



UPPSALA
UNIVERSITET

The SNIC Science Cloud (SSC) and Glenna2

Andreas Hellander

Associate professor in Scientific Computing, Uppsala University

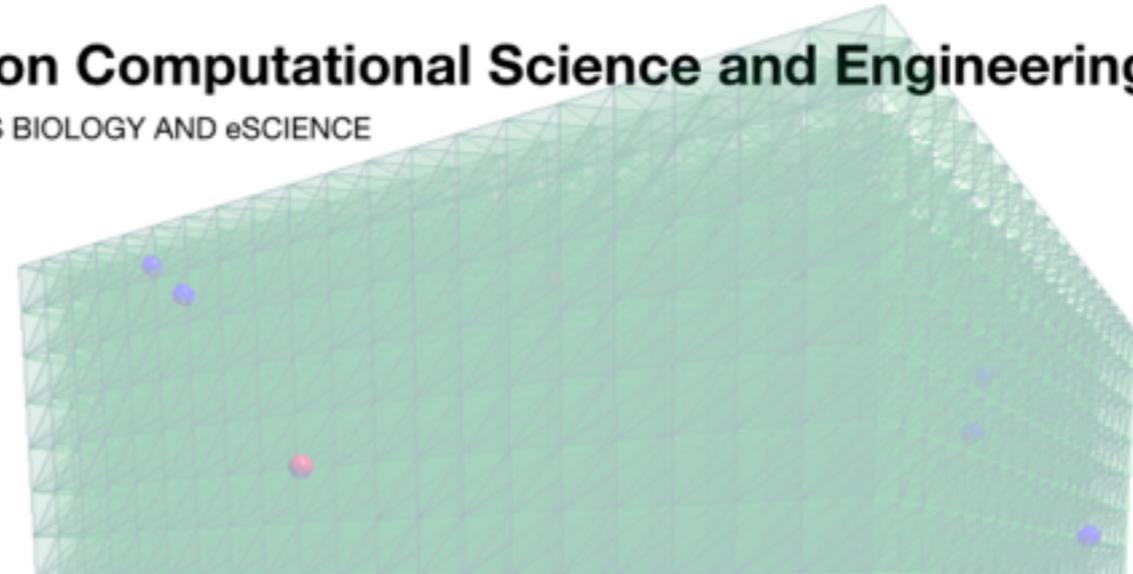
Project manager, SNIC Science Cloud

Steering group member Glenna2



Hellander Lab on Computational Science and Engineering

COMPUTATIONAL SYSTEMS BIOLOGY AND eSCIENCE



Computational systems biology:

- Multiscale methods for quantitative simulations of cellular pathways and cell-cell interaction.
- Algorithm & theory development for reaction-diffusion kinetics.
- Modeling of specific systems with collaborators in biology

e-Science Tools & Techniques:

- Applied Cloud Computing and data-intensive computing
- Computational Science and Engineering (CSE) software
- Smart systems for scientific discovery

Courses in e-Science on MSc. and PhD level:

- Large Datasets for Scientific Applications (Tools&Techniques for Big Data analysis, Hadoop, Spark etc.)
- Applied Cloud Computing (cloud fundamentals, contextualization, orchestration, automation etc.)

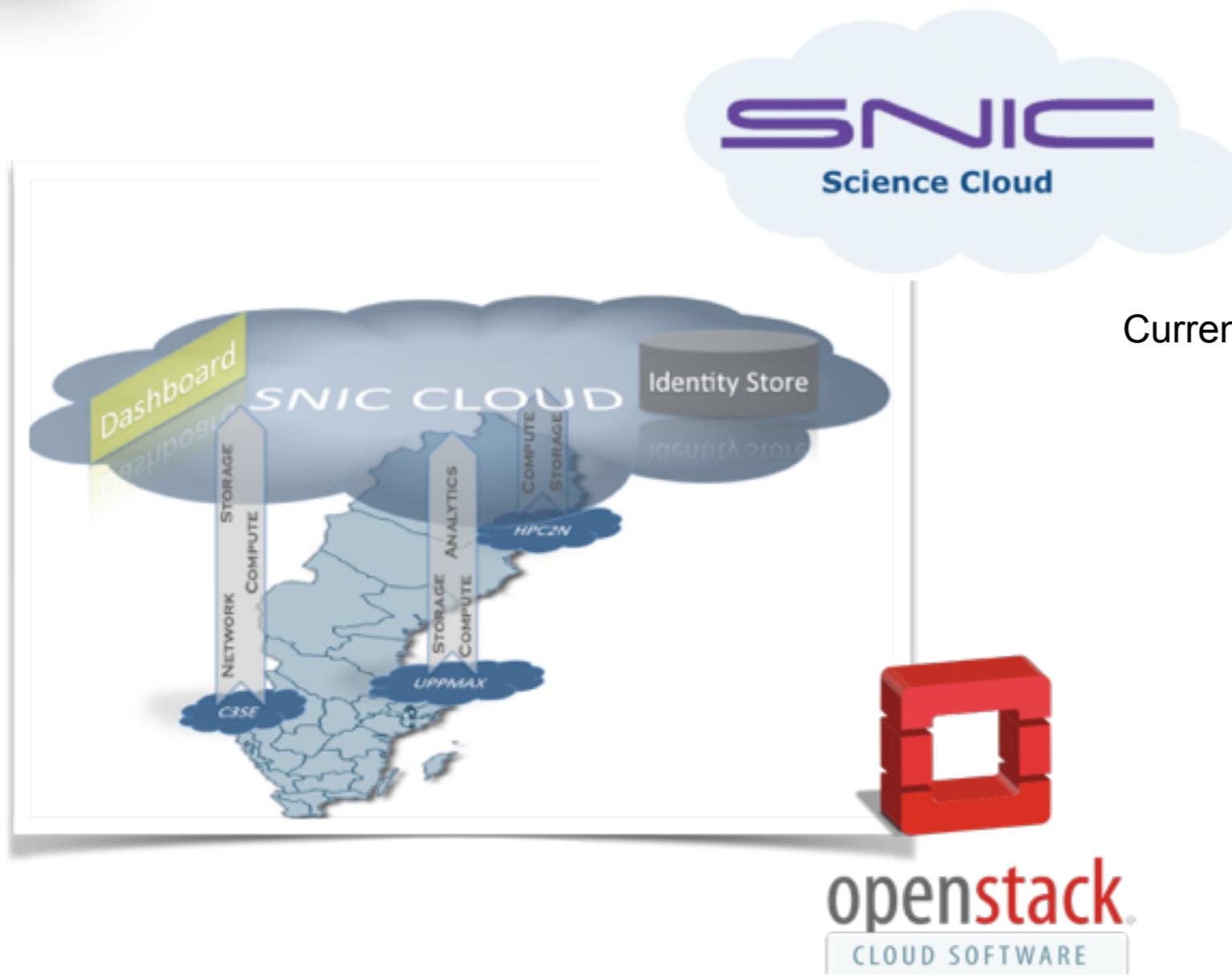
SNIC Science Cloud

- An e-infrastructure capable of meeting the needs of *new and emerging use cases / “long tail of science”*
 - Requires flexibility, scalability
 - Technological focus on OpenStack

The project has run in two main phases:

1. 2013-2014: A small-scale pilot system and a small number (3-5) of predefined pilot use-cases, chosen to highlight the utility of cloud resources. Advanced user support in the form of SNIC Application Experts (AEs) were available to assist the pilot users.
2. 2015-2016: A scaled up infrastructure with more resources, and an open process for user-initiated project requests. In this phase, we have concentrated on training workshops and creating open source tutorials instead of targeted advanced support.

The SNIC Science Cloud (SSC)



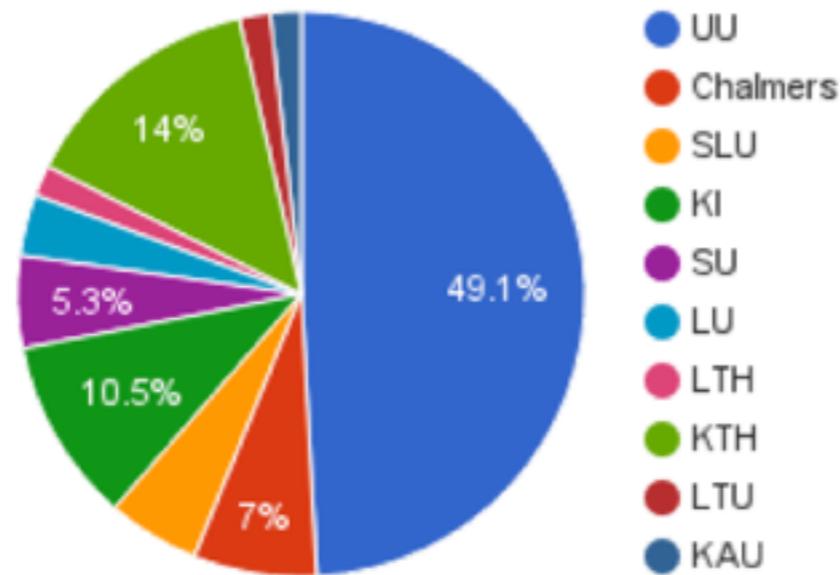
Current status:

- Cloud IaaS in “preproduction”
- Federated access in Sweden, working on Nordic Federation
- Full production planned for August 2017
- Currently 2 regions, ~1500 vCPUs
- Soon: 3 regions, ~5000-6000vCPUs
- **Appears to be a highly cost effective way of providing IaaS for science compared to commercial IaaS providers (still under investigation)**

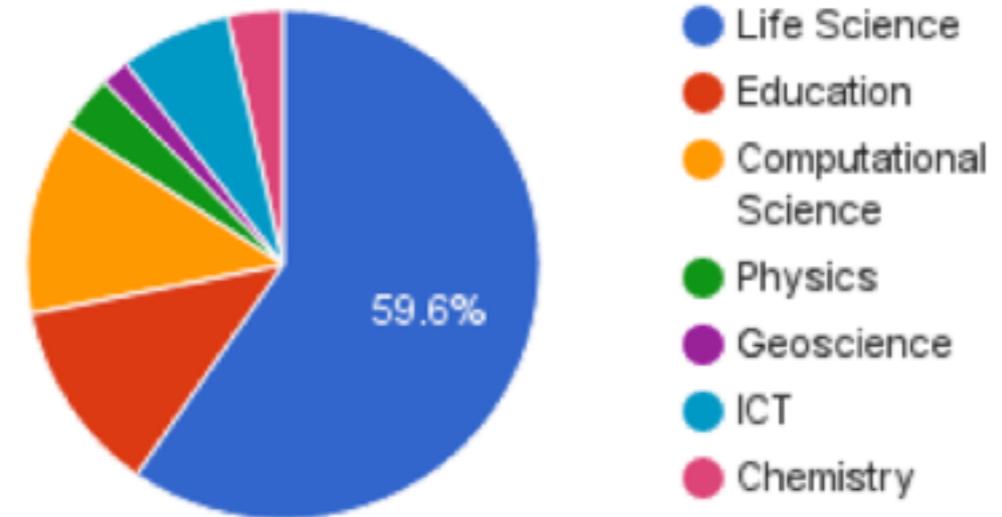


Who are the users of SSC?

SSC projects per institution



SSC projects per area



User stories:

- Elastic proteomics analysis (researchers in Sweden and Switzerland).
- Processing ozone data from the Odin satellite at Chalmers University of Technology.
- Estimation of failure probabilities with applications in underground porous media flows.
- Virtual Research Environment for Clinical Metabolomics (H2020 project Phenomenal).

<https://cloud.snic.se/index.php/2016/10/11/who-are-the-pilot-users-of-the-snic-science-cloud/>

Some current and future challenges

- Flexible and performant data analysis
 - Complex problem, complex political landscape
 - Users not accustomed with “modern” solutions
 - Multi-cloud
 - “Convergence” between Cloud/HPC
- “Pricing model” that encourages good resource usage
- Ensuring competency on OpenStack and PaaS for “Big Data
- Educating users how to consume resources more dynamically
- Security
- Future funding for HW, fragmentation

Many of these challenges are best met in collaboration.

The NeIC Glenna project (2014-2016)

Glenna: Share knowledge and set best practices on managing cloud services and to create a Nordic federated cloud service, driven by the need of the Nordic researchers.

- Funded by the Nordic e-Infrastructure collaboration (NeIC) / Nordforsk (Federated Cloud from the Nordic eScience Action Plan)
- Sweden, Finland, Norway, Island, Denmark
- Focus on knowledge exchange,
- Cloud IaaS / PaaS for Big data analytics

Key outcomes of the project:

- Drastically increase in awareness and competency on cloud computing and platforms for analytics
- Increased stability of cloud IaaS initiatives
- Pilot PaaS based on Kubernetes and Apache Spark

The Glenna Environment

<https://neic.nordforsk.org/activities/glenna/>



Plan: Federate most important (all, but optional) services using eduGain

Glenna 2 (2017-2020)

Project directive approved by NeIC board: https://wiki.neic.no/w/ext/img_auth.php/c/c5/Glenna_2_Project_Directive_Final.pdf

Aims to provide added value to the national cloud and data-intensive computing initiatives by:

- Supporting national cloud initiatives to sustain affordable IaaS cloud resources through financial support, knowledge-exchange and pooling competency on cloud operations.
- Using such national resources to establish an internationally leading collaboration on data-intensive computing in collaboration with user-communities.
- Leveraging the pooled competency to take responsibility for assessing future hybrid cloud technology and communicate that to the national initiatives.
- Supporting use of resources by pooling national cloud application expert support and create a Nordic support channel for cloud and big data. The mandate is to sustain a coordinated training and dissemination effort, creating training material and providing application-level support to cloud users in all countries.

Vision: collaboration on scalable data analytics

Open and reproducible data analysis on a massive scale offers many challenges. Through collaboration we will facilitate for researchers to have **access to, analyze, present** and **share** large datasets.

Some concrete aims/activities in Glenna2:

- Create mechanisms to simplify the process of importing/staging data from external data sources (public cloud storage, EUDAT, OSDC, national provider infrastructure, DeIC OwnCloud, SeaFile etc.) onto scalable and resilient, compute-located cloud storage within the IaaS infrastructure (such as the Swift Object Store).
- Provide modern means to present and reproduce complex analysis through e.g. web-based notebooks (such as Jupyter and Apache Zeppelin). Here, cPouta blueprints (<https://github.com/CSC-IT-Center-for-Science/pouta-blueprints>) developed during Glenna should form the basis.
- Create a variety of demonstrators of how a user can leverage Nordic cloud resources to provision computing resources close to data (i.e. instead of a user importing data to the national cloud provider, the user will provision virtual compute resources with the cloud provider that most effectively can deliver that particular dataset irrespective of where that is).
- Assess, in close collaboration with national providers and in particular building on the experience in Norway, how to deliver affordable cloud storage as a backend for computing with good performance.

Vision: Convergence Cloud/HPC

Flexible infrastructure ecosystem meeting needs of HPC/MTC/HTC/
data-intensive use-cases:

- Performance important
- Cost efficiency important
- Should be service oriented
- Should be “data-friendly”
- Reconfigurable

Glenna2: Mapping out future generations of hybrid e-infrastructure

Some concrete aims/activities in Glenna2:

- Establish a test-bed for hybrid infrastructure using existing OpenStack technology such as OpenStack Ironic (bare-metal provisioning).
- Collaborate with the OpenStack scientific working group which is focusing on high-performance clouds for scientific use cases (https://wiki.openstack.org/wiki/Scientific_working_group).
- Using the national IaaS resources, provide platform-level infrastructure for microservices-development, such as OpenShift or Kubernetes. Ideally, showcase how Dev/Test can be done in cloud IaaS and production seamlessly deployed in UNINETT Sigma 2's bare-metal clusters. Proof-of-concept has been established in Glenna I.