

3 B's Session 10

Introduction to HPC

Lizel Potgieter

lizel.potgieter@slu.se

Department of Plant Breeding, SLU Alnarp



Who we are

SLU bioinformatics Infrastructure

Weekly online drop-in (Wednesdays at 13.00)

slubi@slu.se, slubi.se

Alnarp: Lizel Potgieter (Dept. of Plant Breeding)

Statistics at SLU

SLU statistics center

Free consultations for all SLU staff

statistics@slu.se

Alnarp: Jan-Eric Englund and Adam Flöhr (Dept. of Biosystems and Technology)



What is HPC

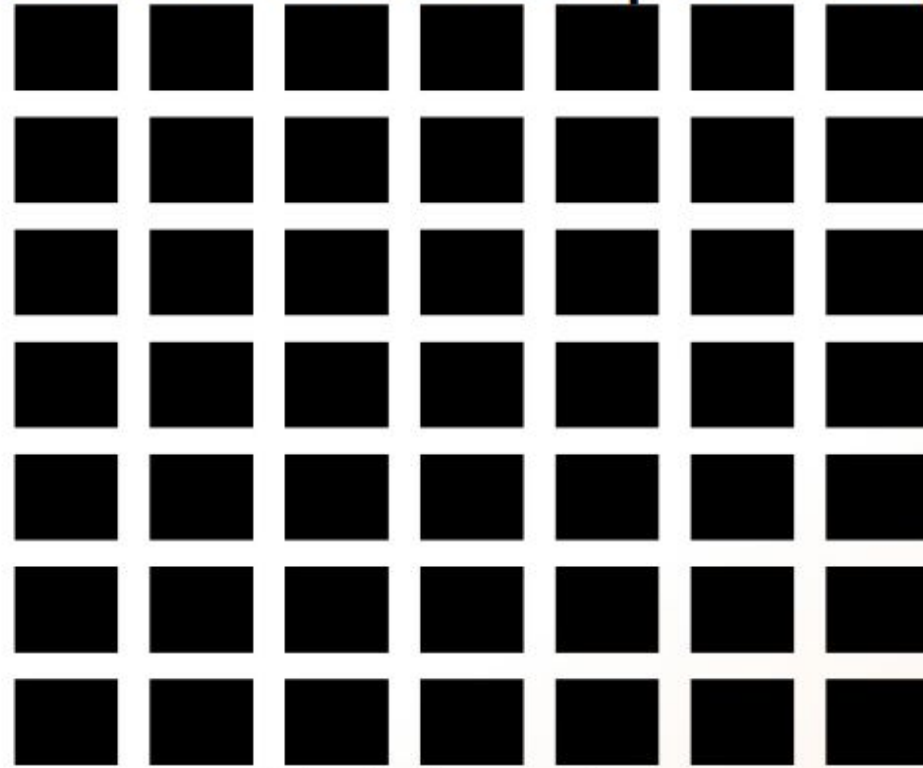
High Performance Computing most generally refers to the practice of aggregating computing power in a way that delivers much higher performance than one could get out of a typical desktop computer or workstation in order to solve large problems in science, engineering, or business.

HPC facilities in Sweden

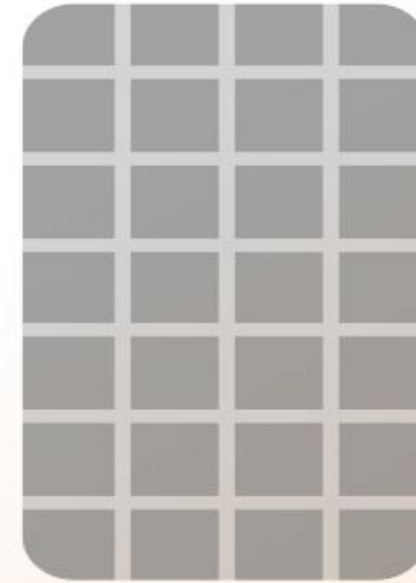
- Uppsala University
 - Uppmax
 - Rackham
 - Snowy
 - Bianca
- KTH Royal Institute of Technology
 - Dardel

Cluster name	Amount of RAM per node [GB]	Number of cores per node	Amount of RAM per core [GB]	Flag to use node
bianca	112	16	7	
	256			-C mem256GB
	512			-C mem512GB
snowy	128	16	8	
	256			-C mem256GB
	512			-C mem512GB
miarka	384	48	8	
	2048			-p fat -C 1TB
	4096			-p fat -C 4tB
rackham	128	20	6.4	
	256	16	16	
	256			-C mem256GB
	1024			-C mem1TB

The basic structure of supercomputer



Calculation nodes



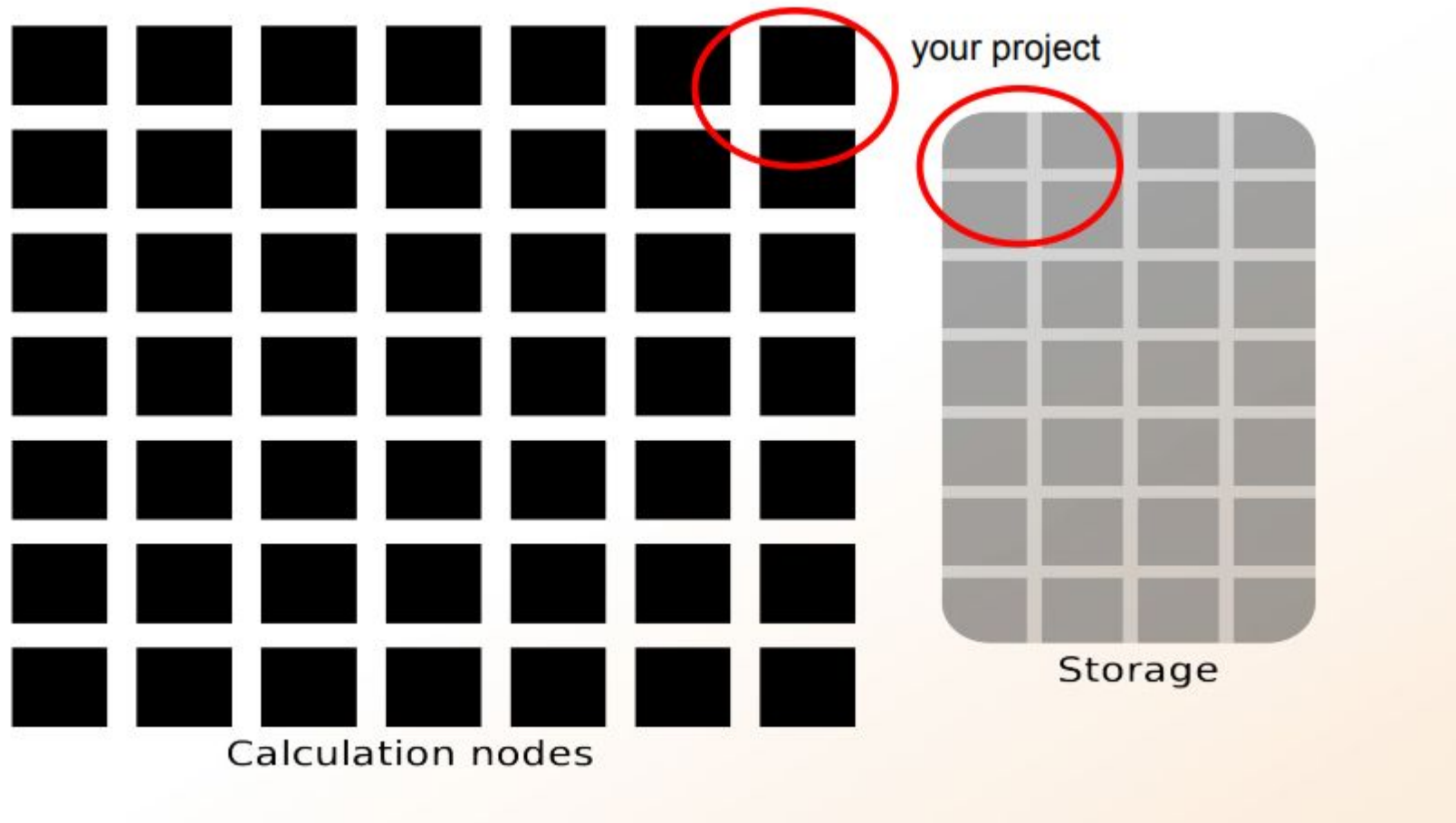
Storage



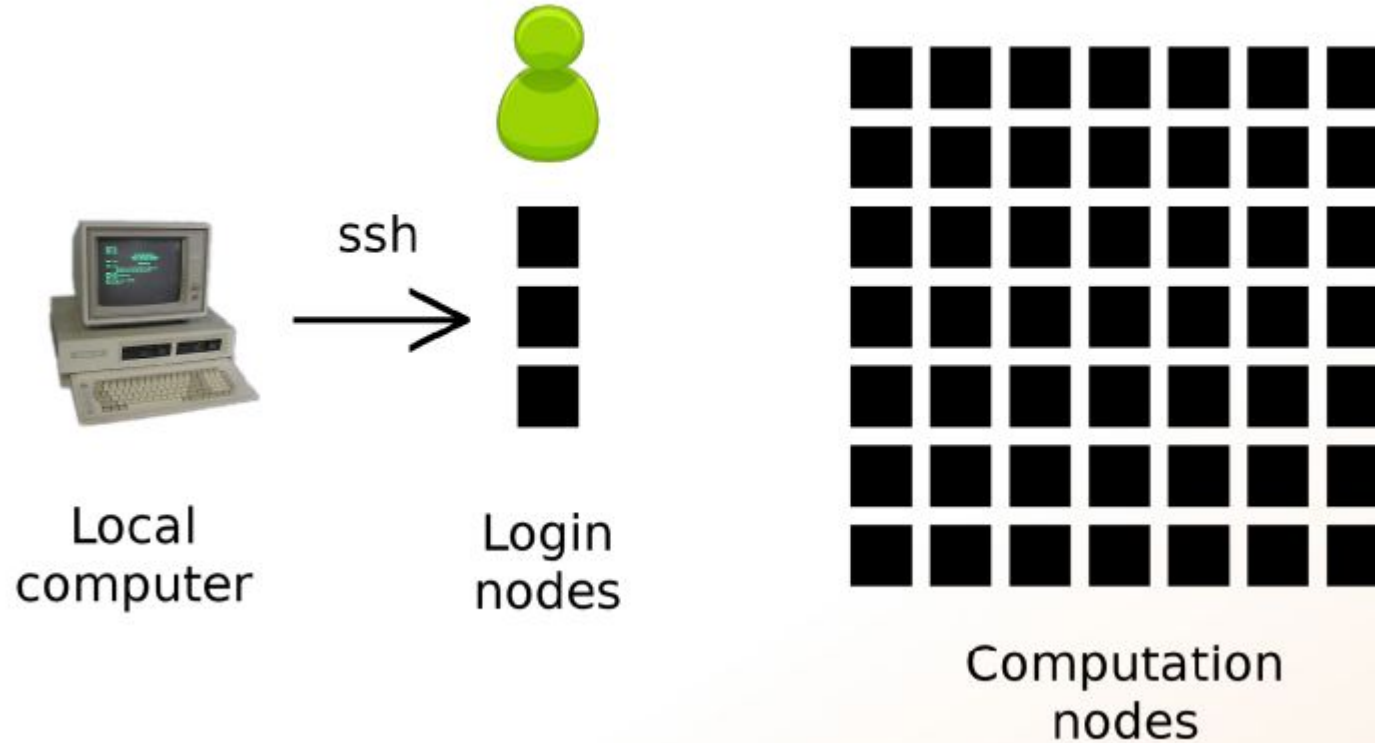
Login nodes

Compute and Storage

Why you need to apply for two different projects

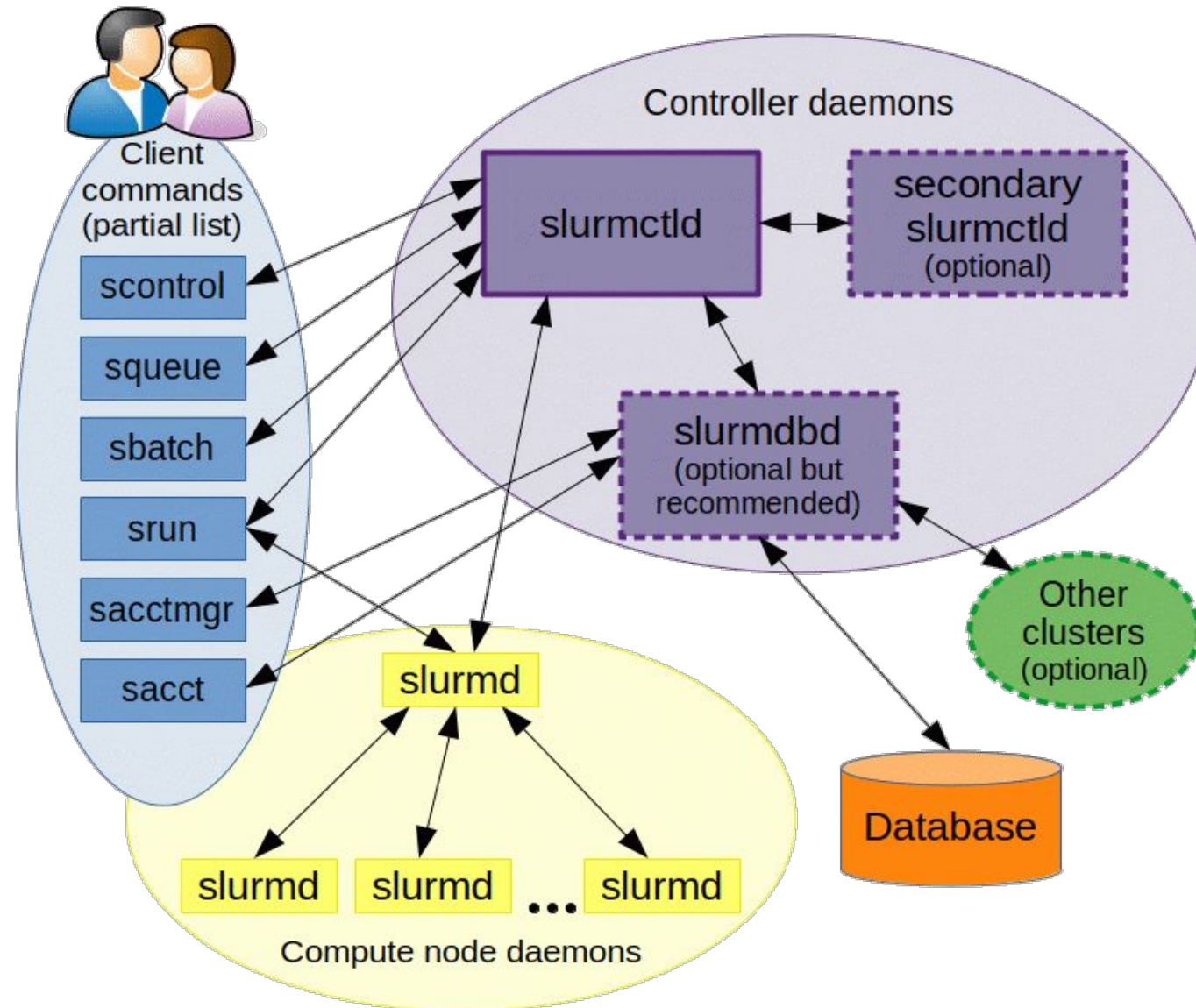


How to access an HPC



```
ssh -Y your_username@cluster_name.uppmax.uu.se
```


Slurm: Workload Manager System





interactive

Launches an interactive session that uses your computation time for the month

Runs like a normal shell

Use this when you are running computationally intensive jobs that don't run through scripts

-n is the number of cores

-t is the time you request dd:hh:mm:ss format

-A is your project

interactive -n 1 -t 2:00:00 -A project_number



sbatch

You must have this header in your script!

There are many more options supported by Slurm

-A is your project

-p is which resource you want (core or node)

-n is how many of the resource you want

-t is time

-J is the job name

Execute with **sbatch script_name.sh**

```
#!/bin/bash -l
#SBATCH -A b2010999
#SBATCH -p core -n 8
#SBATCH -t 1:00:00
#SBATCH -J mapping
```



Other commands

`jobinfo -u username`

What your job is doing (running or queue)

`scancel job_number`

Cancels that particular job (running or queue)

`finishedjobinfo`

Info about finished jobs

`projinfo -y proj_number`

Info about resource usage of project



Modules

Rackham doesn't have most programs in the path but in modules.

List: <https://www.uppmax.uu.se/resources/software/installed-software/>

Loading modules: **module load bioinfo-tools**

This must match exactly to the module name (often includes version numbers)

Loading bioinfo-tools at the start of your session is a good habit to get into as you need to load this module before other modules

Finding modules: **module spider fastqc**

Useful if you know you need to use fastqc and want to see which versions are installed



For more information

<https://www.slubi.se/uppmax.html>

https://nbisweden.github.io/workshop-ngsintro/2403/home_contents.html

<https://www.pdc.kth.se/hpc-services/computing-systems/about-the-dardel-hpc-system-1.1053338>

<https://slurm.schedmd.com/documentation.html>



SCIENCE AND
EDUCATION **FOR**
SUSTAINABLE
LIFE