

HIGH PERFORMANCE COMPUTING AT CBS

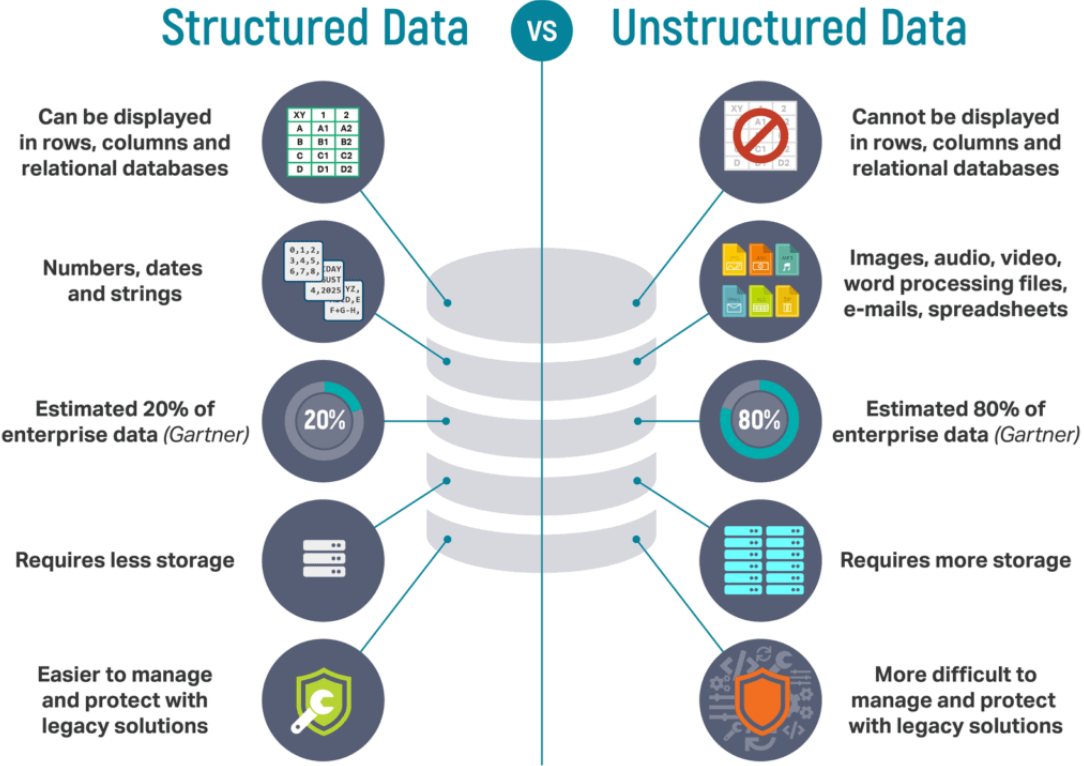
New cloud computing possibilities for researchers & students

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

“Detecting Social Media Hate Speech Surrounding Refugees Using Deep Learning”

“Asset pricing via Machine Learning”

“Money in Politics at Work: A Population-level Analysis of Employee Campaign Donations”

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

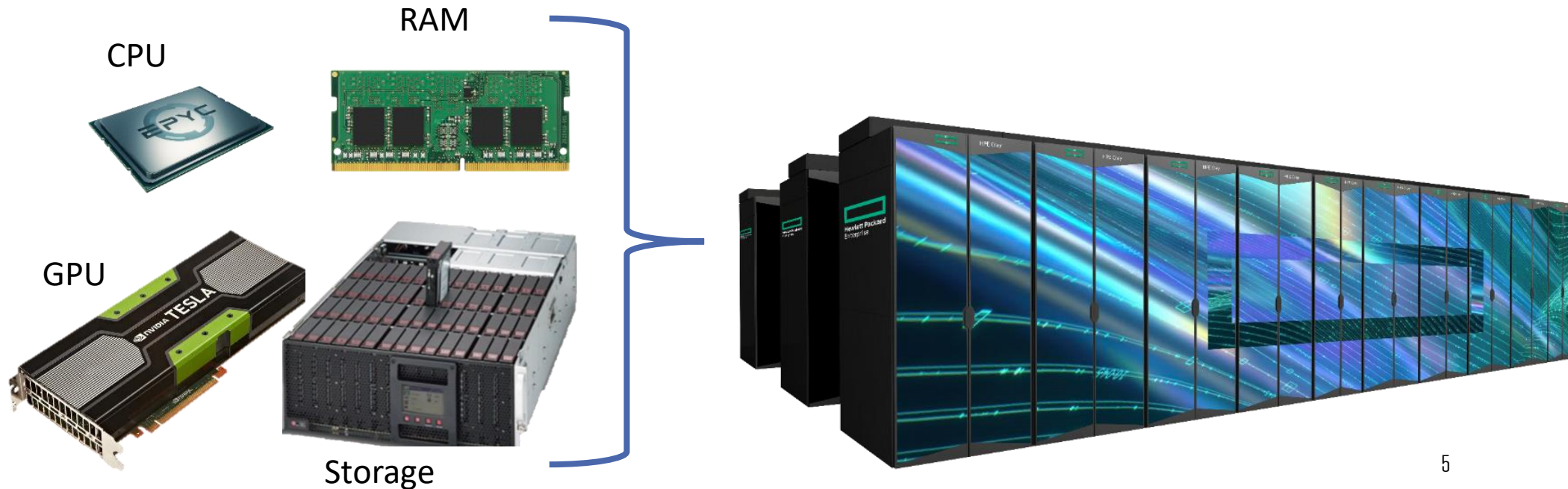
- **Core:** Processing unit on a single machine.
- **Node:** A single machine.
- **Cluster:** Network of multiple nodes.

Message Passing Interface (MPI)

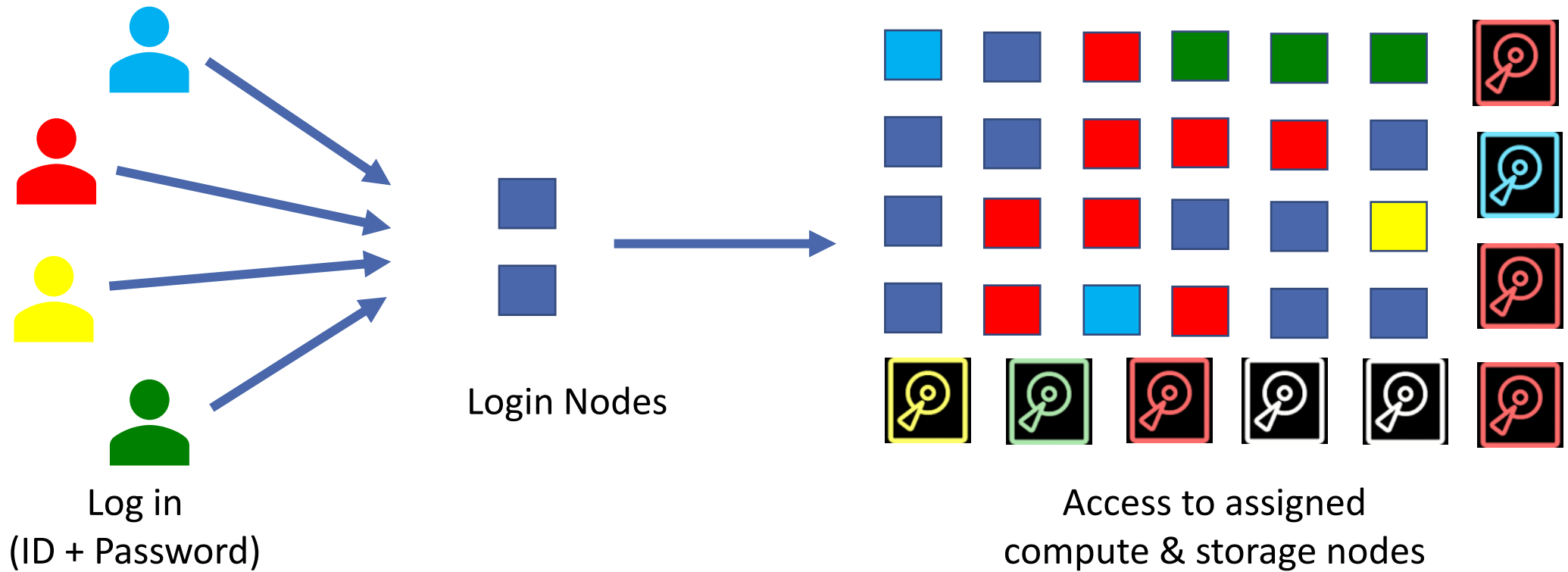
- A standard protocol for passing data and other messages between **nodes** in a **cluster**.

Simple Linux Utility for Resource Management (SLURM)

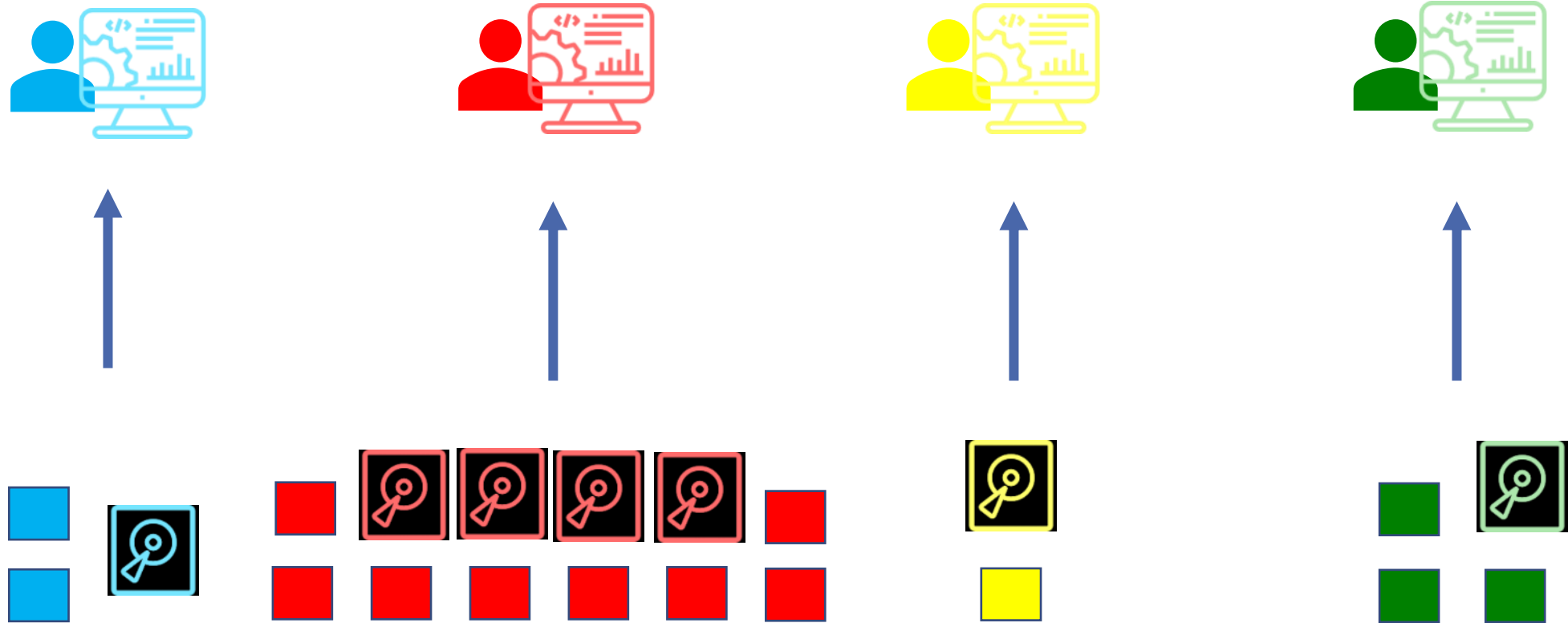
- A free MPI framework for Linux and Unix-like kernels.



Accessing an HPC...



Accessing an HPC...



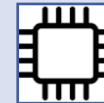
National HPC facilities

- Collaboration between Universities and DeIC (Danish e-Infrastructure Cooperation)
- **EuroHPC consortium:** Finland, Belgium, the Czech Republic, **Denmark (3%),** Estonia, Norway, Poland, Sweden and Switzerland.
- LUMI is financed 50% by the EuroHPC.



Type 1 (SDU, AAU)

[Interactive HPC](#)



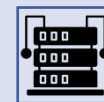
Type 2 (AU, KU & DTU)

[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
- 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 |

Type 1: AAU

- Primarily GPU resources
- Virtual Machines
- SSH/terminal Access



Type 1 (SDU, AAU)

[Interactive HPC](#)

| Name | vCPU | Memory (GB) | GPU | Price |
|--|------|-------------|------|----------------|
| ——— DeIC Interactive HPC (AAU): uc-general ——— | | | | |
| uc-general-large | 16 | 64 | None | 1,36 DKK/hour |
| uc-general-medium | 8 | 32 | None | 0,67 DKK/hour |
| uc-general-small | 4 | 16 | None | 0,33 DKK/hour |
| ——— DeIC Interactive HPC (AAU): uc-t4 ——— | | | | |
| uc-t4-1 | 10 | 40 | 1 | 8,49 DKK/hour |
| uc-t4-2 | 20 | 80 | 2 | 16,99 DKK/hour |
| uc-t4-4 | 40 | 160 | 4 | 33,99 DKK/hour |



Arch Linux (Virtual M...
latest
by AAU
VIRTUAL MACHINES



CentOS (Virtual Mac...
latest
by AAU
VIRTUAL MACHINES



Debian 11 (Virtual Ma...
11
by Debian authors
VIRTUAL MACHINES



Ubuntu (CUDA + Jup...
20.04
by AAU
VIRTUAL MACHINES



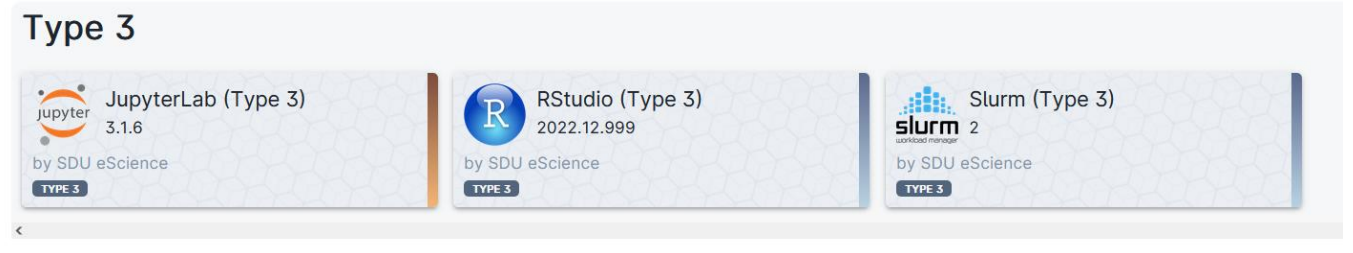
Ubuntu (CUDA) (Virt...
18.04
by AAU
VIRTUAL MACHINES



Ubuntu (Virtual Mach...
22.04
by AAU
VIRTUAL MACHINES

Type 3: Large Memory HPC

- Single node applications with large amounts of RAM memory demand (up to 4TB).
- SSH/terminal access.
- Sensitive data (ISO 27001-compliance)
- MPI (Slurm Cluster)
- Traditional/Interactive HPC



Type 3 (SDU)

Large Memory HPC

| Name | vCPU | Memory (GB) | GPU | Price |
|--|---------------------|-------------|------|-----------------------|
| — DeIC Large Memory HPC (SDU): hippo-hm1 — | | | | |
| hippo-hm1-1 | 1 (AMD EPYC 7742) | 32 | None | 1 Core hour(s)/hour |
| hippo-hm1-2 | 2 (AMD EPYC 7742) | 64 | None | 2 Core hour(s)/hour |
| hippo-hm1-4 | 4 (AMD EPYC 7742) | 128 | None | 4 Core hour(s)/hour |
| hippo-hm1-8 | 8 (AMD EPYC 7742) | 256 | None | 8 Core hour(s)/hour |
| hippo-hm1-16 | 16 (AMD EPYC 7742) | 512 | None | 16 Core hour(s)/hour |
| hippo-hm1-32 | 32 (AMD EPYC 7742) | 1024 | None | 32 Core hour(s)/hour |
| hippo-hm1-64 | 64 (AMD EPYC 7742) | 2048 | None | 64 Core hour(s)/hour |
| hippo-hm1-128 | 128 (AMD EPYC 7742) | 4096 | None | 128 Core hour(s)/hour |

Support at CBS

Local CBS support

- Lars Nondal & Kristoffer Gulmark Poulsen
- Contact: rdm@cbs.dk or directly to Kristoffer (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/>

Applying for HPC resources

CBS Front Office: Type 1, Type 3 (few resources only this year)

Deic Grant Applications: Type 1 - 3

LUMI Grant Applications: Type 5 (LUMI)

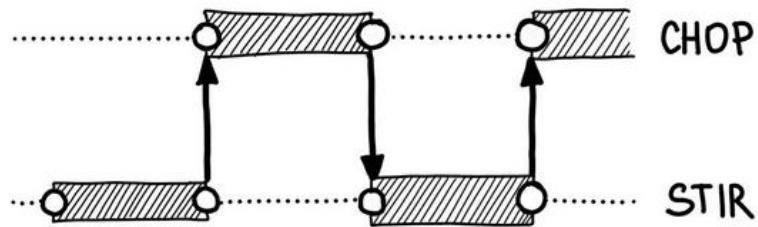
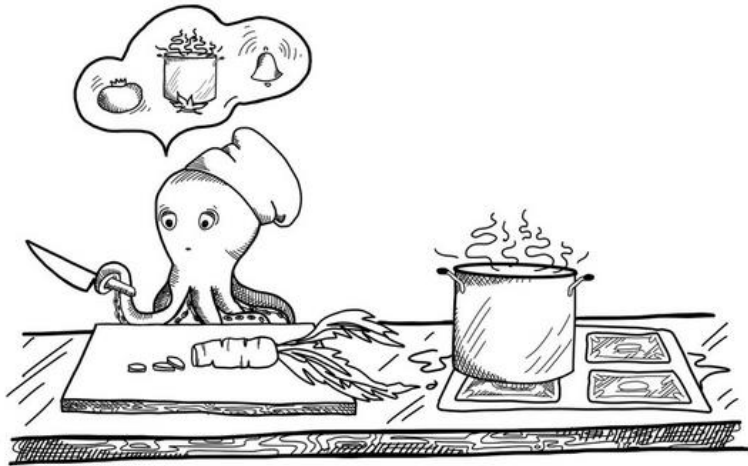
Contact: rdm@cbs.dk or directly to Kristoffer (kgp.lib@cbs.dk)

PARALLEL PROGRAMMING

Working on Laptop vs HPC

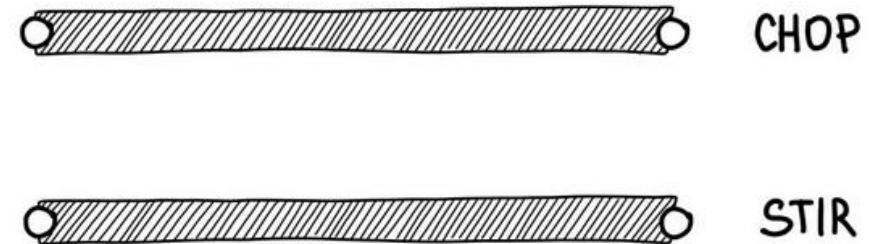
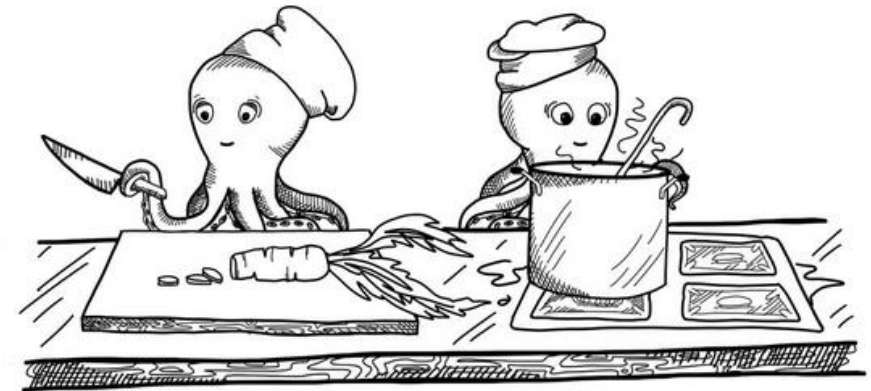
Sequential Computing

- 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

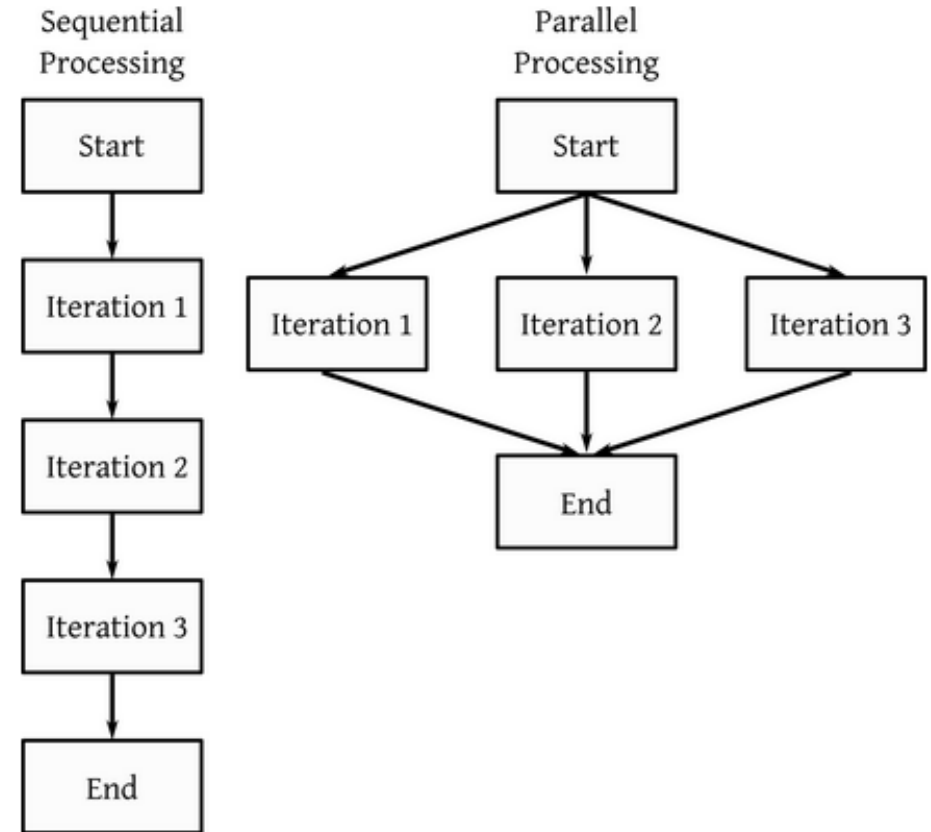


Parallel Programming

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**.

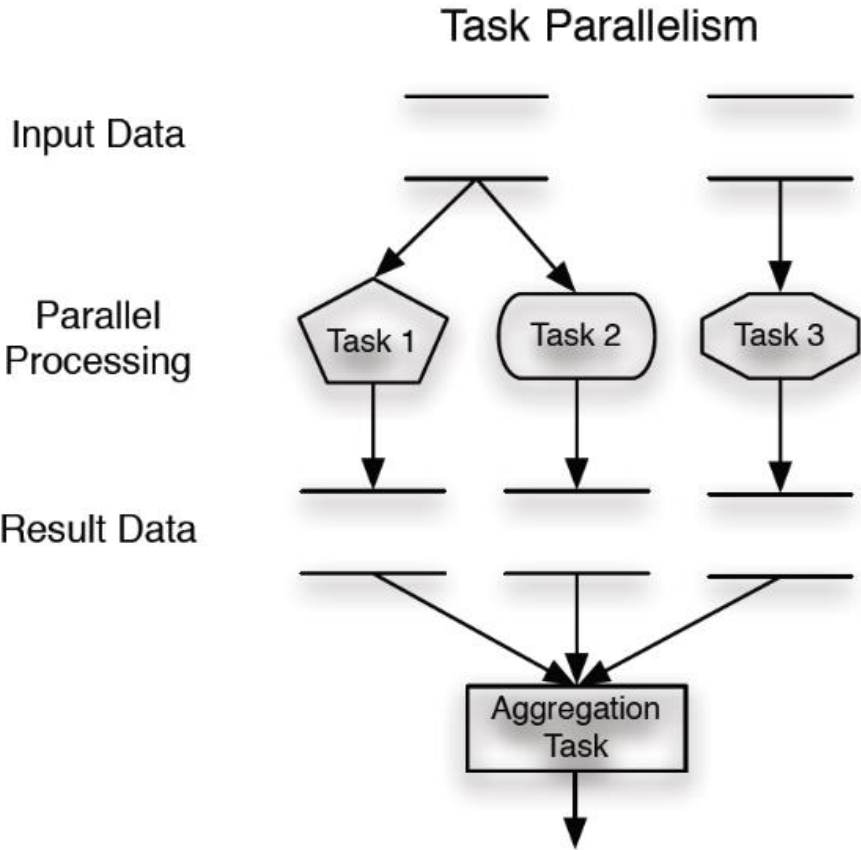
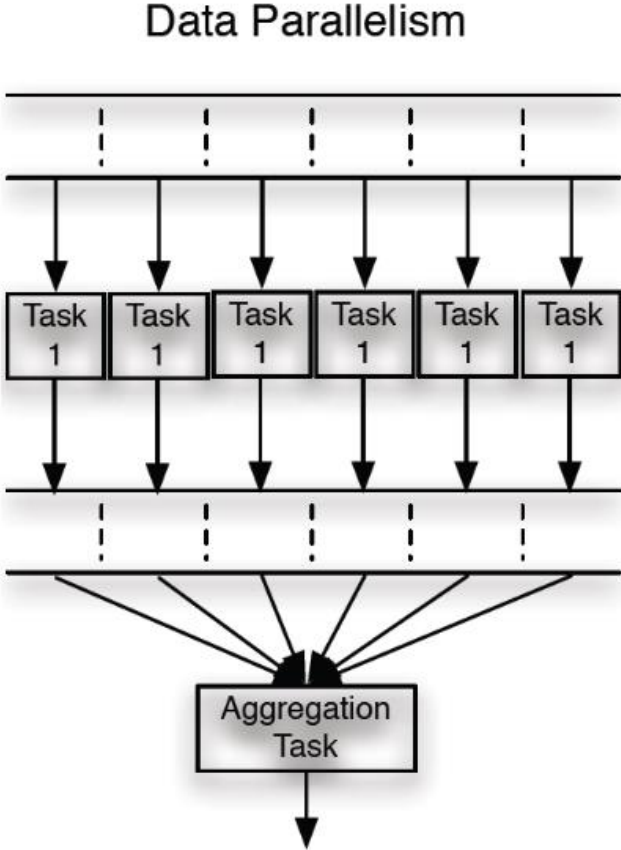
Distributed Memory Parallelism work is divided between **multiple machines** with its own private memory.



Parallel Programming

Data vs Task Parallelism

Multi-Threading vs Multi-Processing

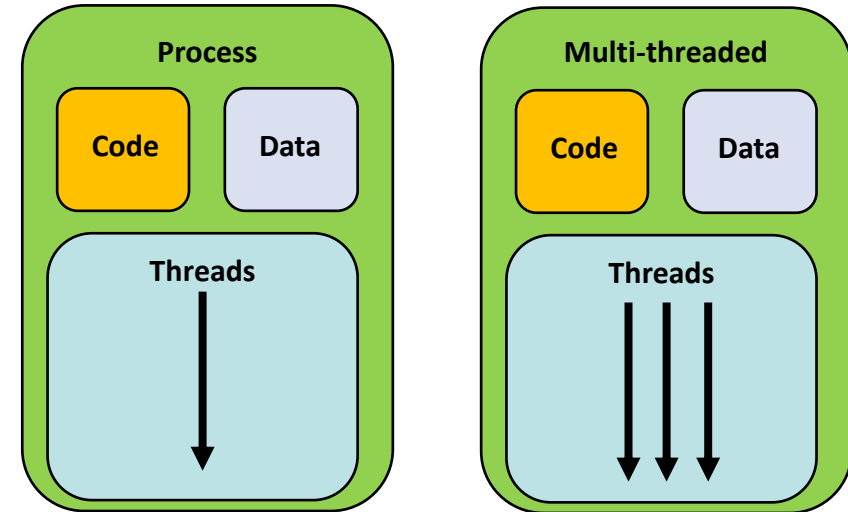


Multi-Threading

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 |

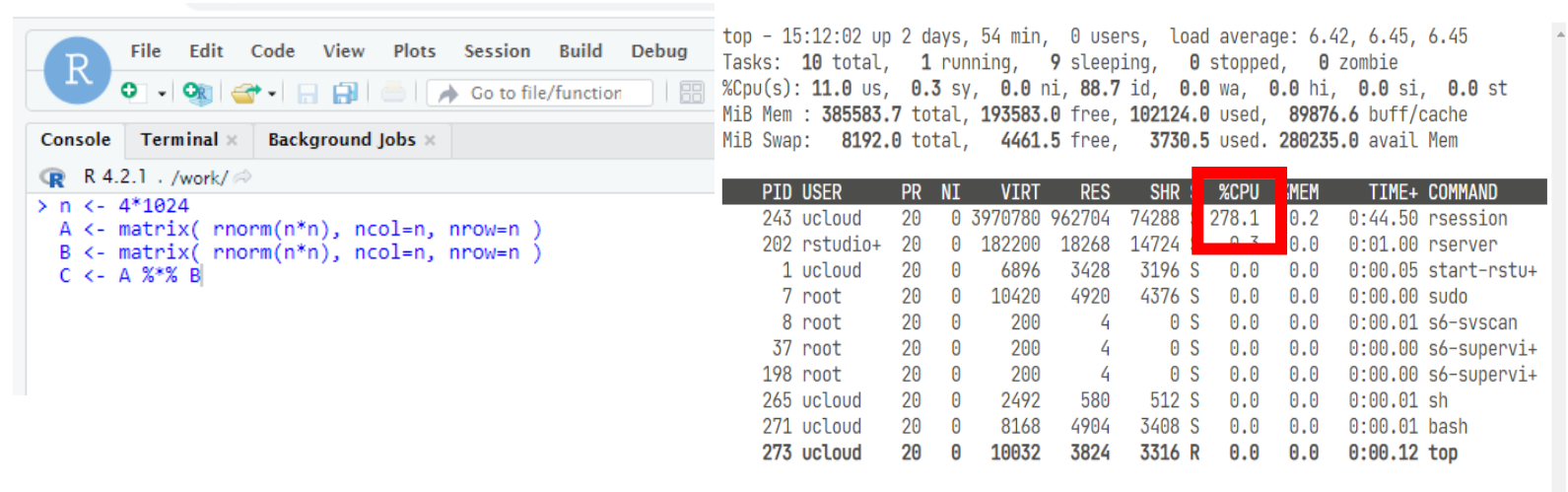
Multi-Threading & SIMD in Python and R

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 the RStudio interface. The console displays the following R code:

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

The terminal window shows the output of the `top` command, displaying system statistics and a list of processes. The `%CPU` column is highlighted in red for the first process, `rsession`, which is using 278.1% CPU.

```
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
```

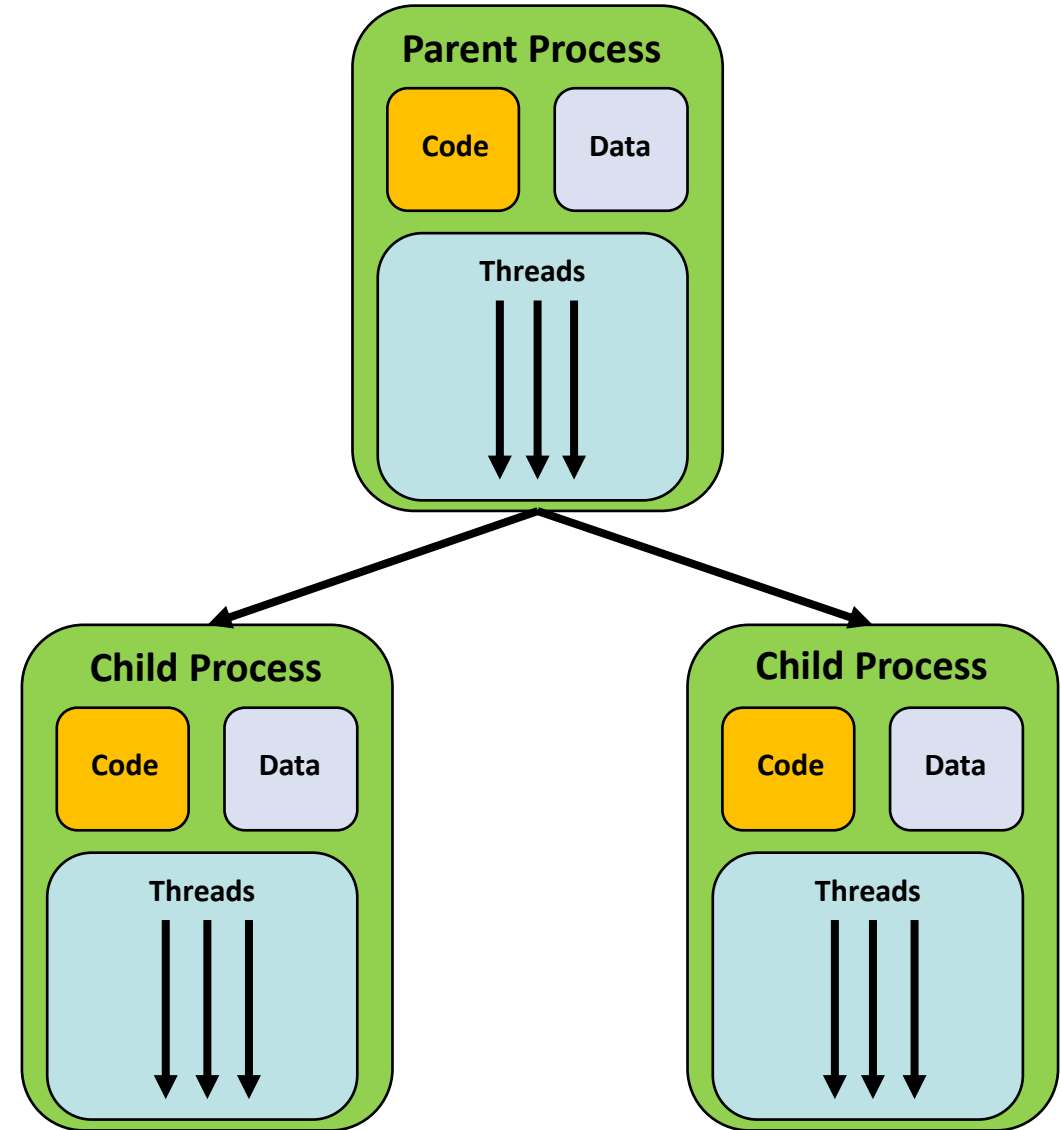
| PID | USER | PR | NI | VIRT | RES | SHR | %CPU | MEM | TIME+ | COMMAND |
|-----|----------|----|----|---------|--------|--------|-------|-----|---------|-------------|
| 243 | ucloud | 20 | 0 | 3970780 | 962704 | 74288 | 278.1 | 0.2 | 0:44.50 | rsession |
| 202 | rstudio+ | 20 | 0 | 182200 | 18268 | 14724 | 0.3 | 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

UCloud Dashboard

The dashboard features a blue header with the UCloud logo, a dropdown menu for 'Type 1 - CBS', and a search bar. A left sidebar contains navigation options: Files, Projects, Resources, Apps, and Runs. The main content area is divided into several panels:

- News:** Information about 'Maintenance on the DeiC Interactive HPC (AAU) provider' on 13/03/2023 at 08:00. It states that the provider will be performing scheduled maintenance on 21/03/23 between 09:00-15:00. A list of affected machine types is provided: uc-general, uc-a10, and uc-t4.
- Favorites:** A section indicating 'No favorites' and providing an 'Explore files' button.
- Recent runs:** A list of recent runs with status indicators (checkmarks for success, crosses for failure) and durations. Items include 'affaff' (MATLAB 2022a-1, 3 months), 'machine_01' (Ubuntu (Virtual Machine) 20.04, 5 months), 'testing' (Ubuntu (Virtual Machine) 20.04, 5 months), 'aff' (Ubuntu Xfce 22.04, 5 months), and 'LAC_live_test' (JupyterLab 3.4.5, 6 months).
- Recent notifications:** A list of notifications, including 'Type 1 - CBS' grant application reviews and 'Information' messages about role changes for users EmilBegrup-Bright#9222, StefanBernidMöhler#0398, PeterMistol#7044, and PeterMistol#7044.
- Resource allocations:** A table showing resource allocations for various HPC configurations and their associated costs or license counts.
- Resource usage:** A section showing 'Past 30 days' usage for 'Compute' at 114,05 DKK.
- Grant applications:** A section indicating 'No recent outgoing applications' and providing an 'Apply for resources' button.
- Providers:** A section showing active providers: 'DeiC Interactive HPC (SDU)' and 'DeiC Interactive HPC (AAU)', both with green checkmarks.

At the bottom left, there is a status bar with the following items: Type 1/Type 1 - CBS, KristofferGulmarkP..., UCloud Docs, and SDU Data Protection.

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



Getting Started

Documentation and Tutorials: <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](#)

Using Conda to manage Libraries



- Available for installation as “**Anaconda**” or “**Miniconda**”.
- **Package Management:** Conda helps manage software packages, allowing you to easily install, update, and remove libraries, tools, and applications.
- **Package Repositories** to store and distribute a wide range of pre-built packages and libraries for various programming languages, including Python, R, and more.
- **Dependency Resolution:** Conda can automatically handle package dependencies.
- **Virtual Environments:** It enables the creation of isolated environments where you can install specific packages and dependencies without interfering with other projects.

Using Conda to manage Libraries

```
# Shows already installed environments (Now "myenv" is available):  
which conda  
  
conda env list  
  
# Ini Conda  
conda init && bash -i  
  
# Create symbolic link for R environment between the two conda  
installations:  
sudo ln -s /work/miniconda3/envs/myenv /opt/conda/envs  
  
# Shows already installed environments (Now "myenv" is available):  
conda env list  
  
# Activate environment:  
conda activate myenv  
  
# make myenv available in Jupyterla  
python -m ipykernel install --user --name myenv --display-name "myenv"
```

UCloud Dashboard

The dashboard features a blue header with the UCloud logo, a search bar, and user profile icons. A left sidebar contains navigation options: Files, Projects, Resources, Apps, and Runs. The main content area is divided into several panels:

- News:** Information about maintenance on the DeIC Interactive HPC (AAU) provider on 21/03/23. Lists affected machine types: uc-general, uc-a10, and uc-t4.
- Favorites:** No favorites are currently listed.
- Recent runs:** A list of recent runs with status indicators (checkmarks for success, crosses for failure) and durations (e.g., 3 months, 5 months, 6 months).
- Recent notifications:** Alerts about new grant applications and role changes for users like EmilBegtrup and StefanBernidMöhler.
- Resource allocations:** A table showing various HPC configurations and their associated costs or license counts.
- Resource usage:** A summary of compute resource usage for the past 30 days, showing 114,05 DKK.
- Grant applications:** A section for outgoing applications, currently showing none.
- Providers:** A list of active providers: DeIC Interactive HPC (SDU) and DeIC Interactive HPC (AAU).

At the bottom left, there is a status bar with information about the user (Type 1/Type 1 - CBS, KristofferGulmarkP...), UCloud Docs, 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 ?