Federation

the sound of inevitability

Martin Brandt, Markus van Dijk

SURF SARA

Past: SARA → SURFsara

1971: The two Amsterdam universities and the Mathematical Centre create SARA, a facility for their academic compute services. Other universities and research institutes join later.

1984: SARA starts hosting the National Supercomputer (a CDC Cyber 205).

2013: SARA joins the SURF cooperation and is renamed to SURFsara.

Present: SURF and HPC clouds

SURF clouds

SURFCumulus

SURFsara HPC Cloud

Life Science Cloud

Grid Services Park

data clouds: SURFdrive, Beehub

Federations

EOSC "the European Open Science Cloud for research"

Elixir

Federated Cloud Groningen/NIKHEF/SURFsara

Developments

EGI Federated Cloud

Helix Nebula "the science cloud"

SURFsara Research Cloud

HTC-PaaS

iRODS

SURFcumulus

"your institution can move its data centre to the cloud"

Reliable hosting of services, IaaS.

- email
- websites

Providers:

- SURFnet
- KPN
- Vancis

Future: Azure, AWS, Dimension Data

SURF SARA

SURFsara HPC Cloud

High Performing Computing IaaS for researchers

OpenNebula, VM creation by end-users. Some pre-build images available for import.
Originally optimised for MPI applications. No VM lifetime limit. Direct Internet access.

Currently 286 users in 192 projects, 560 VMs running. Since 2011: 780 users in 530 projects.

Biggest possible VMs:
  80 cores, 500 GB RAM
  40 cores, 2000 GB RAM

Totals:
  3300 cores, 32 GPUs, 18600 GB RAM
  82 TB host-local SSD, 900 TB fast remote image storage (CEPH)
  Switch allows 2 Gbit non-blocking interconnect between hosts.

SURF SARA

Life Science Cloud

Evolution of the Life Science Grid

Targeted specifically for researchers in the life sciences.

A series of connected computer clusters, fully virtualised.

Locations: Amsterdam, Leiden, Radboud, Nijmegen, Maastricht, Wageningen, etc.

Traditional grid cluster on top.

Virtualisation: OpenStack/Ceph based.

Life Science Grid Community

Grid Services Park

Intended for internal services and SaaS

Grid services: UIs, WMS, VOMS, BDII, MyProxy, LFC.

Generic services: Picas (CouchDB), Topos, Softdrive (CVMFS).

SaaS services: e-Ecology, TwiNL.

Dedicated UIs for big projects: LOFAR, TROPOMI, ProjectMine.

Virtualisation: OpenStack/Ceph based.

SURF SARA

Elixir

"A distributed infrastructure
for life-science information"

Responsibilities SURFsara within the Elixir - EXCELERATE project:

WP4: Compute, Data access and Exchange Services

WP9: Secure Archiving, Dissemination and Analysis of Human Access-controlled Data

WP11: ELIXIR Training Programme

Helix Nebula Science Cloud

"The Science Cloud"

Goal: European hybrid cloud platform that will support high-performance, data-intensive scientific use-cases

Pre-Commercial Procurement (PCP) tender

Buyers: 10 institutes

Suppliers: 3 consortia in the current prototype phase

Main challenges:
- Transparent data access
- Federated AAI
- Defining expectations

SURFsara Research cloud

"Long tail of science"

SURFsara internal development for future services.

Goal: ready-to-go scientific computing environment for ad-hoc collaboration.

Parts:

Catalog for services

Catalog for data sets

Connected resources

- Local at SURFsara
- Public clouds
- Resources at members

Controller to orchestrate a research environment on connected resources.

HTC-PaaS

High Throughput Computing platform-as-a-service

Under development.

Cloud tailored to efficiently deploy HTC-PaaS services.

1st iteration: run Grid cluster (SARA-MATRIX) without loss of performance.

2nd iteration: PaaS facilities.

Future plans: join to National HT-cloud, together with NIKHEF and RUG.

Virtualisation: OpenStack/Ceph/SWIFT based.

SURF SARA

EGI Federated Cloud

Remarkable progress since our review 2014.

Applications on Demand

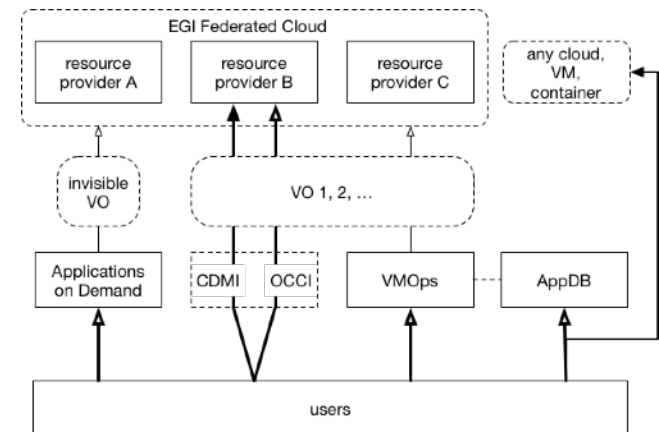Access Service for PaaS (Chipster, Galaxy and Jupyter), EC3 portal.

Promising for "the long tail of science".

VMops

Create and manage topologies of VMs.

AppDB

Download a (trusted) appliance for your own deployment.

SURF SARA

The Promise

Commercial IaaS providers are more efficient at owning and operating hardware.

- Total Cost of Ownership
- Elasticity
- Uptime and reliability

Federation solves all your problems

- Total Cost of Ownership
- Elasticity
- Collaboration between users/institutes/disciplines
- Single sign-on
- Easy user collaboration

Provider Concerns

Commercial IaaS providers are commercial.

    Different priorities.

    European tendering.

    Accounting and traceability.

    Exit: risk of lock-in, discontinued service.

Federated resources need managing.

    Ownership and responsibility of HW, SW, services.

    User authentication, authorisation, AUP — one size fits all?

    Anonymous users may gain superuser rights on your hardware and may not behave.

    Traceability of actions on behalf of end-users.

    SLA: we must do what you promise.

    Who maintains which service?

    Who is responsible that a service works on a provider?

    Synchronised updates of services.

    Helpdesk nightmare.

User Concerns

Commercial IaaS providers are commercial.

    Not optimised for research.

    Data transfer slow and/or costly.

    Confidentiality of data, algorithms, research efforts.

    Legal concerns (medical data).

Federated resources need managing.

    Consistency over providers — upload or attach data, open external ports.

    Which service runs where with what speed for what cost.

    Trust and correctness of software.

    Reproducibility of computation, traceability.

    Constraints on provider selection — my data may (not) leave the EU.

   Source and result data transfer. Backups? Petabytes?

Inevitable?

From the SURFsara long-term plan:

A federated approach to data and service delivery is integral part of the future of higher education and research community, where local, institutional, national and international services and data are provisioned, baring in mind the distributed and collaborative nature of data, ICT-resources, research equipment, and researchers, and the variety in ownership.

Translation: yes, federation is inevitable.

… and if done right, will have great benefits for users and providers.
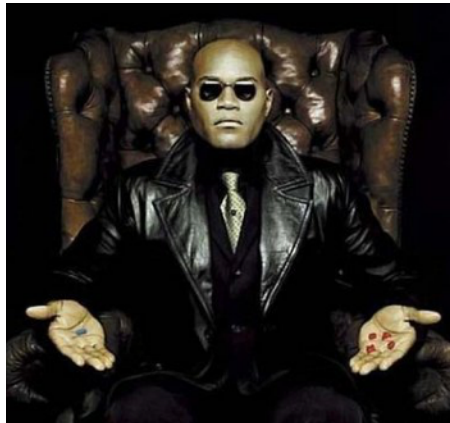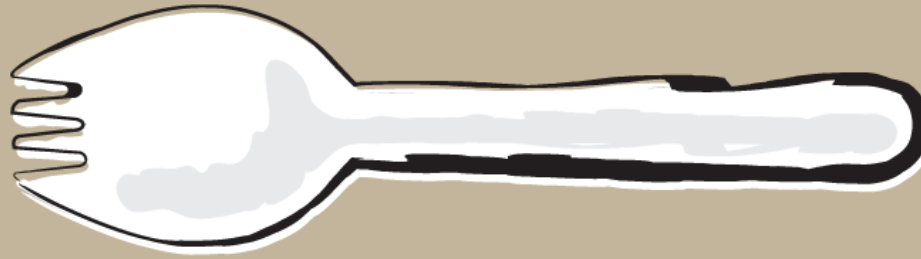
SURF SARA

Ceci n'est pas une cuillère.

Martin.Brandt@surfsara.nl

Markus.vanDijk@surfsara.nl

SURF SARA