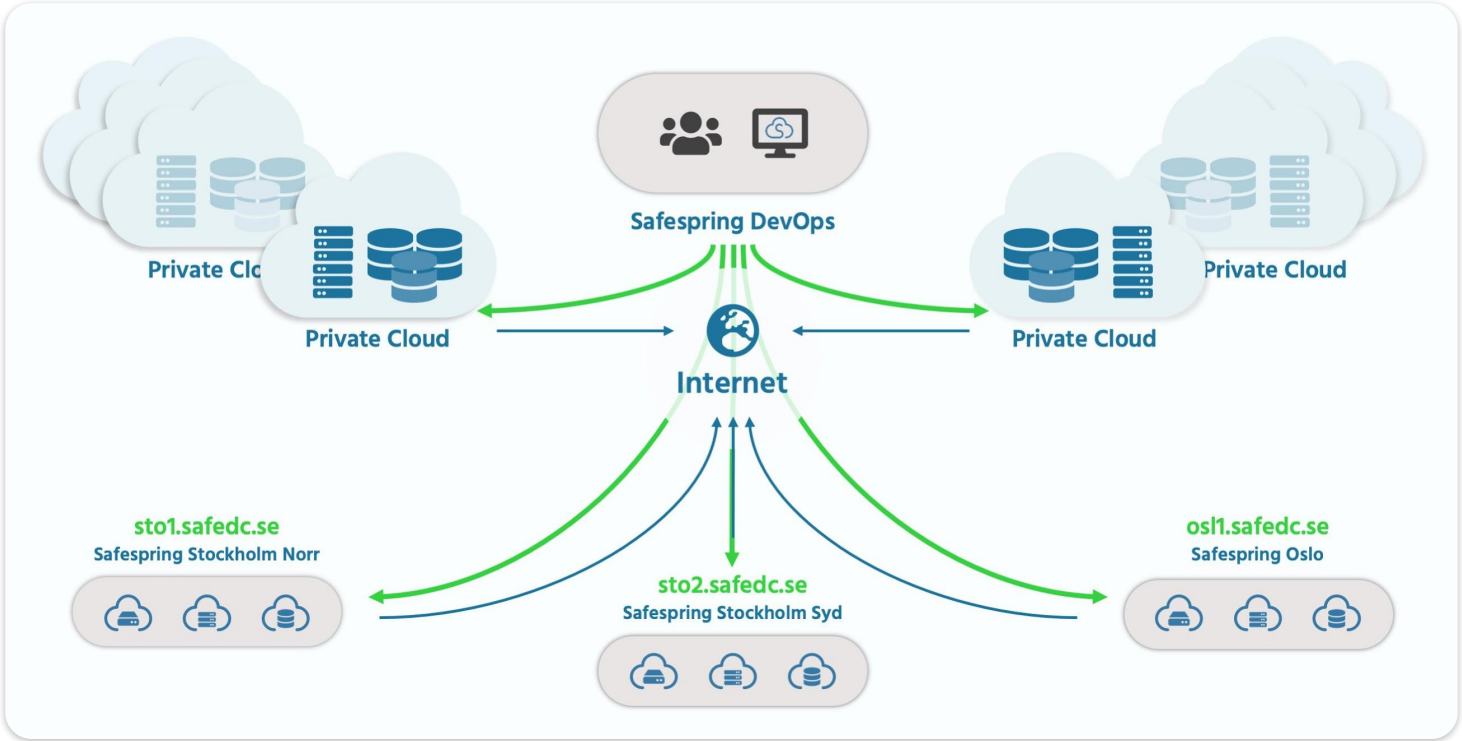
Hardware designed for Compute intensive application, HPC & Datacenter

- Dual CPU (64 cores per CPU 1:2HT)
- Local NvME disk - Fast IOPS
- Reference VM 16 vCPU and 500GB disk

Second gen IaC - Safespring DevOps - Advantages

- Update systems **faster**
  - Lower barrier to changes
- **Reproduce** systems as needed
  - Build everything with as few dependencies as possible
- **Add or change** easily
  - Target the affected nodes easily
- **Verify** that software works as intended
- **Scales better** with many operators

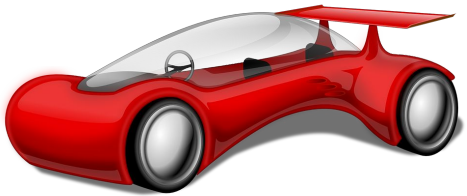
Hybrid Cloud



On the drawing board: GPU power

Two variants:

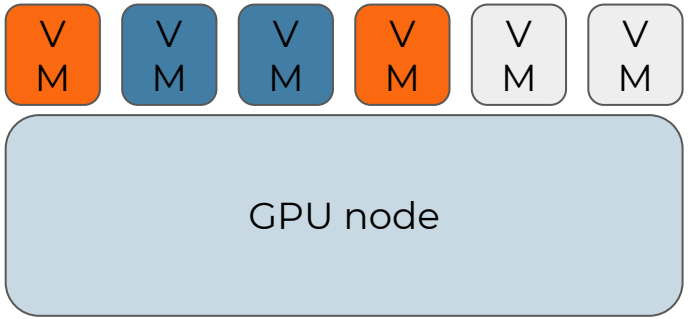
- Virtualized for sharing GPU resources between different projects
- Physical dedicated to one project at a time
- Based on sector requirements



On the drawing board: GPU power virtualized

Virtualized

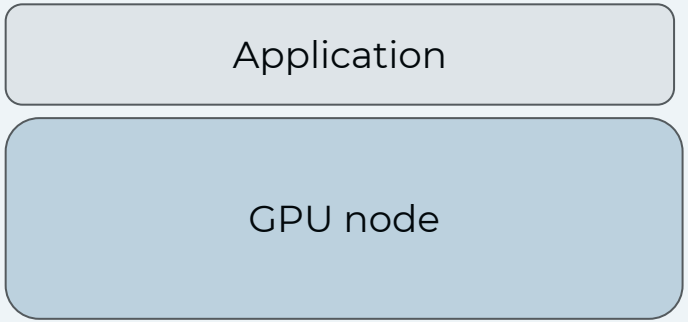
- Good for testing
- Many users can share the same resources
- Using KWM and Hardware Assist
- Lower performance



On the drawing board: GPU power - physical

Physical

- Production
- Dedicated to one project
- Using OpenStack Ironic for bare-metal provisioning
- High Performance = More expensive



Reference group

SUNET is organizing a technical review group. It will work with the project to review the processes, designs and SLA.

It is crucial to SUNET & Safespring as a vendor to make sure the service meets expectations and requirements.

https://wiki.sunet.se/display/SDL

Current project status

The SUNET Object Storage project is currently validating the service designs by building a site together with SUNET.

This work will let us develop the service further by combining our experience at all stages of delivery.

The goal is to have a MVP (minimum viable product) solution on air by 2019-04-12.


QA
Follow us
linkedin.com/company/safespring
twitter.com/safespring