

# HIGH PERFORMANCE COMPUTING AT CBS

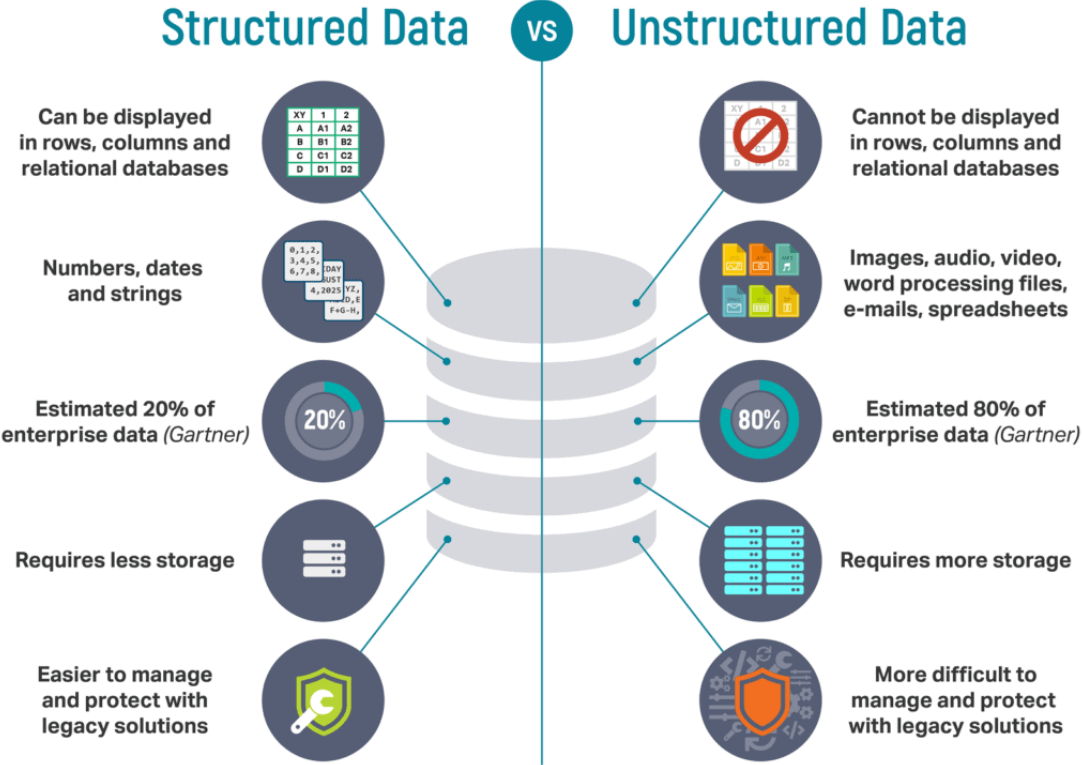
New cloud computing possibilities for researchers & students

Kristoffer Gulmark Poulsen & Lars Nondal  
Research Data Management Support  
CBS Library



# About You?

- 1. What **type and size** of data do you work with?
- 2. What programming languages do you work in (e.g. R, Python..)?
- 3. Are you familiar with parallel programming?
- 4. Are you familiar with high performance computing?



# Use Cases from CBS

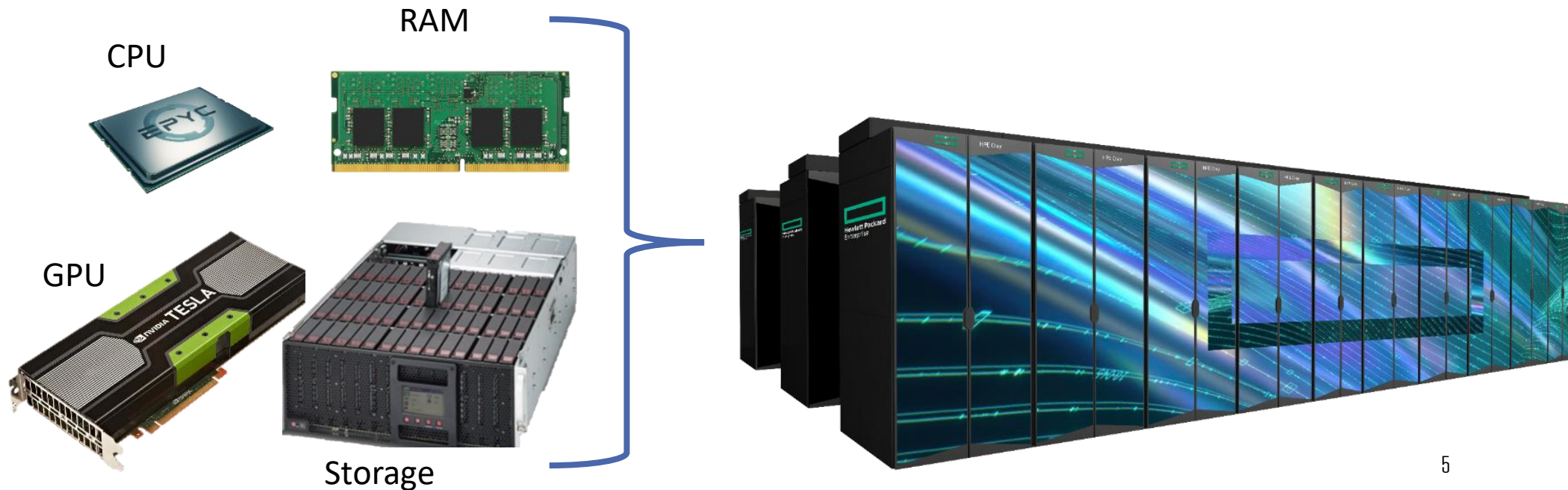
## HPC might be useful when:

- Applying ML/AI
- Running other simulation and resampling techniques
- Working with large datasets
- My laptop runs out of memory
- My workflow is running very slow

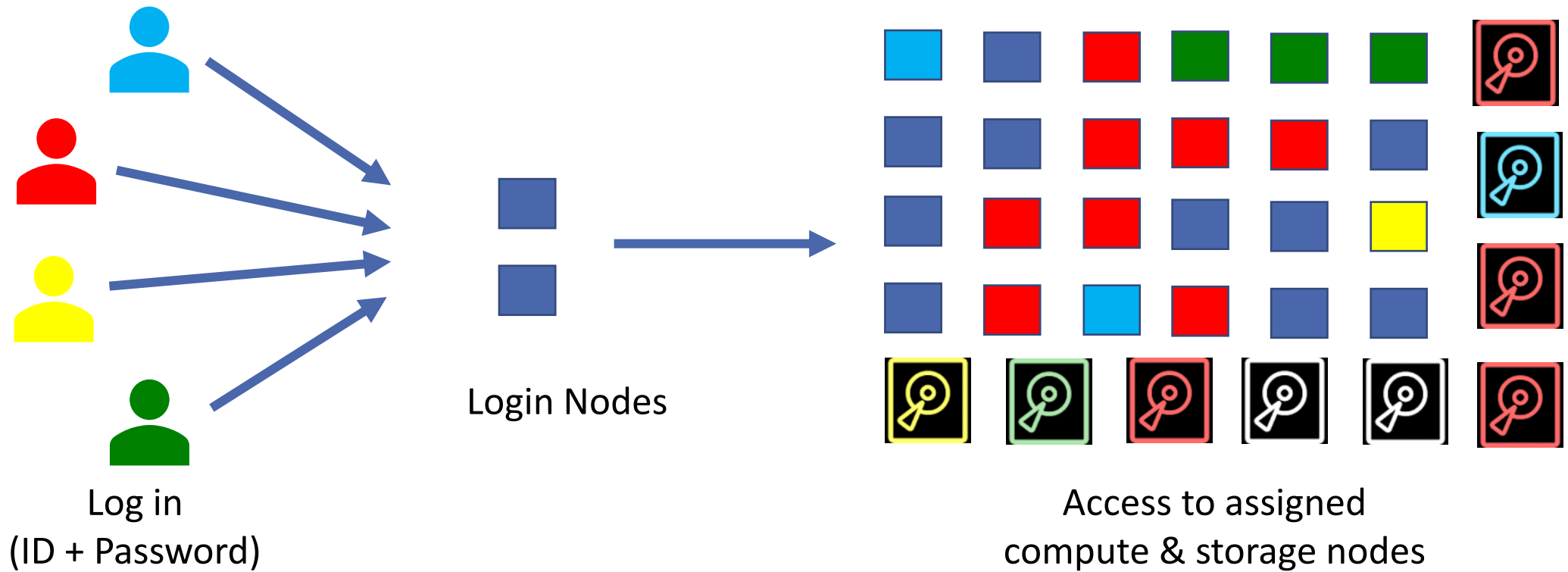
# What is High Performance Computing (supercomputer)?

## Hardware

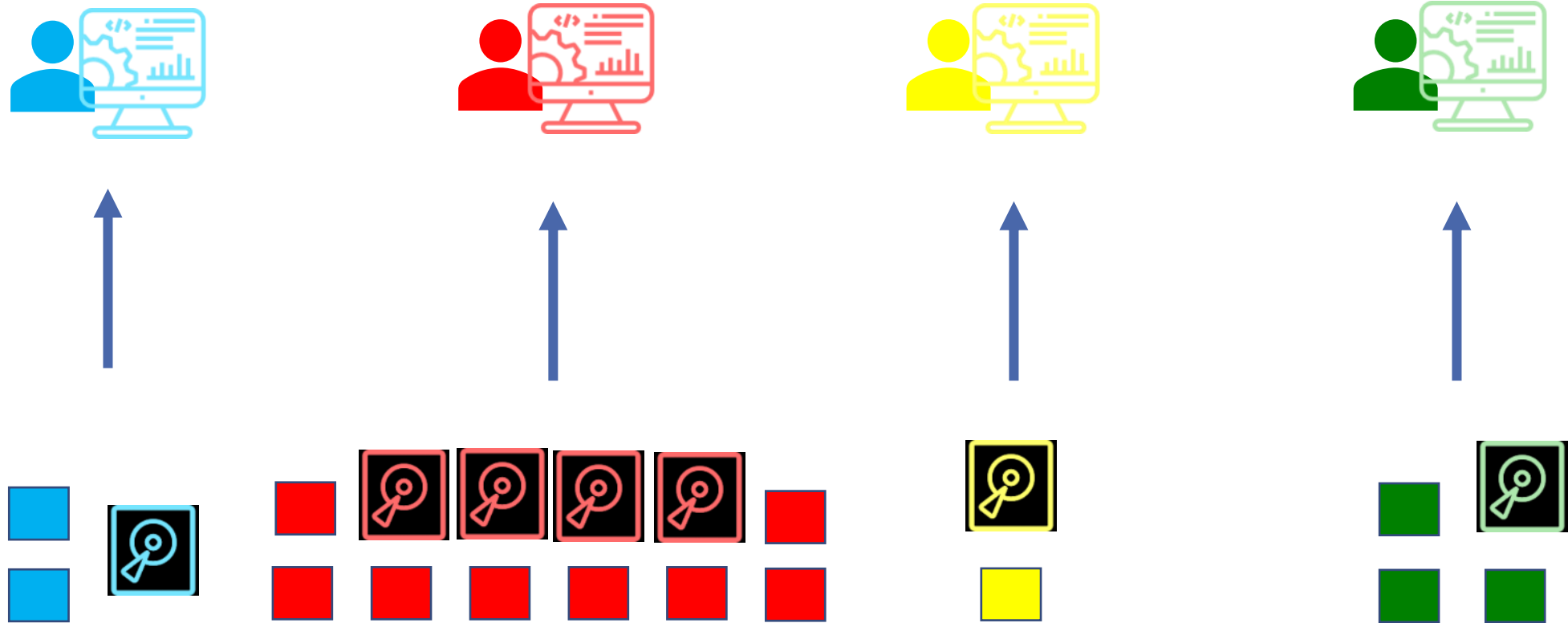
- **Core:** Processing unit on a single machine.
- **Node:** A single machine.
- **Cluster:** Network of multiple nodes.



# Accessing an HPC...



# Accessing an HPC...



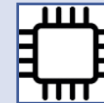
# National HPC facilities

- Collaboration between Universities and DeIC (Danish e-Infrastructure Cooperation)



Type 1 (SDU, AAU)

[Interactive HPC](#)



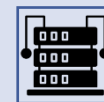
Type 2 (AU, KU & DTU)

[Throughput HPC](#)



Type 3 (SDU)

[Large Memory HPC](#)



Type 5 (EuroHPC Consortium)

[LUMI Capability HPC](#)

<https://www.deic.dk/en/supercomputing/national-hpc-facilities>

# Type 1: Interactive HPC

Cloud-based (HPC) systems (e.g. similar to google colab, amazon aws)

User friendly with Graphical User Interface (GUI).

Lots of preinstalled software (R, Python, Stata & Matlab)

Collaborative projects – work & share files with others

GDPR-Compliant

Access with university credentials from <https://cloud.sdu.dk>

- [xxx@student.cbs.dk](mailto:xxx@student.cbs.dk)
- [xxx@cbs.dk](mailto:xxx@cbs.dk)
- 1000 DKK Free credit.



Type 1 (SDU, AAU)

[Interactive HPC](#)



Deic  
Type 1

WAYF  Login  
More login options

UCloud

AALBORG UNIVERSITY DENMARK AARHUS UNIVERSITY SDU










# Type 1: SDU

- CPU resources
- GUI based
- Wide range of applications
- Slurm and Spark Cluster



Type 1 (SDU, AAU)

[Interactive HPC](#)

Name	vCPU	Memory (GB)	GPU	Price
——— DeIC Interactive HPC (SDU): u1-standard ———				
 u1-standard-1	1 (Intel Xeon Gold 6130)	6	None	0,07 DKK/hour
 u1-standard-2	2 (Intel Xeon Gold 6130)	12	None	0,16 DKK/hour
 u1-standard-4	4 (Intel Xeon Gold 6130)	24	None	0,33 DKK/hour
 u1-standard-8	8 (Intel Xeon Gold 6130)	48	None	0,67 DKK/hour
 u1-standard-16	16 (Intel Xeon Gold 6130)	96	None	1,36 DKK/hour
 u1-standard-32	32 (Intel Xeon Gold 6130)	192	None	2,74 DKK/hour
 u1-standard-64	64 (Intel Xeon Gold 6130)	384	None	5,49 DKK/hour

# Support at CBS

## Local CBS support

- Lars Nondal & Kristoffer Gulmark Poulsen
- Contact: [rdm@cbs.dk](mailto:rdm@cbs.dk) or directly to Kristoffer ([kgp.lib@cbs.dk](mailto:kgp.lib@cbs.dk))

**User support:** Advising and granting resources, technical problems.

**Consultation:** Code development etc.

**Teaching:** “[High Performance Computing](#)”, “[HPC & Parallel Programming in R](#) and [Python](#)” and “[Train your ML/AI Model on GPUs](#)”.

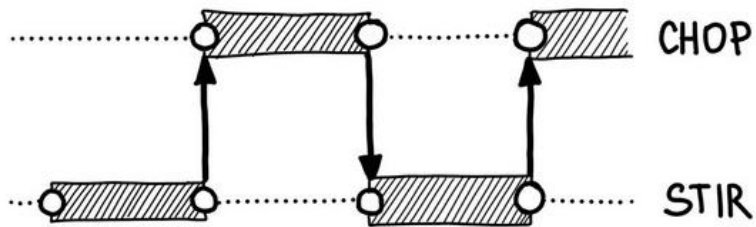
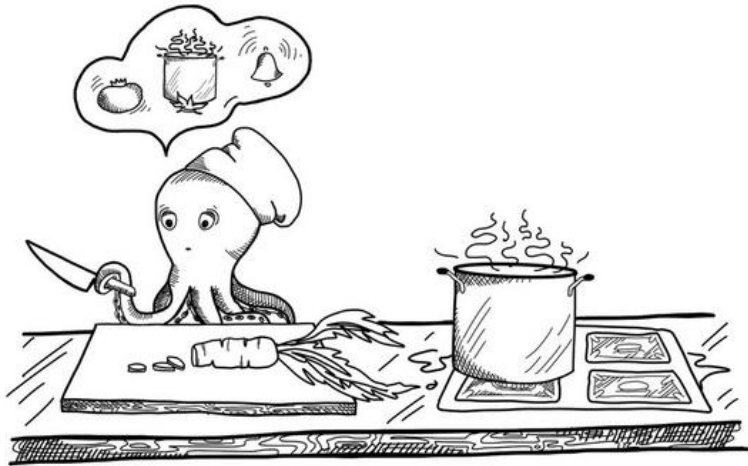
Documentation and Tutorials: <https://cbs-hpc.github.io/>

# PARALLEL PROGRAMMING

# Working on Laptop vs HPC

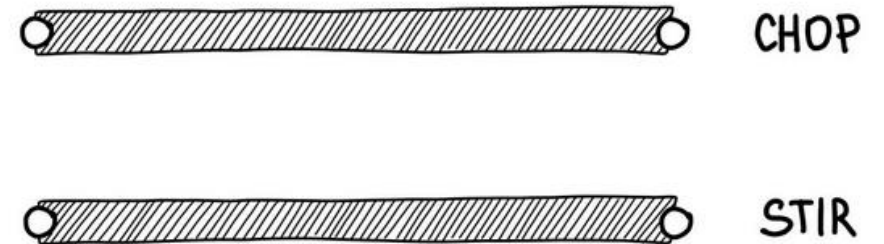
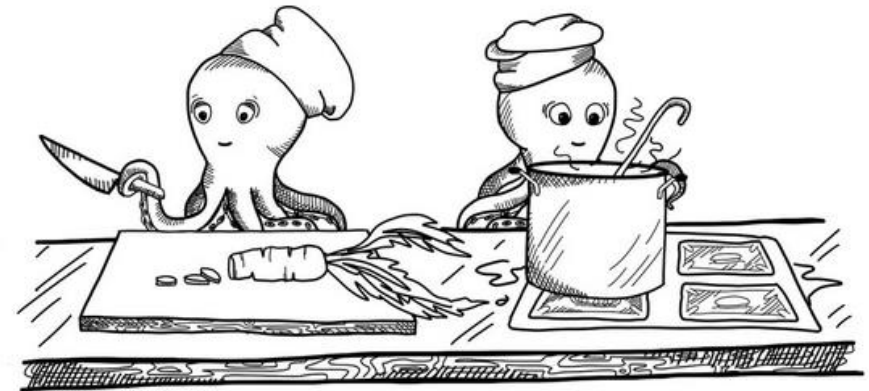
## Sequential Computing

- Single core processor
- Multiple tasks which runs overlapping but **not** at same time
- Synchronous tasks



## Parallel Computing

- Multi-core processor
- Multiple tasks which runs overlapping.
- Synchronous/Asynchronous



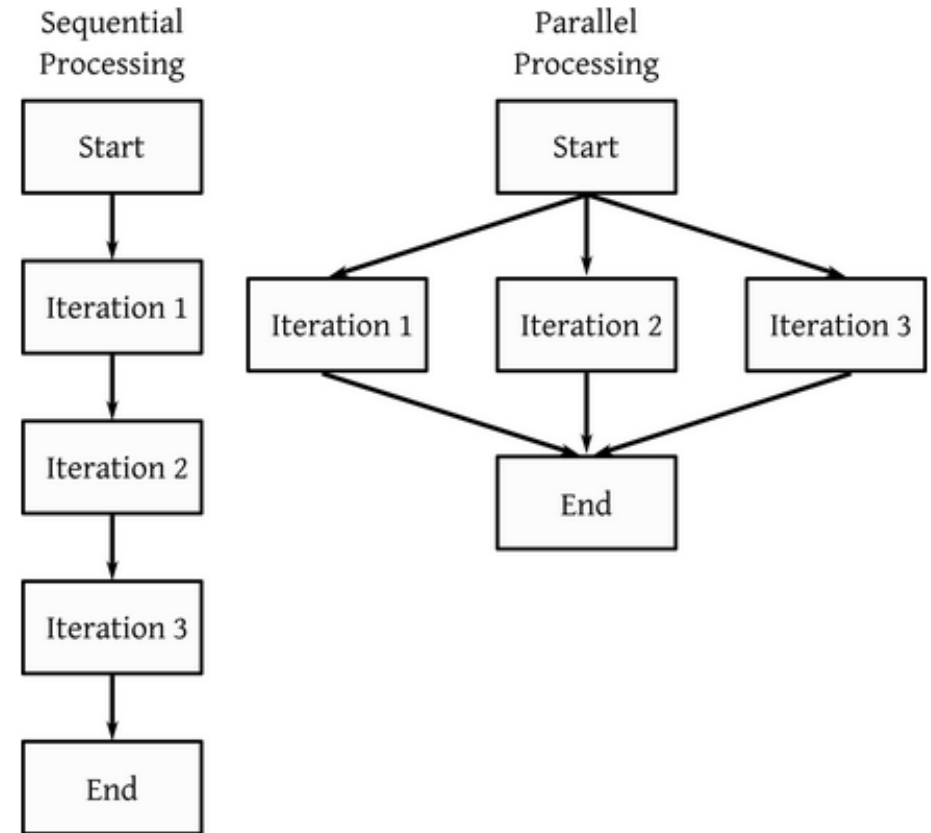
# Parallel Programming

**Single Instruction, Multiple Data (SIMD)** - single thread/processor where same each processing unit performs the same instruction on different data.

**Used in Vectorization.**

**Shared Memory Parallelism (SMP)** work is divided between **multiple threads/processes** running on a **single machine**.

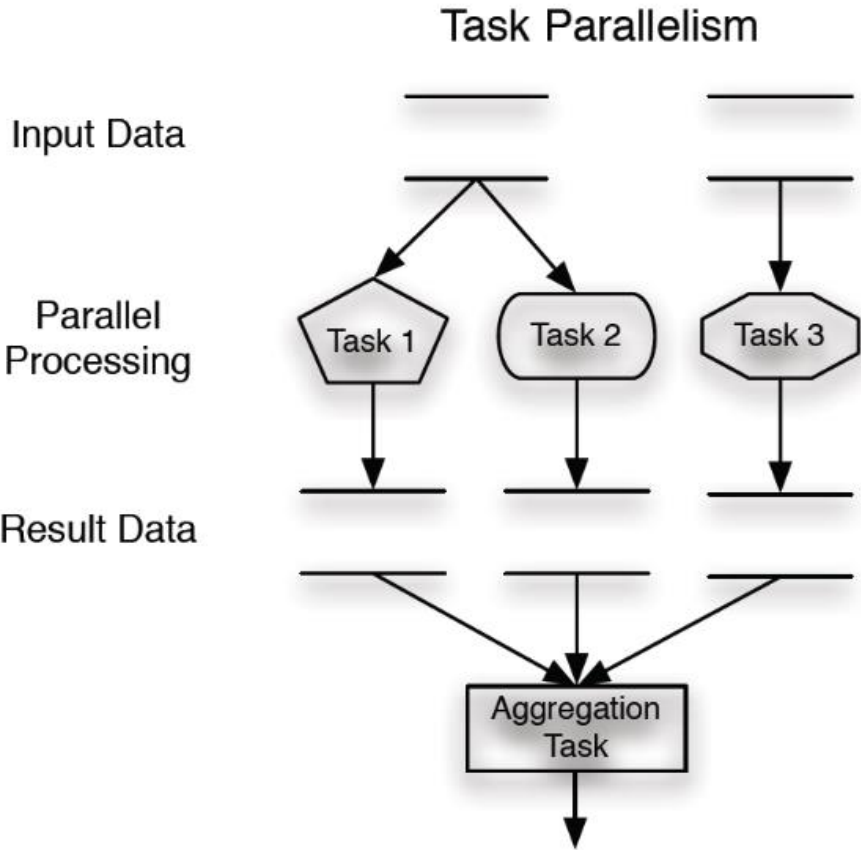
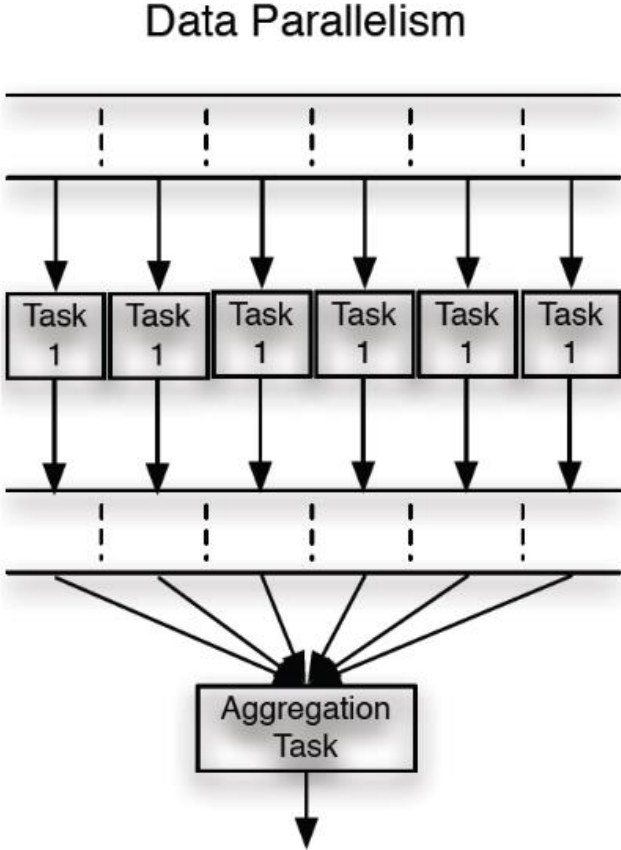
**Distributed Memory Parallelism** work is divided between **multiple machines** with its own private memory.



# Parallel Programming

## Data vs Task Parallelism

## Multi-Threading vs Multi-Processing

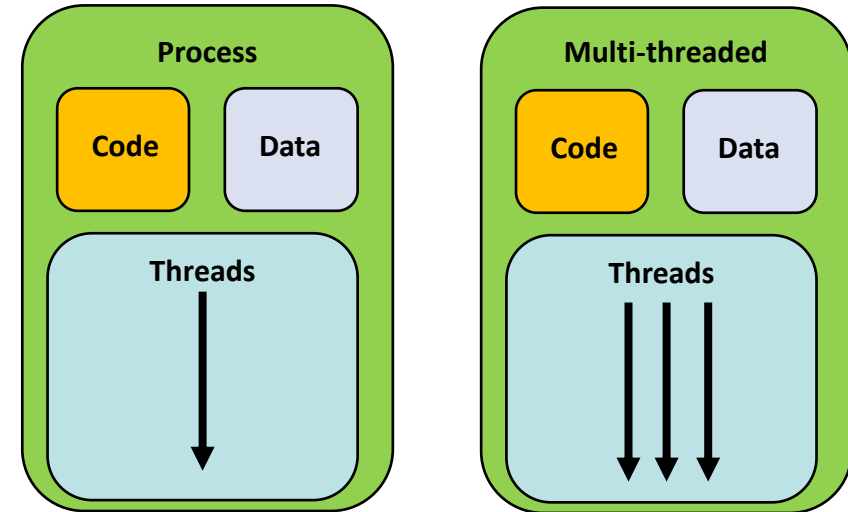


# Multi-Threading

**Threads** are multiple paths of execution within a single process.

- Appears as a single process.
- Not **hyperthreading**: a single core appears as two cores.

Python and R are examples of single-threaded programming languages.



```
top - 15:12:02 up 2 days, 54 min, 0 users, load average: 6.42, 6.45, 6.45
Tasks: 10 total, 1 running, 9 sleeping, 0 stopped, 0 zombie
%Cpu(s): 11.0 us, 0.3 sy, 0.0 ni, 88.7 id, 0.0 wa, 0.0 hi, 0.0 si, 0.0 st
MiB Mem : 385583.7 total, 193583.0 free, 102124.0 used, 89876.6 buff/cache
MiB Swap: 8192.0 total, 4461.5 free, 3730.5 used, 280235.0 avail Mem
```

PID	USER	PR	NI	VIRT	RES	SHR	%CPU	MEM	TIME+	COMMAND
243	ucloud	20	0	3970780	962704	74288	278.1	0.2	0:44.50	rsession
202	rstudio+	20	0	182200	18268	14724	0.7	0.0	0:01.00	rserver
1	ucloud	20	0	6896	3428	3196	S 0.0	0.0	0:00.05	start-rstu+
7	root	20	0	10420	4920	4376	S 0.0	0.0	0:00.00	sudo
8	root	20	0	200	4	0	S 0.0	0.0	0:00.01	s6-svscan
37	root	20	0	200	4	0	S 0.0	0.0	0:00.00	s6-supervi+
198	root	20	0	200	4	0	S 0.0	0.0	0:00.00	s6-supervi+
265	ucloud	20	0	2492	580	512	S 0.0	0.0	0:00.01	sh
271	ucloud	20	0	8168	4904	3408	S 0.0	0.0	0:00.01	bash
273	ucloud	20	0	10032	3824	3316	R 0.0	0.0	0:00.12	top

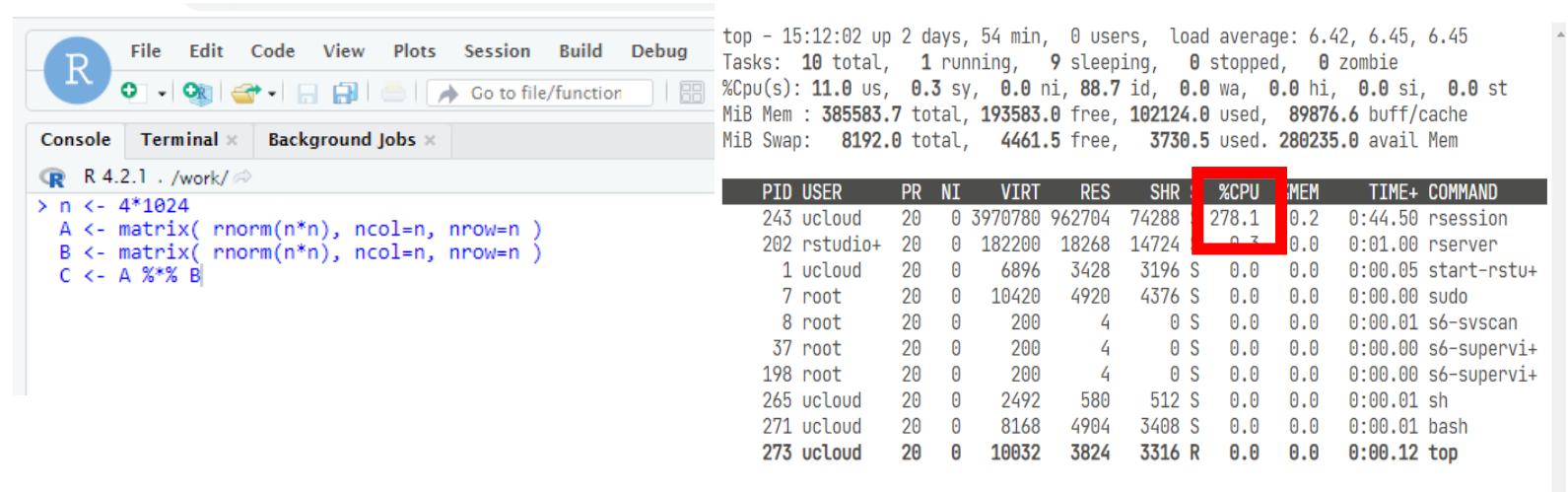
# Multi-Threading & SIMD in Python and R

**Multi-Threading** and **SIMD** is achieved through external libraries written in other languages (e.g. C, C++, Fortran) that run multi-threaded.

**Basic Linear Algebra Subprograms (BLAS)** – Allows vectorized calculations in R and Python.

**Python** e.g. NumPy and Pandas.

**R** e.g. some built-in functions.



The screenshot shows an R IDE interface. The terminal window displays the following R code:

```
R 4.2.1 . /work/  
> n <- 4*1024  
> A <- matrix( rnorm(n*n), ncol=n, nrow=n )  
> B <- matrix( rnorm(n*n), ncol=n, nrow=n )  
> C <- A %*% B
```

To the right of the terminal, the output of the `top` command is visible. The `%CPU` column is highlighted with a red box, showing a value of 278.1 for the process with PID 243.

PID	USER	PR	NI	VIRT	RES	SHR	%CPU	MEM	TIME+	COMMAND
243	ucloud	20	0	3970780	962704	74288	278.1	0.2	0:44.50	rsession
202	rstudio+	20	0	182200	18268	14724	0.3	0.0	0:01.00	rserver
1	ucloud	20	0	6896	3428	3196 S	0.0	0.0	0:00.05	start-rstu+
7	root	20	0	10420	4920	4376 S	0.0	0.0	0:00.00	sudo
8	root	20	0	200	4	0 S	0.0	0.0	0:00.01	s6-svscan
37	root	20	0	200	4	0 S	0.0	0.0	0:00.00	s6-supervi+
198	root	20	0	200	4	0 S	0.0	0.0	0:00.00	s6-supervi+
265	ucloud	20	0	2492	580	512 S	0.0	0.0	0:00.01	sh
271	ucloud	20	0	8168	4904	3408 S	0.0	0.0	0:00.01	bash
273	ucloud	20	0	10032	3824	3316 R	0.0	0.0	0:00.12	top

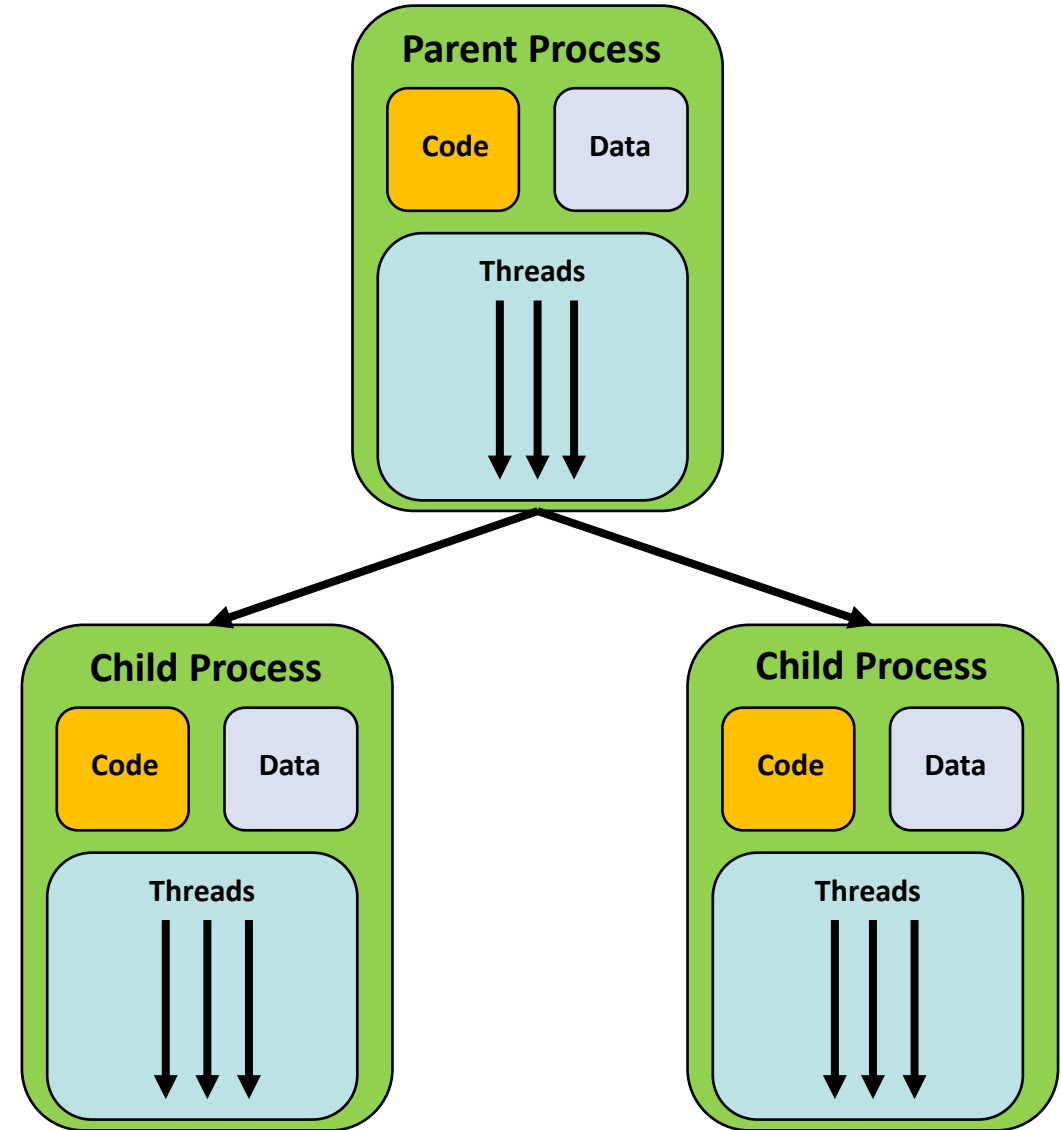


# Multi-Processing

R packages: *parallel*, *doParallel*, *future* and *Tidymodels...*

Python libraries: *multiprocessing*, *threading*, *Joblib*, *Dask* and *Ray...*

Other Frameworks: *Tensorflow*, *Torch* and *Apache Spark...*



# UCLOUD - TYPE-1 INTERACTIVE HPC

# Getting Started

**Tutorials and Documentation** <https://cbs-hpc.github.io/>

[Getting Started with HPC \(UCloud\)](#)

[Use Conda to manage R-packages and Python-libraries](#)

[Batch Processing on Ucloud](#)

[Rsync - Large data transfer to UCloud](#)

[Synchronization on UCloud](#)

[Tutorials and Documentation](#)

# UCloud Dashboard

The dashboard is titled "UCloud" and is for "Type 1 - CBS". It features a search bar and a navigation sidebar with "Files", "Projects", "Resources", "Apps", and "Runs".

- News:** "Maintenance on the DeiC Interactive HPC (AAU) provider" on 13/03/2023 at 08:00. Affects machine types: uc-general, uc-a10, uc-t4.
- Favorites:** No favorites.
- Recent runs:** List of jobs including affaff (MATLAB 2022a-1), machine\_01 (Ubuntu), testing (Ubuntu), aff (Ubuntu Xfce), and LAC\_live\_test (JupyterLab).
- Recent notifications:** Multiple notifications from "Type 1 - CBS" about grant applications and "Information" messages about role changes for users like EmilBegtrup and StefanBernidMöhler.
- Resource allocations:** Table of resources:

DeiC Interactive HPC (AAU) / uc-general	44.980,56 DKK
DeiC Interactive HPC (AAU) / uc-t4	494.528,06 DKK
DeiC Interactive HPC (SDU) / u1-standard	149.494,22 DKK
DeiC Interactive HPC (SDU) / sas-9-4-cbs	9997 License(s)
DeiC Interactive HPC (SDU) / STATA-MP17-CBS	997 License(s)
DeiC Interactive HPC (SDU) / STATA16-CBS-TEST	997 License(s)
DeiC Interactive HPC (SDU) / public-ip	97 Public IP(s)
- Resource usage:** Compute usage of 114,05 DKK for the past 30 days.
- Grant applications:** No recent outgoing applications.
- Providers:** DeiC Interactive HPC (SDU) and DeiC Interactive HPC (AAU) are both active.

Bottom left shows "Type 1/Type 1 - CBS" and "SDU Data Protection".

<https://cloud.sdu.dk/app/dashboard>

# User Pit Falls

**Wait in queue when starting job-** at “rush hours” you might have to wait in queue.

**The job stopped while I was working** - Remember to set enough hours for the job!

**My results disappeared** - Remember to work in the right folder!

**Remember to stop the application after use!**

# QUESTIONS ?