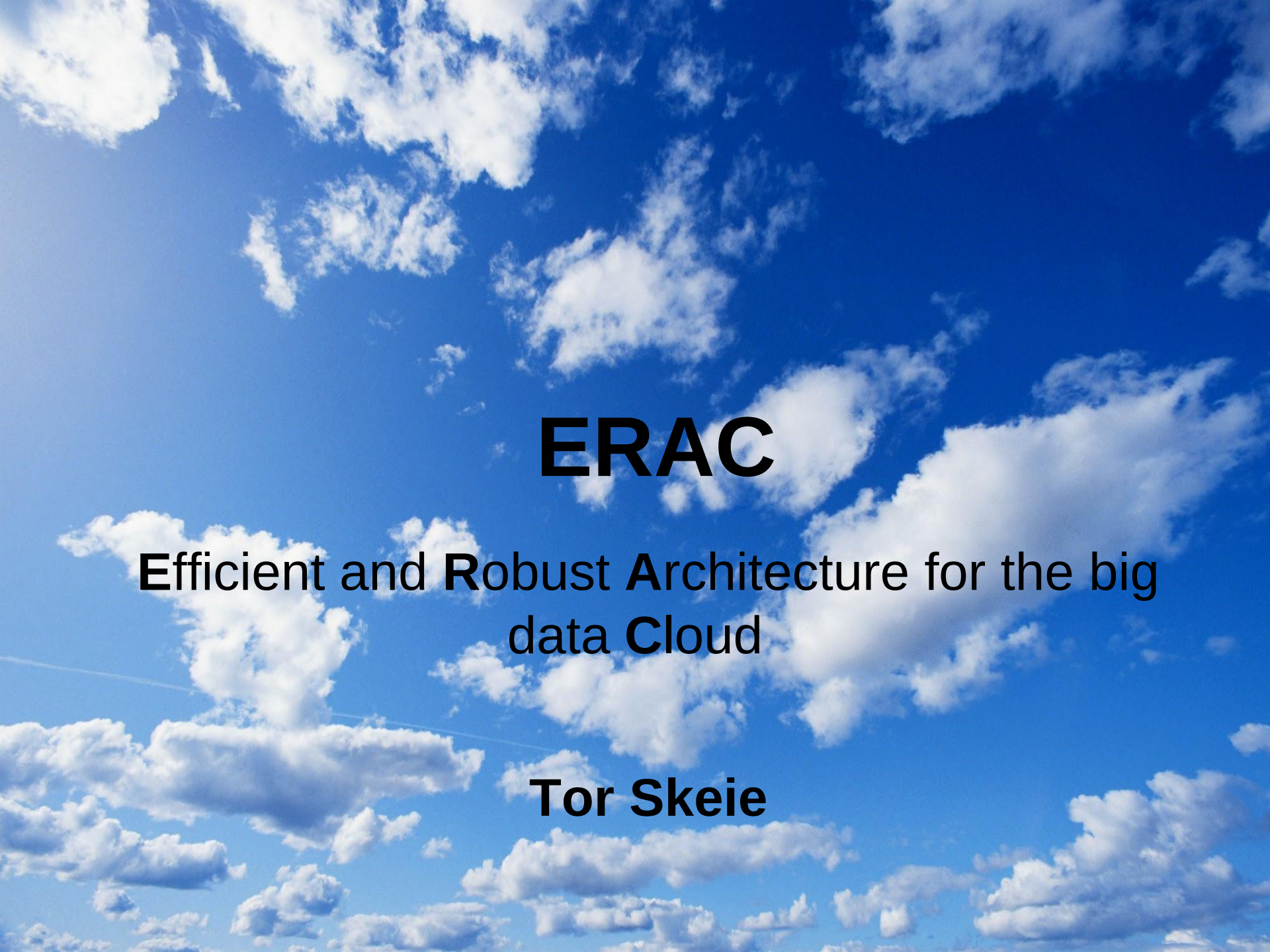
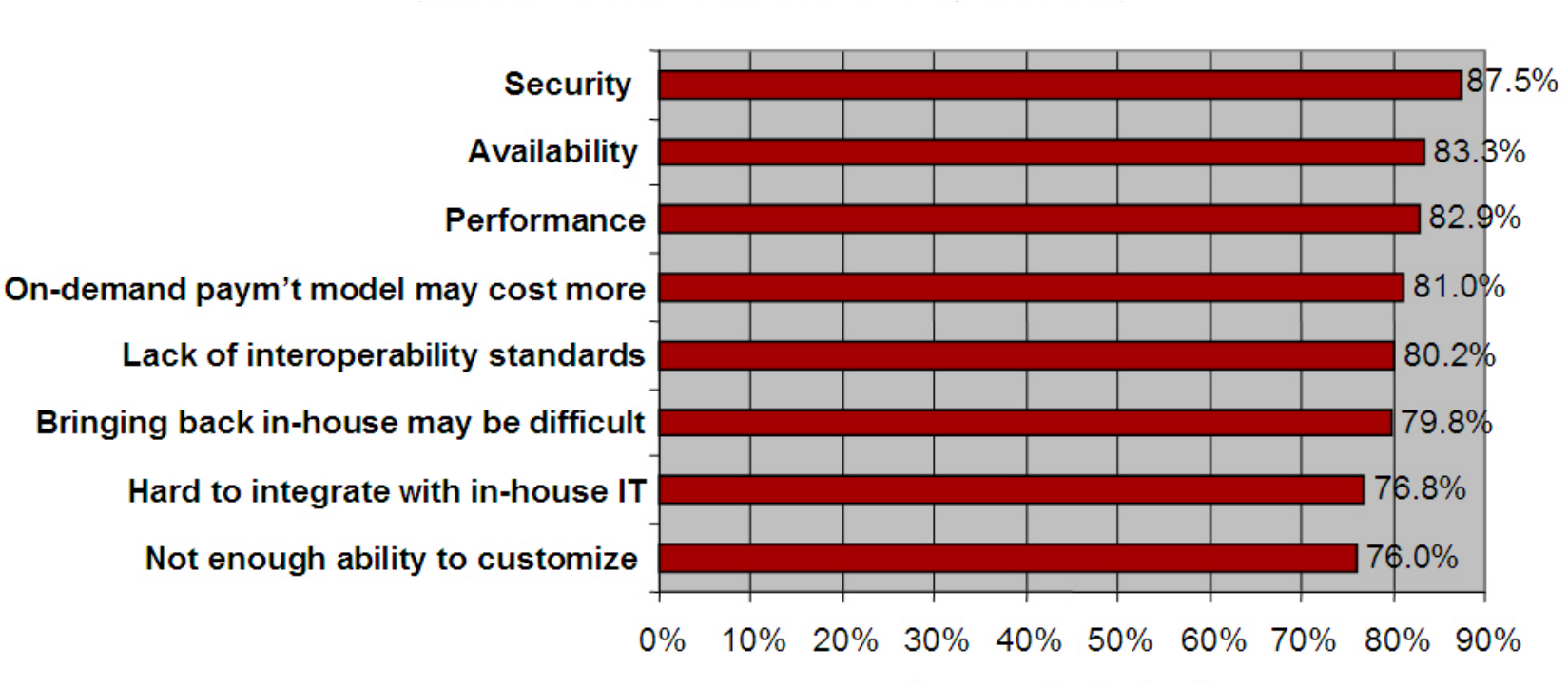

ERAC
Efficient and Robust Architecture for the big data Cloud
Tor Skeie

Cloud computing challenges

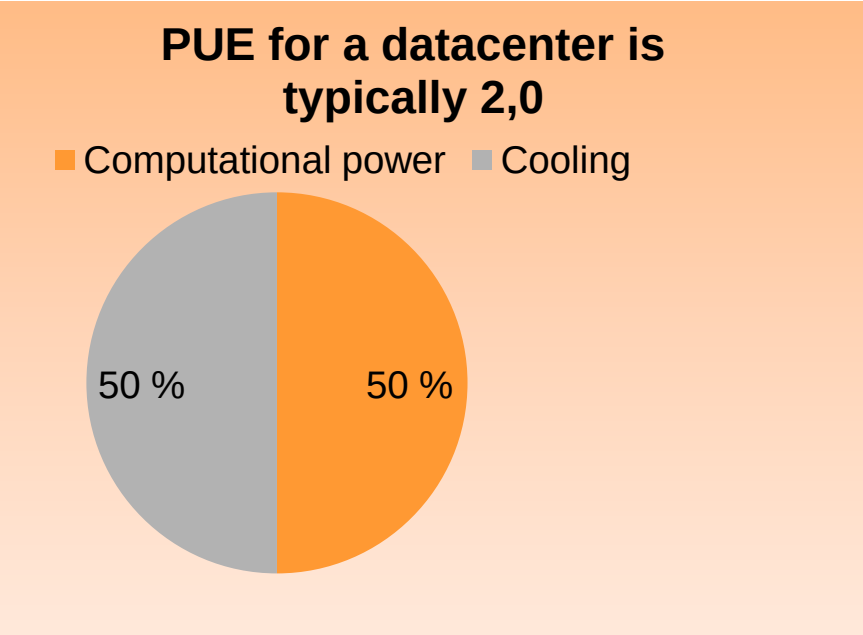
Results of a survey among enterprises, conducted by IDC



Source: IDC Enterprise Panel 2009

Performance requires power

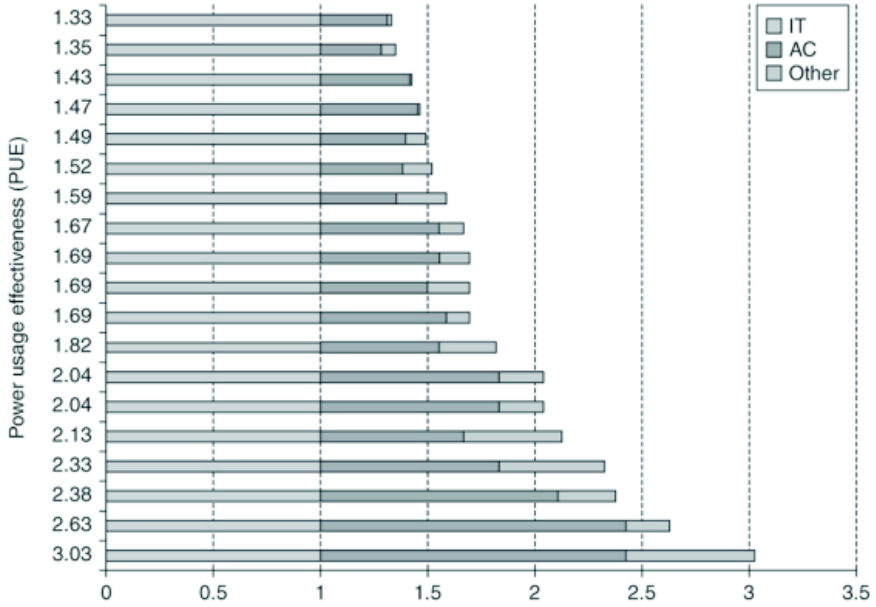
- The IT industry is responsible for 2,0% of the CO_2 emission worldwide, the same amount as the aircraft industry.
- Consumes 2% of the global power production and that is increasing significantly, assumed to be 4% in 2015
- **PUE** (Power Usage Effectiveness) is a measure of how efficiently a computer data center uses its power; specifically, how much of the power is actually used by the computing equipment in contrast to cooling and other overhead



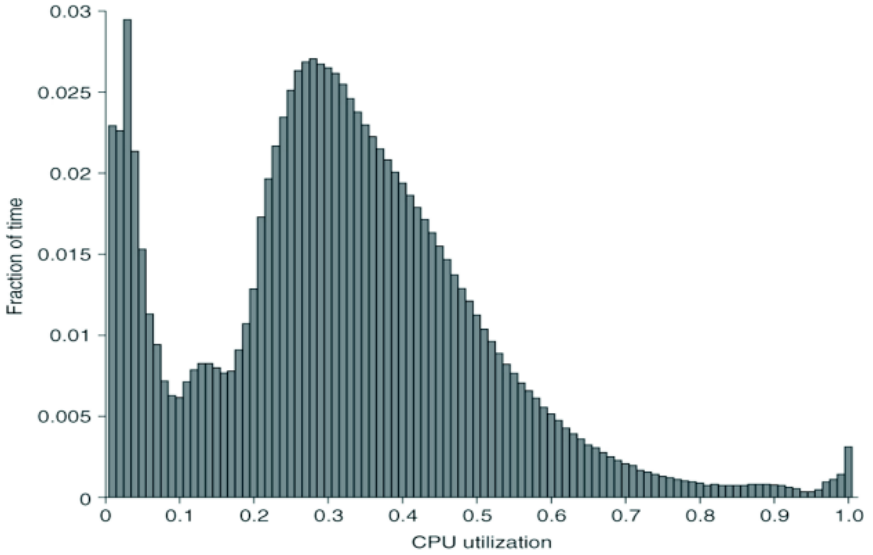
- *If all the data centers in the world had a PUE around 1.6, the Kyoto-agreement would be met*
- *A green data center should have a PUE of not more than 1.2*

Cloud datacenters have poor utilization – efficiency today

Power efficiency of 19 datacenters



CPU utilization of 5000 random Google servers over a period of six months



Source: Computer Architecture: A Quantitative Approach. By John L. Hennessy, Stanford University and David A. Patterson, University of California at Berkeley

On robustness of datacenters

- On April 21st 2011 the Amazon EC2 cloud computing service was down for 36 hours taking with it more than 70 websites, including parts of the New York Times and popular social networking sites
  - Amazon blamed the problem on ”network errors”, but no future details have been disclosed

- Then, on April 22nd, Sony Corporation´s PlayStation Network (PSN), one of the worlds largest social networks related to gaming, music, and movies, was taken out of service and Sony announced that the PSN had been compromised
  - The PSN was out of service for three weeks and around 77 million registered users where affected by the outage. Moreover, account information and credit card details for these users might have been exposed

The above incidents, on two consecutive days, show real world examples of the vulnerability and lack of robustness in the communication technology and infrastructure (the cloud)

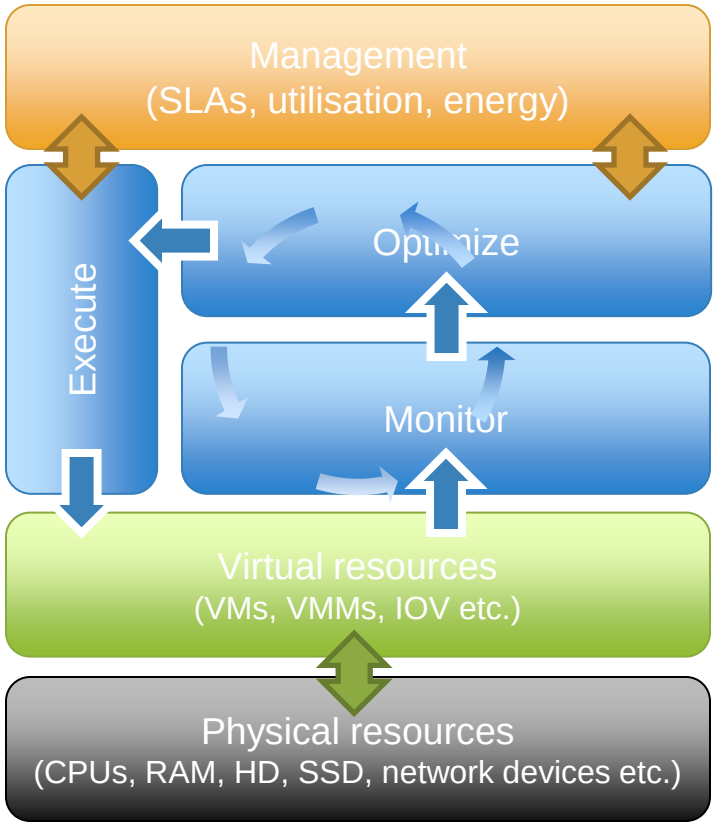
Motivation

Cloud computing has the potential to change how we do computing in the future, but for cloud technologies to reach their full potential more research is required. We have identified the following challenges:

- System utilization
- Rapid elasticity
- Robust computing
- Secure computing
- Energy efficiency

In addition, this expertise can complement the existing plans for several Norwegian data centres driven by the Norwegian industry

Research vision



- Collect performance and faults data.
- Optimize resource provisioning based on collected data and management directives.
- Create new system configuration including any manual overrides.
- Reconfigure resources.

Research topics

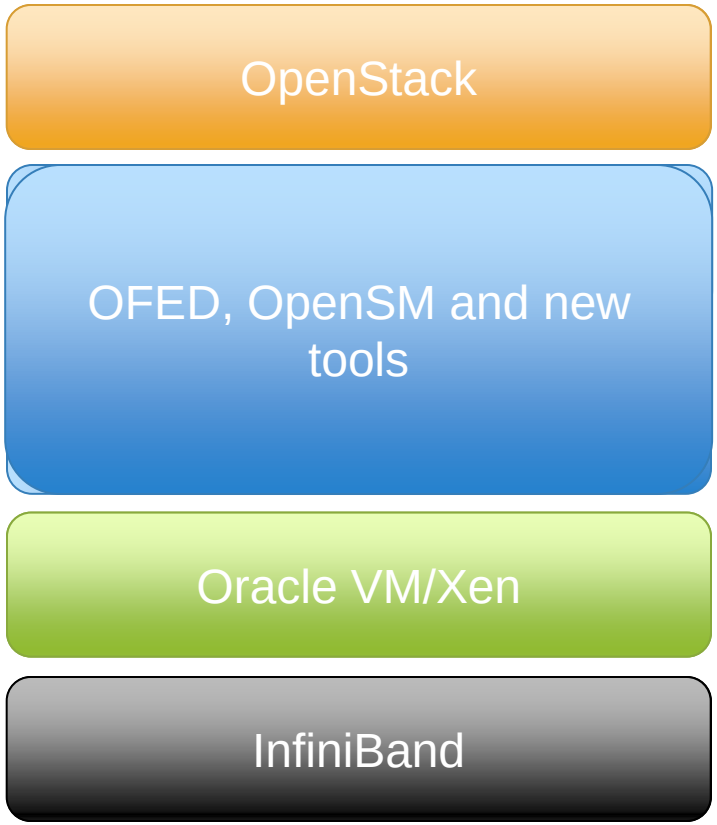
- Dynamic fabric routing and resource allocation (a self-adaptive network)
  - Methods for real-time monitoring of compute resources.
  - Automatically adapt routing tables, path selection algorithms, and priority settings based on existing workload and faults.
- Dynamic provisioning of virtual resources (a fully virtualized cloud)
  - Methods for real-time monitoring of communication resources.
  - Virtual machine placement and migration.
  - Improved virtualization of network resources (e.g. RDMA devices).
  - Increased VM flexibility in order to rapidly rearrange and scale resources.
- Secure storage and sharing (a secure cloud, performed by UiS)
  - Guarantee the authenticity of data.
  - Avoid unauthorized altering of data.
  - Prevent data leakage when stored at a 3rd party.

Cloud test bed I

We will work experimentally and creating proof-of-concept systems using:

- an InfiniBand based private cloud
- a prototype for each research topic:
  - a self-adaptive network
  - a fully virtualized cloud
  - a secure cloud
- *Might* add DCB Ethernet to increase technological impact

Cloud test bed II



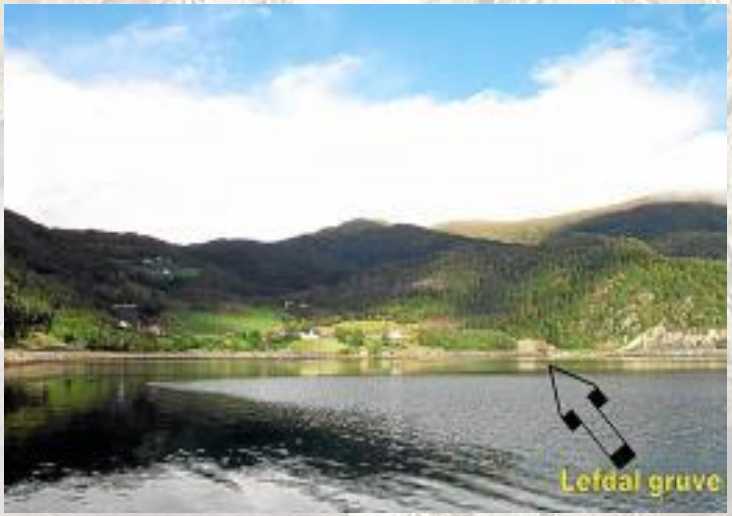
- Collect performance and faults data.
- Optimize resource provisioning based on collected data and management directives.
- Create new system configuration including any manual overrides.
- Reconfigure resources.

Project details

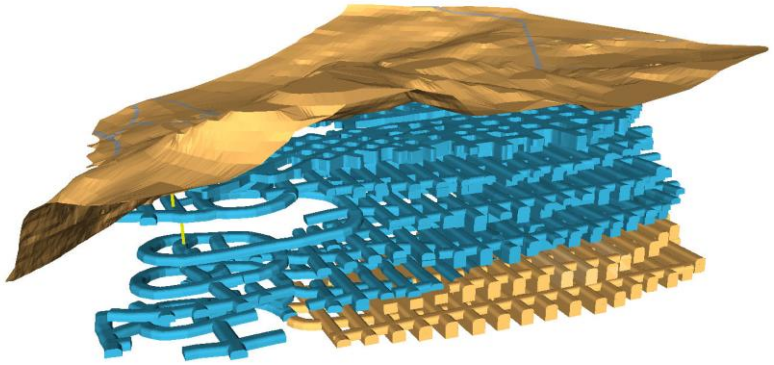
- Academic partners: Simula, UiO, and UiS.
- Industry partners: Oracle and Lyse.
- Budget of NOK 15.7 million over 4 years:
  - 11.5 million from the Research Council.
  - 2.6 millions from Oracle.
  - 1.6 millions from Lyse.
  - In addition Oracle will contribute with in-kind
- There will be 3 Ph.D. students and a post doc. plus senior personnel working on the project

Norway well-suited for hosting green datacenters

- Cold climate, offering free cooling
- Empty mountain halls and abandoned mines, being bomb proof
- Cold (sea) water for effective cooling
- Competitive electricity prices, made from hydroelectric power
- Stable geology
- Economical and political stability
- *Source: The ICT-Norway report ”Sky og fjordane”*



Lefdal mine is green datacenter pilot




Thanks for listening!