

SURF e-Infrastructure

USER SUPPORT PROGRAMME SPACE RESEARCH GO – JUNE 20TH – THE HAGUE



Dr. Jeroen Engelberts

jeroen.engelberts@surfsara.nl



Outline

- SURF / SURFsara
- Computing Services
- Storage / Data Management
- Access to SURF Compute
- HPC Trainings/courses
- If time permits:
 - Discussion/Possible Use cases

SURF: National e-infrastructure in the Netherlands



SURFsara

- 1971 – founded as SARA
- 1984 – first national supercomputer
- 1995 – independent
- 2008 – split Vancis / SARA
- 2012 – new name SURFsara
- 2013 – merger to co-operative SURF

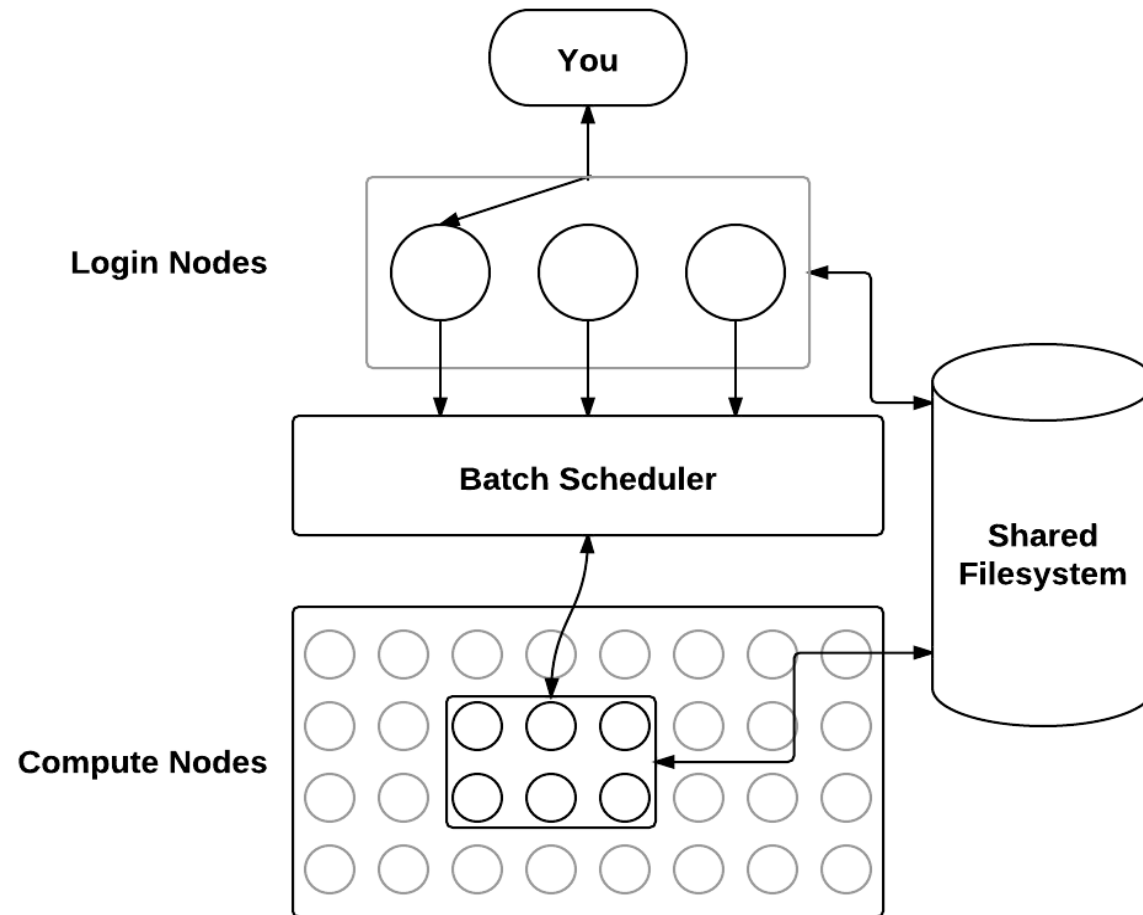
Computing services

Service	Intended for	Examples	Required experience
Cartesius	Large scale analysis, capability computing	Climate models Machine learning	**
LISA (cluster)	Batch processing	Genomic imputations	**
HPC Cloud	Interactive applications, service hosting,	Galaxy, interactive visualisations	*
Grid	Parameter sweeps, trivial parallel applications	Genomic alignment	*****
Big data services	Exploration of large amounts of data	Searching web archives, Twitter analysis	**

Differences Lisa / Cartesius compared to own PC

Cartesius / Lisa	Own computer
Linux	Windows, macOS, Linux (maybe)
Shared with others	Single user access
Access via secure shell (SSH)	Access directly
Mostly text based (Terminal)	Mostly Graphical User Interfaces
Compute/Analyses tasks via queue	Compute/Analyses tasks directly

Schematic overview of Lisa / Cartesius



HPC Cloud

When to use HPC Cloud

- "Big Laptop" required
- More cores
- More memory
- (Limited) possibility to have Virtual machines (VMs) communicate

Workflow

- Login to HPC Cloud portal (website)
 - Create or upload you Virtual Machine
 - Start it from within the portal
 - Use it interactively with a viewer (Remote Desktop)
- Check <https://doc.hpccloud.surfsara.nl/> for more information!

Big Data Services

- Scalable data analytics with Spark
- Simple but powerful interface for researchers
- Text mining, machine learning, genomics, geoinformatics
- Flexible, custom infrastructure with Kubernetes container orchestration
- SURF Jupyter Notebook service for courses and reproducible research
- Spark courses and workshops























kubernetes



jupyter

National (pre-funded) e-infrastructure allocation quota

Service	Max cpu request	Max storage request	User support	Delivery Model	Request through
Cartesius	30M SBU	100s of TBs	4 hours*	  	
LISA	3M SBU	200 GB per user	4 hours*	  	
HPC Cloud	70K core hours 2 TB		4 hours	  	
Grid	4M core hours	200 TB disk, 250 TB tape	250 hours	  	
Big data services	-	25 TB	4 hours	  	

* Separate funding opportunities available for code optimisation



B2C, Apply as individual



B2C, Apply as project



B2B, Institutional subscription

Data services

Service	Features	Intended for	Long/ short term
Data ingest service	Disk read-out system	Large data volumens transported on hard drives	Short
ResearchDrive	Webdav, ownCloud	General storage functionality, provides access to several resources	Short/ long term
SURFdrive	DropBox like	Online collaboration	Short term
SURFfilesender	WeTransfer like with encryption	Online collaboration	Short term
HPC central data archive	Tiered, Dual copies, supports NFs/ ssh/scp/gridFTP	For use with super compute facilities	Long term preservation
EUDAT B2- services (e.g. B2Share)	...	Curation/Preservation	Long term preservation
PID service	Identifier service for datasets	Publishing data	Long term

Access to SURF Compute

- Requests for (pilot) projects Lisa & Cartesius via NWO:
 - Using the ISAAC portal:
<https://www.nwo.nl/en/funding/our-funding-instruments/ew/access-to-the-national-computer-facilities/index.html>
- Direct Access to SURF e-Infra
 - <https://e-infra.surfsara.nl>
- Via your local, own university
 - <https://userinfo.surfsara.nl/systems/shared/rccs>
- Documentation
 - <http://userinfo.surfsara.nl>
- Questions? → helpdesk@surfsara.nl

Training & Support

Request information about training courses

- Introduction to big data and Hadoop
- Introduction to Unix
- Introduction to MPI
- Introduction to Intel Xeon Phi training
- Introduction to GPU programming
- Introduction to Visualisation
- Getting started with HPC Cloud
- Getting started with the Hadoop-cluster
- Getting started with Grid computing
- Getting started with the Cartesius supercomputer and the national compute cluster Lisa
- Getting started with iRODS and EUDAT datamanagement



- <http://hpc.uva.nl> (June 2018)
- <http://surfresearchbootcamp.nl> (November 2nd, 2018)
- <https://hpc.labs.vu.nl> (Autumn 2018)

SURF Research Boot Camp

Discover new possibilities
and do more with your data

15 June 2017

location: Eindhoven University of Technology

Many thanks!

