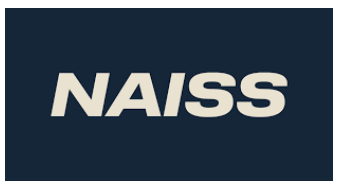
Introduction to HPC2N, Kebnekaise and HPC

Birgitte Brydsö,
Pedro Ojeda May, and others at HPC2N

HPC2N
Umeå University

21. September 2023



UMEÅ UNIVERSITY

HPC2N

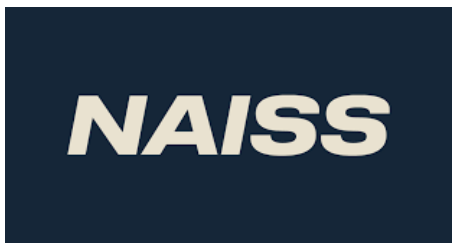
HPC2N (HPC2N at a glance)

- **High Performance Computing Center North (HPC2N)** is a competence center for Scientific and Parallel Computing

HPC2N (HPC2N at a glance)

- **High Performance Computing Center North (HPC2N)** is a competence center for Scientific and Parallel Computing

- A part of **National Academic Infrastructure for Supercomputing in Sweden (NAISS)**



HPC2N (HPC2N at a glance)

Provides state-of-the-art resources and expertise:

- Scalable and parallel **HPC**

HPC2N (HPC2N at a glance)

Provides state-of-the-art resources and expertise:

- Scalable and parallel **HPC**
- Large-scale **storage facilities** (Project storage (Lustre), SweStore, Tape)

HPC2N (HPC2N at a glance)

Provides state-of-the-art resources and expertise:

- Scalable and parallel **HPC**
- Large-scale **storage facilities** (Project storage (Lustre), SweStore, Tape)
- **Grid and cloud** computing (WLCG NT1, SNIC Cloud)

HPC2N (HPC2N at a glance)

Provides state-of-the-art resources and expertise:

- Scalable and parallel **HPC**
- Large-scale **storage facilities** (Project storage (Lustre), SweStore, Tape)
- **Grid and cloud** computing (WLCG NT1, SNIC Cloud)
- National Data Science Node in "Epidemiology and Biology of Infections" (DDLS)

HPC2N (HPC2N at a glance)

Provides state-of-the-art resources and expertise:

- Scalable and parallel **HPC**
- Large-scale **storage facilities** (Project storage (Lustre), SweStore, Tape)
- **Grid and cloud** computing (WLCG NT1, SNIC Cloud)
- National Data Science Node in "Epidemiology and Biology of Infections" (DDLS)
- Software for e-Science applications

HPC2N (HPC2N at a glance)

Provides state-of-the-art resources and expertise:

- Scalable and parallel **HPC**
- Large-scale **storage facilities** (Project storage (Lustre), SweStore, Tape)
- **Grid and cloud** computing (WLCG NT1, SNIC Cloud)
- National Data Science Node in "Epidemiology and Biology of Infections" (DDLS)
- Software for e-Science applications
- All levels of user support
  - Primary, advanced, dedicated
  - Application Experts (AEs)

HPC2N

Primary objective: to raise the national and local level of HPC competence and transfer HPC knowledge and technology to new users in academia and industry.

HPC2N (partners)

HPC2N is hosted by

UMEÅ UNIVERSITY

Partners:

HPC2N (funding and collaborations)

- Funded mainly by **Umeå University**, with contributions from the **other HPC2N partners**

HPC2N (funding and collaborations)

- Funded mainly by **Umeå University**, with contributions from the **other HPC2N partners**
- Involved in several **projects and collaborations**

HPC2N (training and other services)

- **User support** (primary, advanced, dedicated)
  - Research group meetings @ UmU
  - Also at the partner sites

HPC2N (training and other services)

- **User support** (primary, advanced, dedicated)
  - Research group meetings @ UmU
  - Also at the partner sites
- **User training and education program**
  - 0.5 – 3 days; ready-to-run exercises
  - Introduction to HPC2N and Kebnekaise
  - Parallel programming and tools (OpenMP, MPI, debugging, perf. analyzers, Matlab, R, MD simulation, ML, GPU, . . .)
    - **Using Python in an HPC environment**, 1 December 2023
    - **Introduction to Git**, 13-17 November 2023
    - **Introduction to running R, Python, and Julia in HPC**, 17-19 October 2023
    - **Workshop: Matlab in HPC**, 11, 18, 25/26 September 2023
    - **Introduction to Kebnekaise**, 21 September 2023

HPC2N (training and other services)

- **User support** (primary, advanced, dedicated)
  - Research group meetings @ UmU
  - Also at the partner sites
- **User training and education program**
  - 0.5 – 3 days; ready-to-run exercises
  - Introduction to HPC2N and Kebnekaise
  - Parallel programming and tools (OpenMP, MPI, debugging, perf. analyzers, Matlab, R, MD simulation, ML, GPU, . . .)
    - **Using Python in an HPC environment**, 1 December 2023
    - **Introduction to Git**, 13-17 November 2023
    - **Introduction to running R, Python, and Julia in HPC**, 17-19 October 2023
    - **Workshop: Matlab in HPC**, 11, 18, 25/26 September 2023
    - **Introduction to Kebnekaise**, 21 September 2023
- Workshops and seminars

HPC2N (training and other services)

- **User support** (primary, advanced, dedicated)
  - Research group meetings @ UmU
  - Also at the partner sites
- **User training and education program**
  - 0.5 – 3 days; ready-to-run exercises
  - Introduction to HPC2N and Kebnekaise
  - Parallel programming and tools (OpenMP, MPI, debugging, perf. analyzers, Matlab, R, MD simulation, ML, GPU, . . .)
    - **Using Python in an HPC environment**, 1 December 2023
    - **Introduction to Git**, 13-17 November 2023
    - **Introduction to running R, Python, and Julia in HPC**, 17-19 October 2023
    - **Workshop: Matlab in HPC**, 11, 18, 25/26 September 2023
    - **Introduction to Kebnekaise**, 21 September 2023
- Workshops and seminars
- NGSSC / SeSE & university courses

HPC2N (personnel)

Management

- Paolo Bientinesi, director
- Björn Torkelsson, deputy director
- Lena Hellman, administrator

HPC2N (personnel)

Management

- Paolo Bientinesi, director
- Björn Torkelsson, deputy director
- Lena Hellman, administrator

Application experts

- Jerry Eriksson
- Pedro Ojeda May

HPC2N (personnel)

Management

- Paolo Bientinesi, director
- Björn Torkelsson, deputy director
- Lena Hellman, administrator

Application experts

- Jerry Eriksson
- Pedro Ojeda May

Others

- Mikael Rännar (WLCG coord)
- Research Engineers under DDLS, HPC2N/SciLifeLab
  - System Developer, IT
  - Data Engineer
  - Data Steward

HPC2N (personnel)

Management

- Paolo Bientinesi, director
- Björn Torkelsson, deputy director
- Lena Hellman, administrator

Application experts

- Jerry Eriksson
- Pedro Ojeda May

Others

- Mikael Rännar (WLCG coord)
- Research Engineers under DDLS, HPC2N/SciLifeLab
  - System Developer, IT
  - Data Engineer
  - Data Steward

System and support

- Erik Andersson
- **Birgitte Brydsö**
- Niklas Edmundsson (Tape coord)
- Ingemar Fällman
- Magnus Jonsson
- Roger Oscarsson
- **Åke Sandgren**
- Mattias Wadenstein (NeIC, Tier1)
- **Lars Viklund**

HPC2N (application experts)

- HPC2N provides advanced and dedicated support in the form of **Application Experts (AEs)**:

HPC2N (application experts)

- HPC2N provides advanced and dedicated support in the form of **Application Experts (AEs)**:

  Jerry Eriksson — Profiling, Machine learning (DNN), MPI, OpenMP, OpenACC

HPC2N (application experts)

- HPC2N provides advanced and dedicated support in the form of **Application Experts (AEs)**:

Jerry Eriksson	Profiling, Machine learning (DNN), MPI, OpenMP, OpenACC
Pedro Ojeda May	Molecular dynamics, Profiling, QM/MM, NAMD, Amber, Gromacs, GAUSSIAN, R

HPC2N (application experts)

- HPC2N provides advanced and dedicated support in the form of **Application Experts (AEs)**:

Jerry Eriksson	Profiling, Machine learning (DNN), MPI, OpenMP, OpenACC
Pedro Ojeda May	Molecular dynamics, Profiling, QM/MM, NAMD, Amber, Gromacs, GAUSSIAN, R
Åke Sandgren	General high level programming assistance, VASP, Gromacs, Amber

HPC2N (application experts)

- HPC2N provides advanced and dedicated support in the form of **Application Experts (AEs)**:

Jerry Eriksson	Profiling, Machine learning (DNN), MPI, OpenMP, OpenACC
Pedro Ojeda May	Molecular dynamics, Profiling, QM/MM, NAMD, Amber, Gromacs, GAUSSIAN, R
Åke Sandgren	General high level programming assistance, VASP, Gromacs, Amber

- Contact through regular support
  - If you have a specific problem/question and/or need consultation

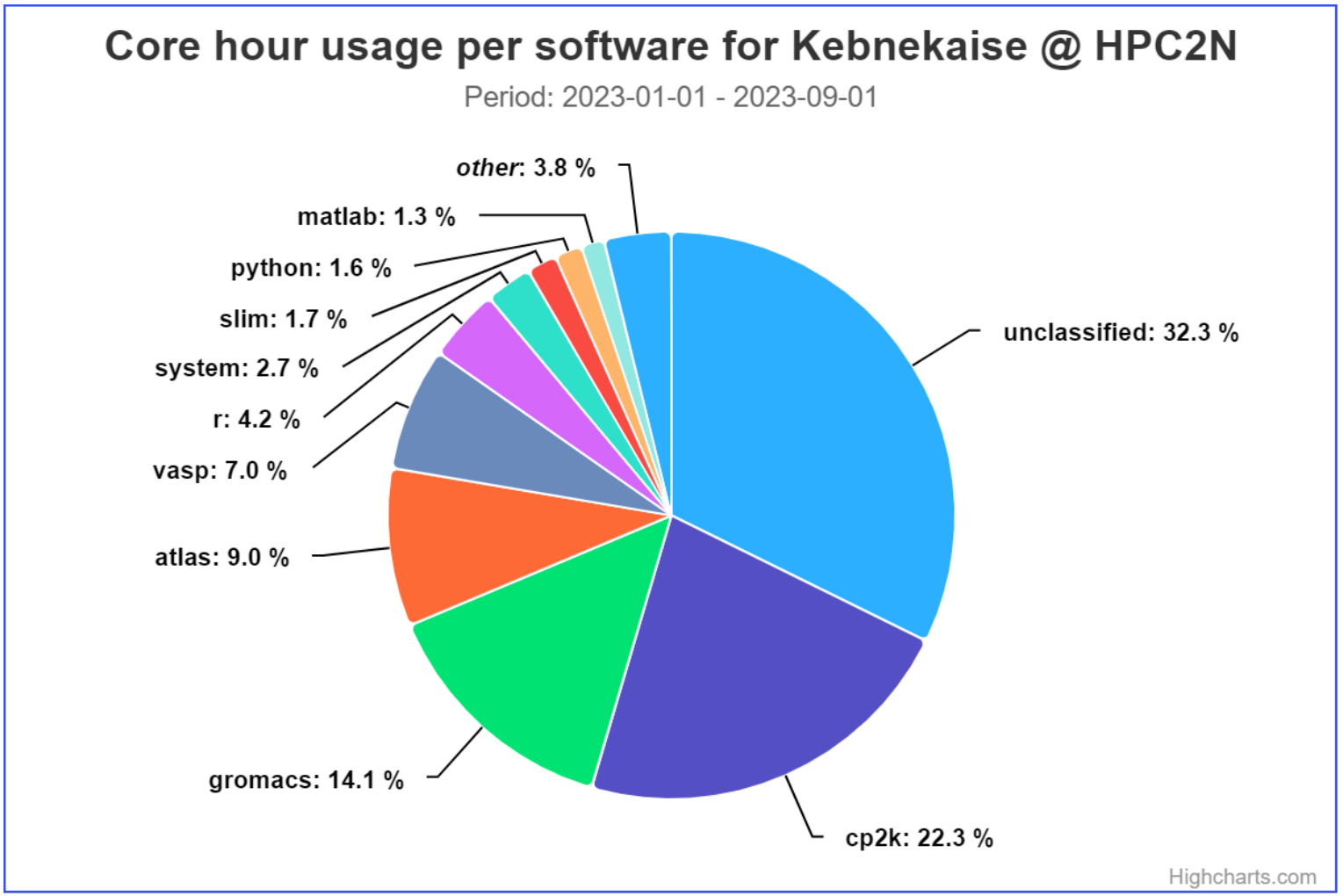
HPC2N (users by discipline)

- Users from several scientific disciplines:
  - Biosciences and medicine
  - Chemistry
  - Computing science
  - Engineering
  - Materials science
  - Mathematics and statistics
  - Physics including space physics
  - ML, DL, and other AI

HPC2N (users by discipline, largest users)

- Users from several scientific disciplines:
  - Biosciences and medicine
  - **Chemistry**
  - Computing science
  - Engineering
  - **Materials science**
  - Mathematics and statistics
  - **Physics including space physics**
  - Machine learning and artificial intelligence (several new projects)

HPC2N (users by software)



Kebnekaise

- **The current supercomputer at HPC2N**

Kebnekaise

- **The current supercomputer at HPC2N**
- Named after a massif (contains some of Sweden's highest mountain peaks)

Kebnekaise

- **The current supercomputer at HPC2N**
- Named after a massif (contains some of Sweden's highest mountain peaks)
- Kebnekaise was
  - delivered by Lenovo and
  - **installed during the summer 2016**

Kebnekaise

- **The current supercomputer at HPC2N**
- Named after a massif (contains some of Sweden's highest mountain peaks)
- Kebnekaise was
  - delivered by Lenovo and
  - **installed during the summer 2016**
- Opened up for general availability on November 7, 2016

Kebnekaise

- **The current supercomputer at HPC2N**
- Named after a massif (contains some of Sweden's highest mountain peaks)
- Kebnekaise was
  - delivered by Lenovo and
  - **installed during the summer 2016**
- Opened up for general availability on November 7, 2016
- In 2018, Kebnekaise was **extended** with
  - 52 Intel Xeon Gold 6132 (Skylake) nodes, as well as
  - 10 NVidian V100 (Volta) GPU nodes

Kebnekaise

- **The current supercomputer at HPC2N**
- Named after a massif (contains some of Sweden's highest mountain peaks)
- Kebnekaise was
  - delivered by Lenovo and
  - **installed during the summer 2016**
- Opened up for general availability on November 7, 2016
- In 2018, Kebnekaise was **extended** with
  - 52 Intel Xeon Gold 6132 (Skylake) nodes, as well as
  - 10 NVidian V100 (Volta) GPU nodes
- In 2023, Kebnekaise was **extended** with
  - 2 dual NVIDIA A100 GPU nodes
  - one many-core AMD Zen3 CPU node

Kebnekaise (compute nodes)

Name	#	Description
Compute-AMD Zen3	1	AMD Zen3 (EPYC 7762), 2 x 64 cores, **1 TB**, EDR Infiniband

Kebnekaise (compute nodes)

Name	#	Description
Compute-AMD Zen3	1	AMD Zen3 (EPYC 7762), 2 x 64 cores, **1 TB**, EDR Infiniband
Compute-skylake	**52**	Intel Xeon Gold 6132, 2 x 14 cores, **192 GB**, EDR Infiniband, **AVX-512**

Kebnekaise (compute nodes)

Name	#	Description
Compute-AMD Zen3	1	AMD Zen3 (EPYC 7762), 2 x 64 cores, **1 TB**, EDR Infiniband
Compute-skylake	**52**	Intel Xeon Gold 6132, 2 x 14 cores, **192 GB**, EDR Infiniband, **AVX-512**
Compute !!! Being phased out !!!	**432**	Intel Xeon E5-2690v4, **2 x 14 cores**, **128 GB**, FDR Infiniband

Kebnekaise (compute nodes)

Name	#	Description
Compute-AMD Zen3	1	AMD Zen3 (EPYC 7762), 2 x 64 cores, **1 TB**, EDR Infiniband
Compute-skylake	**52**	Intel Xeon Gold 6132, 2 x 14 cores, **192 GB**, EDR Infiniband, **AVX-512**
Compute !!! Being phased out !!!	**432**	Intel Xeon E5-2690v4, **2 x 14 cores**, **128 GB**, FDR Infiniband
Large Memory	20	Intel Xeon E7-8860v4, **4 x 18 cores**, **3072 GB**, EDR Infiniband

Kebnekaise (GPU nodes)

Name	#	Description
2 x A100	2	AMD Zen3 (AMD EPYC 7413), 2 x 24 cores, 512 GB, EDR Infiniband, **2 x NVidia A100**, 2 x 6912 CUDA cores, **2 x 432 Tensor cores**

Kebnekaise (GPU nodes)

Name	#	Description
2 x A100	2	AMD Zen3 (AMD EPYC 7413), 2 x 24 cores, 512 GB, EDR Infiniband, **2 x NVidia A100**, 2 x 6912 CUDA cores, **2 x 432 Tensor cores**
GPU-volta	10	Intel Xeon Gold 6132, 2 x 14 cores, 192 GB, EDR Infiniband, **2 x NVidia V100**, 2 x 5120 CUDA cores, 2 x 16 GB VRAM, **2 x 640 Tensor cores**

Kebnekaise (GPU nodes)

Name	#	Description
2 x A100	2	AMD Zen3 (AMD EPYC 7413), 2 x 24 cores, 512 GB, EDR Infiniband, **2 x NVidia A100**, 2 x 6912 CUDA cores, **2 x 432 Tensor cores**
GPU-volta	10	Intel Xeon Gold 6132, 2 x 14 cores, 192 GB, EDR Infiniband, **2 x NVidia V100**, 2 x 5120 CUDA cores, 2 x 16 GB VRAM, **2 x 640 Tensor cores**
4xGPU !!! Being phased out !!!	4	Intel Xeon E5-2690v4, 2 x 14 cores, 128 GB, FDR Infiniband, **4 x NVidia K80** 8 x 2496 CUDA cores, 8 x 12 GB VRAM

Kebnekaise (GPU nodes)

Name	#	Description
2 x A100	2	AMD Zen3 (AMD EPYC 7413), 2 x 24 cores, 512 GB, EDR Infiniband, **2 x NVidia A100**, 2 x 6912 CUDA cores, **2 x 432 Tensor cores**
GPU-volta	10	Intel Xeon Gold 6132, 2 x 14 cores, 192 GB, EDR Infiniband, **2 x NVidia V100**, 2 x 5120 CUDA cores, 2 x 16 GB VRAM, **2 x 640 Tensor cores**
4xGPU !!! Being phased out !!!	4	Intel Xeon E5-2690v4, 2 x 14 cores, 128 GB, FDR Infiniband, **4 x NVidia K80** 8 x 2496 CUDA cores, 8 x 12 GB VRAM
2xGPU !!! Being phased out !!!	32	Intel Xeon E5-2690v4, 2 x 14 cores, 128 GB, FDR Infiniband, **2 x NVidia K80** 4 x 2496 CUDA cores, 4 x 12 GB VRAM

Kebnekaise (in numbers)

- 553 nodes in 15 racks

Kebnekaise (in numbers)

- 553 nodes in 15 racks
- Intel Broadwell and Skylake, AMD Zen3

Kebnekaise (in numbers)

- 553 nodes in 15 racks
- Intel Broadwell and Skylake, AMD Zen3
- NVidia A100, V100, K80 GPUs

Kebnekaise (in numbers)

- 553 nodes in 15 racks
- Intel Broadwell and Skylake, AMD Zen3
- NVidia A100, V100, K80 GPUs
- More than **135 TB memory**

Kebnekaise (in numbers)

- 553 nodes in 15 racks
- Intel Broadwell and Skylake, AMD Zen3
- NVidia A100, V100, K80 GPUs
- More than **135 TB memory**
- 71 switches (Infiniband, Access and Management networks)

Kebnekaise (in numbers)

- 553 nodes in 15 racks
- Intel Broadwell and Skylake, AMD Zen3
- NVidia A100, V100, K80 GPUs
- More than **135 TB memory**
- 71 switches (Infiniband, Access and Management networks)
- 16504 CPU cores

Kebnekaise (in numbers)

- 553 nodes in 15 racks
- Intel Broadwell and Skylake, AMD Zen3
- NVidia A100, V100, K80 GPUs
- More than **135 TB memory**
- 71 switches (Infiniband, Access and Management networks)
- 16504 CPU cores
- 501760 CUDA cores

Kebnekaise (in numbers)

- 553 nodes in 15 racks
- Intel Broadwell and Skylake, AMD Zen3
- NVidia A100, V100, K80 GPUs
- More than **135 TB memory**
- 71 switches (Infiniband, Access and Management networks)
- 16504 CPU cores
- 501760 CUDA cores
- 12800 Tensor cores

Kebnekaise (HPC2N storage)

- Basically four types of storage are available at HPC2N:

Kebnekaise (HPC2N storage)

- Basically four types of storage are available at HPC2N:
  - Home directory
    - `/home/X/Xyz`, `$HOME`, ~
    - 25 GB, user owned

Kebnekaise (HPC2N storage)

- Basically four types of storage are available at HPC2N:
  - Home directory
    - `/home/X/Xyz`, `$HOME`, ~
    - 25 GB, user owned
  - Project storage
    - `/proj/nobackup/abc`
    - Shared among project members

Kebnekaise (HPC2N storage)

- Basically four types of storage are available at HPC2N:
  - Home directory
    - `/home/X/Xyz`, `$HOME`, ~
    - 25 GB, user owned
  - Project storage
    - `/proj/nobackup/abc`
    - Shared among project members
  - Local scratch space
    - `$SNIC_TMP`
    - SSD (170GB), per job, per node, "volatile"

Kebnekaise (HPC2N storage)

- Basically four types of storage are available at HPC2N:
  - Home directory
    - `/home/X/Xyz`, `$HOME`, ~
    - 25 GB, user owned
  - Project storage
    - `/proj/nobackup/abc`
    - Shared among project members
  - Local scratch space
    - `$SNIC_TMP`
    - SSD (170GB), per job, per node, "volatile"
  - Tape Storage
    - Backup
    - **Long term storage**

Kebnekaise (HPC2N storage)

- Basically four types of storage are available at HPC2N:
  - Home directory
    - `/home/X/Xyz`, `$HOME`, ~
    - 25 GB, user owned
  - Project storage
    - `/proj/nobackup/abc`
    - Shared among project members
  - Local scratch space
    - `$SNIC_TMP`
    - SSD (170GB), per job, per node, "volatile"
  - Tape Storage
    - Backup
    - **Long term storage**
- Also SweStore — disk based (dCache)
  - Research Data Storage Infrastructure, for active research data and operated by NAISS, WLCG

Kebnekaise (projects)

- To use Kebnekaise, you must be a member of a **compute project**

Kebnekaise (projects)

- To use Kebnekaise, you must be a member of a **compute project**
  - A compute project has a certain number of **core hours** allocated for it per month

Kebnekaise (projects)

- To use Kebnekaise, you must be a member of a **compute project**
  - A compute project has a certain number of **core hours** allocated for it per month
  - A regular CPU core cost 1 core hour per hour, other resources (e.g., GPUs) cost more

Kebnekaise (projects)

- To use Kebnekaise, you must be a member of a **compute project**
  - A compute project has a certain number of **core hours** allocated for it per month
  - A regular CPU core cost 1 core hour per hour, other resources (e.g., GPUs) cost more
  - Not a hard limit but projects that go over the allocation get lower priority

Kebnekaise (projects)

- To use Kebnekaise, you must be a member of a **compute project**
  - A compute project has a certain number of **core hours** allocated for it per month
  - A regular CPU core cost 1 core hour per hour, other resources (e.g., GPUs) cost more
  - Not a hard limit but projects that go over the allocation get lower priority
- A compute project contains a certain amount of storage
  - If more storage is required, you must be a member of a **storage project**

Kebnekaise (projects)

- To use Kebnekaise, you must be a member of a **compute project**
  - A compute project has a certain number of **core hours** allocated for it per month
  - A regular CPU core cost 1 core hour per hour, other resources (e.g., GPUs) cost more
  - Not a hard limit but projects that go over the allocation get lower priority
- A compute project contains a certain amount of storage
  - If more storage is required, you must be a member of a **storage project**
- As Kebnekaise is a local cluster, you need to be affiliated with UmU, IRF, SLU, Miun, or LTU to use it

Kebnekaise (projects)

- To use Kebnekaise, you must be a member of a **compute project**
  - A compute project has a certain number of **core hours** allocated for it per month
  - A regular CPU core cost 1 core hour per hour, other resources (e.g., GPUs) cost more
  - Not a hard limit but projects that go over the allocation get lower priority
- A compute project contains a certain amount of storage
  - If more storage is required, you must be a member of a **storage project**
- As Kebnekaise is a local cluster, you need to be affiliated with UmU, IRF, SLU, Miun, or LTU to use it
- Projects are applied for through SUPR (https://supr.naiss.se)

Kebnekaise (projects)

- To use Kebnekaise, you must be a member of a **compute project**
  - A compute project has a certain number of **core hours** allocated for it per month
  - A regular CPU core cost 1 core hour per hour, other resources (e.g., GPUs) cost more
  - Not a hard limit but projects that go over the allocation get lower priority
- A compute project contains a certain amount of storage
  - If more storage is required, you must be a member of a **storage project**
- As Kebnekaise is a local cluster, you need to be affiliated with UmU, IRF, SLU, Miun, or LTU to use it
- Projects are applied for through SUPR (https://supr.naiss.se)
- I will cover more details in the next section, where we go more into detail about HPC2N and Kebnekaise.

What is HPC?

High Performance Computing (definition)

"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."[1]

[1] https://insidehpc.com/hpc-basic-training/what-is-hpc/

High Performance Computing (opening the definition)

- **Aggregating computing power**
  - 533 nodes in 15 racks totalling 16504 cores
  - Compared to 4 cores in a modern laptop

[2] 728 trillion (billion)
[3] 200 billion (milliard)

High Performance Computing (opening the definition)

- **Aggregating computing power**
  - 533 nodes in 15 racks totalling 16504 cores
  - Compared to 4 cores in a modern laptop
- **Higher performance**
  - More than 728 000 000 000 000 arithmetical operations per second[2]
  - Compared to 200 000 000 000 Flops in a modern laptop[3]

[2] 728 trillion (billion)
[3] 200 billion (milliard)

High Performance Computing (opening the definition)

- **Aggregating computing power**
  - 533 nodes in 15 racks totalling 16504 cores
  - Compared to 4 cores in a modern laptop
- **Higher performance**
  - More than 728 000 000 000 000 arithmetical operations per second[2]
  - Compared to 200 000 000 000 Flops in a modern laptop[3]
- **Solve large problems**
  - When does a problem become large enough for HPC?
  - Are there other reasons for using HPC resources? (Memory, software, support, etc.)

[2] 728 trillion (billion)
[3] 200 billion (milliard)

High Performance Computing (large problems)

- A problem can be large for two main reasons:
  1. Execution time: The time required to form a solution to the problem is very long
  2. Memory / storage use: The solution of the problem requires a lot of memory and/or storage

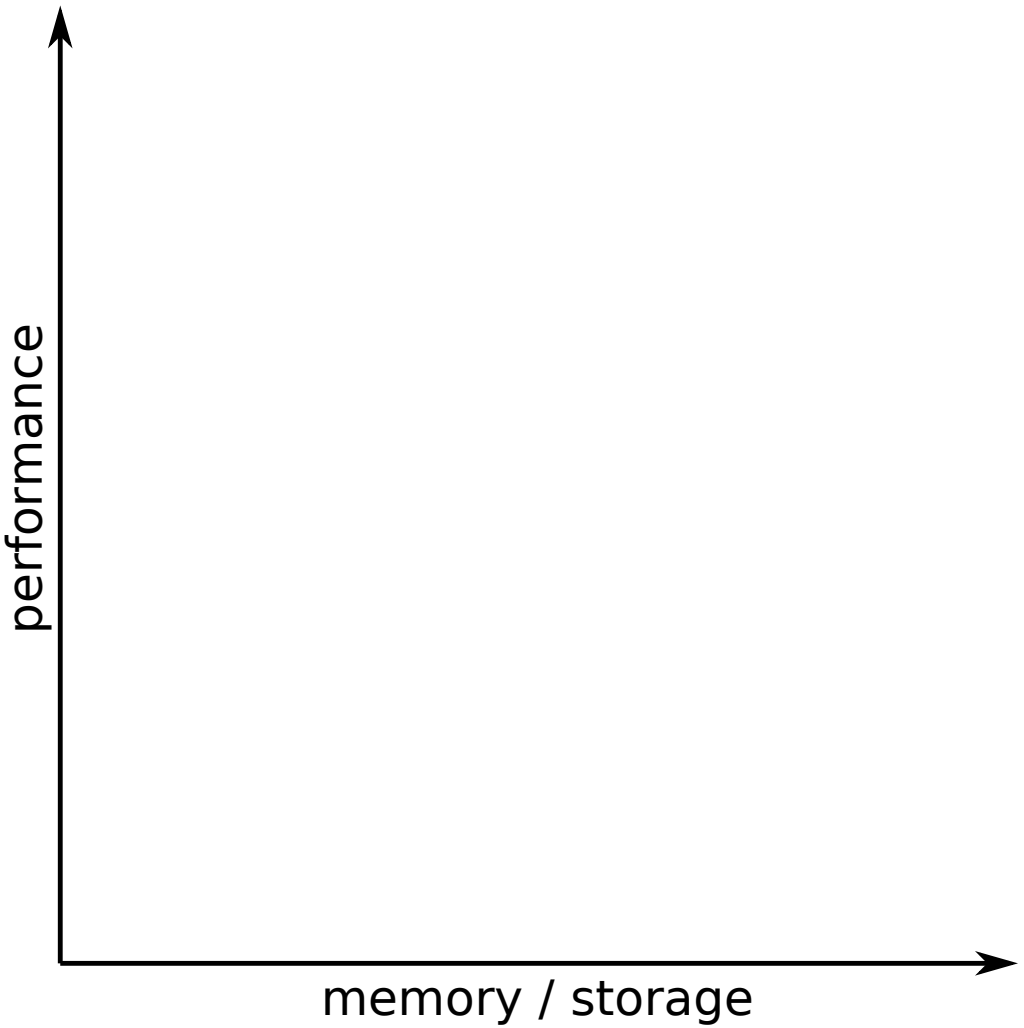
High Performance Computing (large problems)

- A problem can be large for two main reasons:
  1. Execution time: The time required to form a solution to the problem is very long
  2. Memory / storage use: The solution of the problem requires a lot of memory and/or storage
- The former can be remedied by **increasing the performance**
  - More cores, more nodes, GPUs, ...

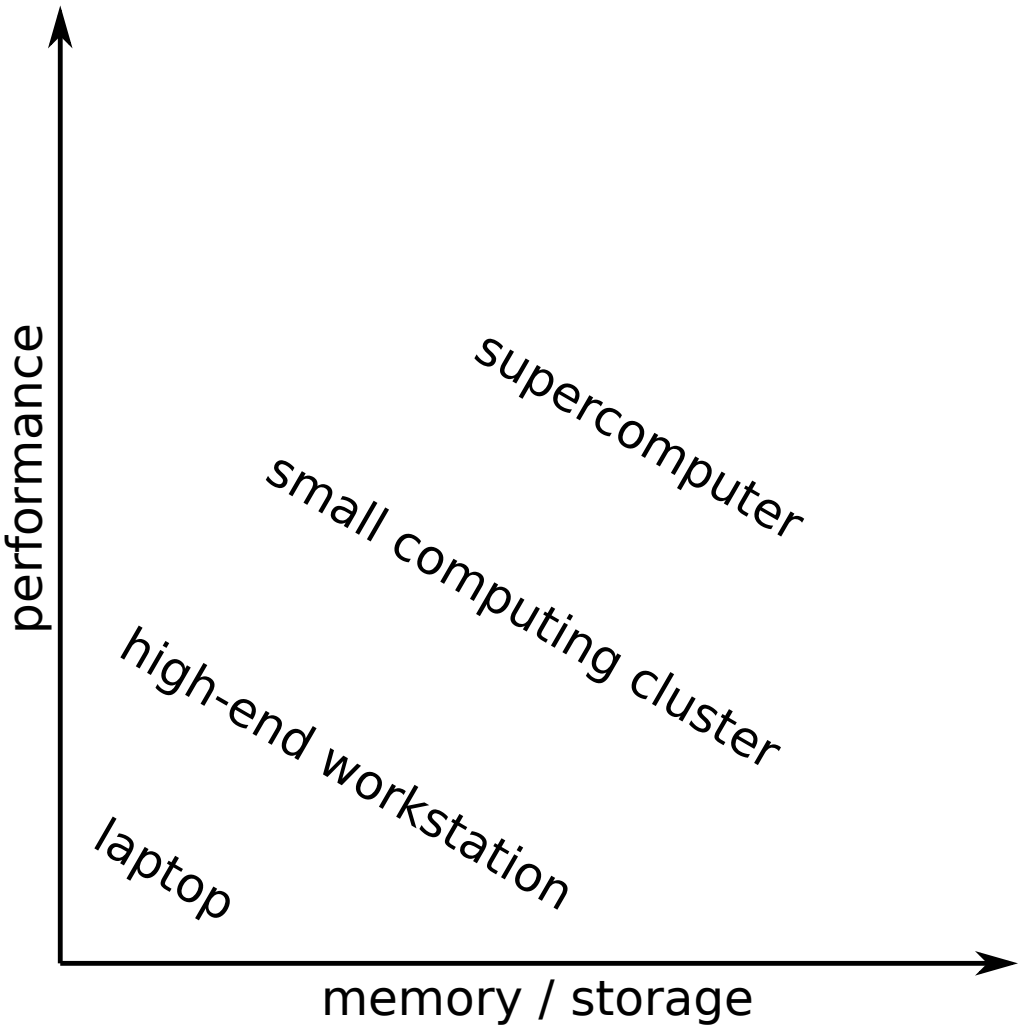
High Performance Computing (large problems)

- A problem can be large for two main reasons:
  1. Execution time: The time required to form a solution to the problem is very long
  2. Memory / storage use: The solution of the problem requires a lot of memory and/or storage
- The former can be remedied by **increasing the performance**
  - More cores, more nodes, GPUs, ...
- The latter by **adding more memory / storage**
  - More memory per node (including large memory nodes), more nodes, ...
  - Kebnekaise: 128GB - 192GB, 512GB, 3TB
  - Large storage solutions, ...

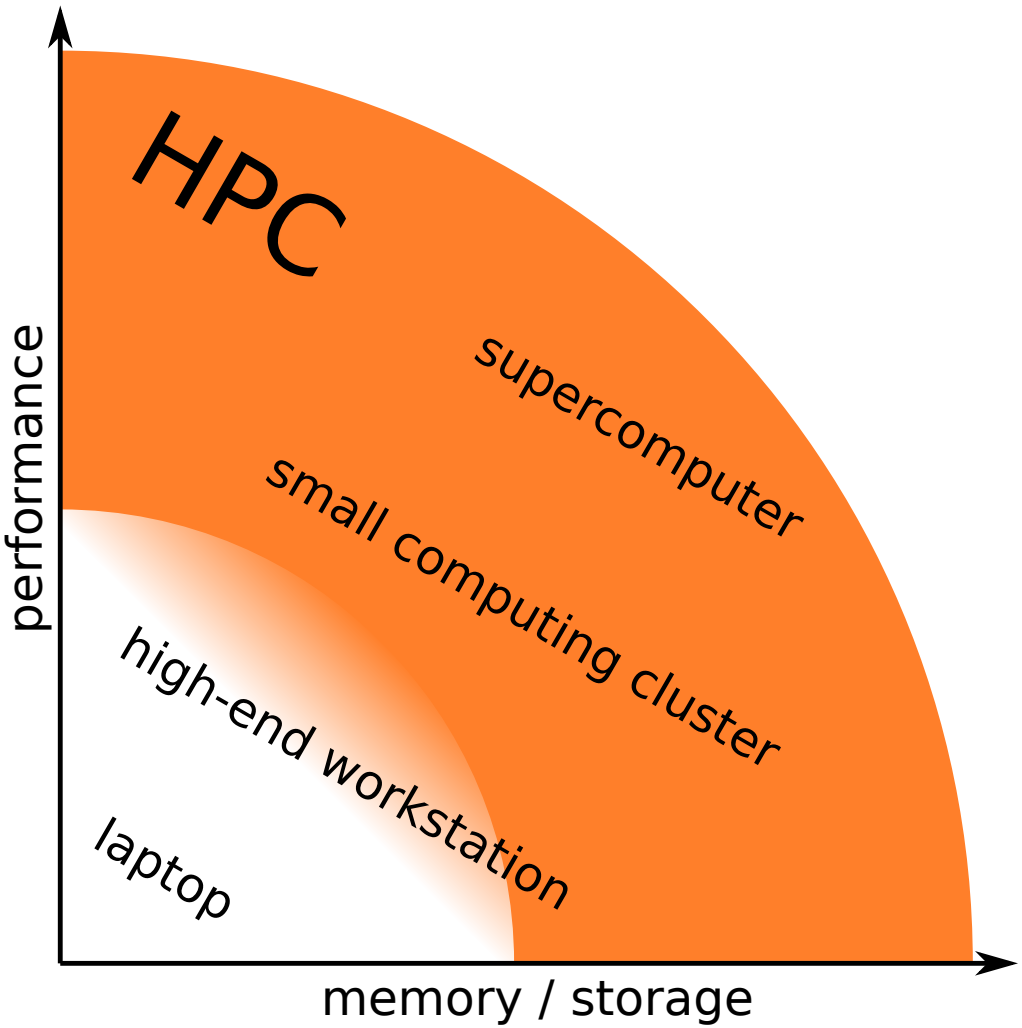
High Performance Computing (what counts as HPC)



High Performance Computing (what counts as HPC)



High Performance Computing (what counts as HPC)



High Performance Computing (other reasons)

- Specialized (expensive) hardware

High Performance Computing (other reasons)

- Specialized (expensive) hardware
  - GPUs, **Nvidia Tesla V100/A100 GPUs** are optimized for AI

High Performance Computing (other reasons)

- Specialized (expensive) hardware
  - GPUs, **Nvidia Tesla V100/A100 GPUs** are optimized for AI
  - High-end CPUs (AVX-512 etc) and ECC memory

High Performance Computing (other reasons)

- Specialized (expensive) hardware
  - GPUs, **Nvidia Tesla V100/A100 GPUs** are optimized for AI
  - High-end CPUs (AVX-512 etc) and ECC memory
- Software
  - HPC2N holds **licenses** for several softwares
  - Software is **pre-configured and ready-to-use**

High Performance Computing (other reasons)

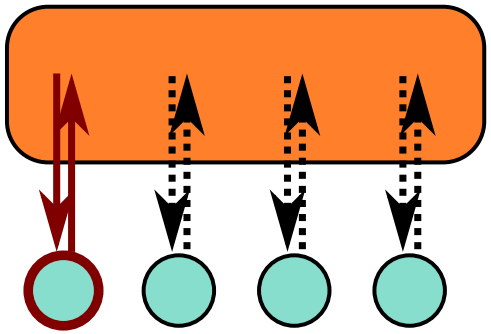
- Specialized (expensive) hardware
  - GPUs, **Nvidia Tesla V100/A100 GPUs** are optimized for AI
  - High-end CPUs (AVX-512 etc) and ECC memory
- Software
  - HPC2N holds **licenses** for several softwares
  - Software is **pre-configured and ready-to-use**
- **Support and documentation**

High Performance Computing (memory models)

- Two memory models are relevant for HPC:

High Performance Computing (memory models)

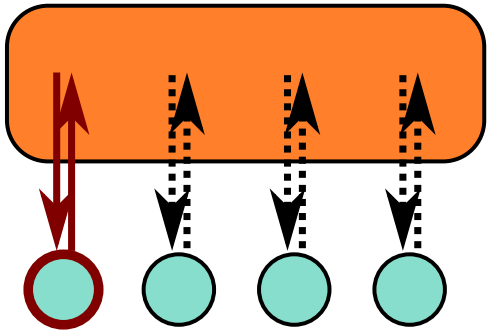
- Two memory models are relevant for HPC:
  - Shared memory: Single memory space for all data.



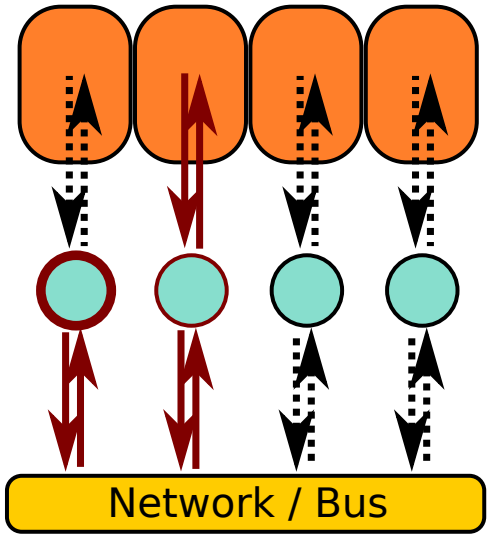
- **Everyone can access the same data**
- Straightforward to use

High Performance Computing (memory models)

- Two memory models are relevant for HPC:
  - Shared memory: Single memory space for all data.

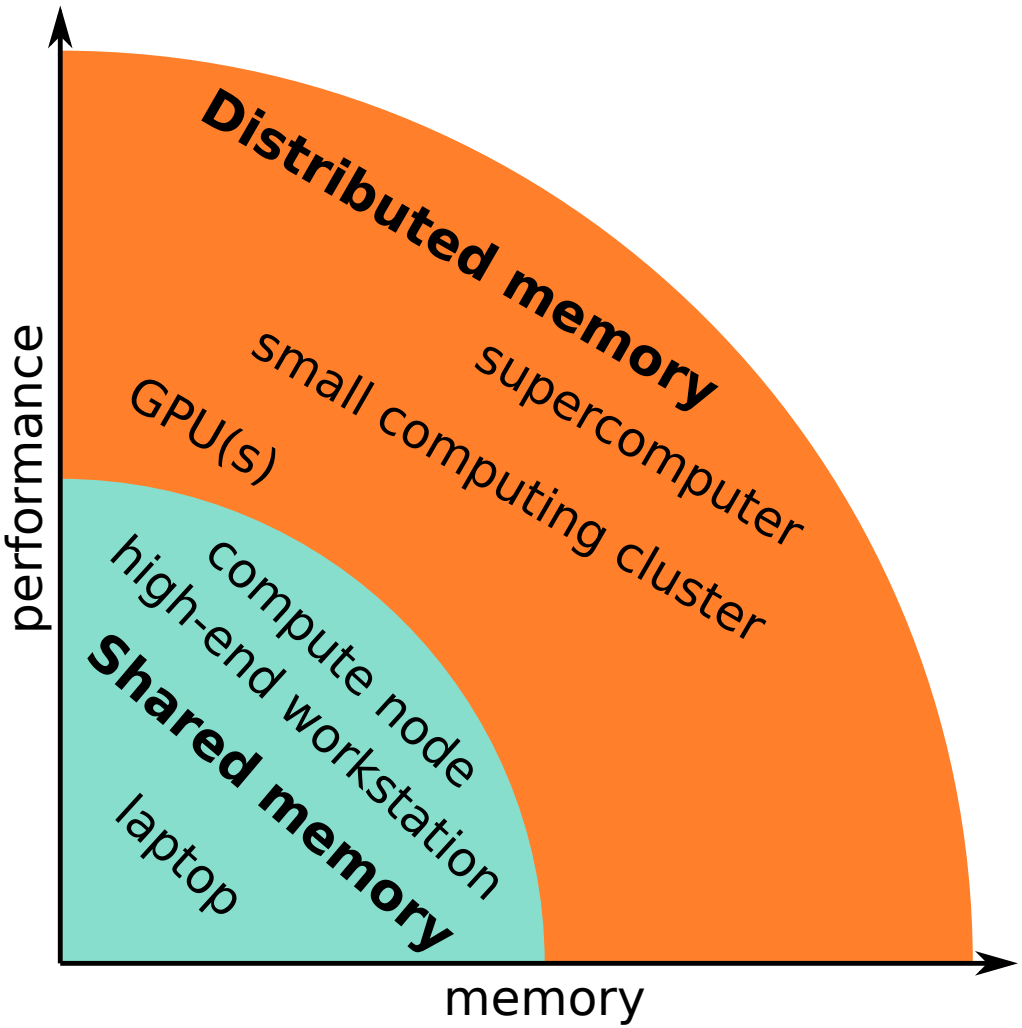


  - **Everyone can access the same data**
  - Straightforward to use
  - Distributed memory: Multiple **distinct** memory spaces.



  - Everyone has direct access **only to the local data**
  - Requires **communication**

High Performance Computing (memory models)



High Performance Computing (programming models)

- The programming model changes when we aim for extra performance and/or memory:

High Performance Computing (programming models)

- The programming model changes when we aim for extra performance and/or memory:
  1. Single-core: Matlab, Python, C, Fortran, ...
     - Single stream of operations

High Performance Computing (programming models)

- The programming model changes when we aim for extra performance and/or memory:
  1. Single-core: Matlab, Python, C, Fortran, ...
     - Single stream of operations
  2. Multi-core: Vectorized Matlab, pthreads, **OpenMP**
     - **Multiple streams** of operations

High Performance Computing (programming models)

- The programming model changes when we aim for extra performance and/or memory:
  1. Single-core: Matlab, Python, C, Fortran, ...
     - Single stream of operations
  2. Multi-core: Vectorized Matlab, pthreads, **OpenMP**
     - **Multiple streams** of operations
     - **Work distribution**, **coordination** (synchronization, etc), ...

High Performance Computing (programming models)

- The programming model changes when we aim for extra performance and/or memory:
  1. Single-core: Matlab, Python, C, Fortran, ...
     - Single stream of operations
  2. Multi-core: Vectorized Matlab, pthreads, **OpenMP**
     - **Multiple streams** of operations
     - **Work distribution**, **coordination** (synchronization, etc), ...
  3. Distributed memory: **MPI**, ...
     - Multiple streams of operations
     - Work distribution, coordination (synchronization, etc), ...

High Performance Computing (programming models)

- The programming model changes when we aim for extra performance and/or memory:
  1. Single-core: Matlab, Python, C, Fortran, ...
     - Single stream of operations
  2. Multi-core: Vectorized Matlab, pthreads, **OpenMP**
     - **Multiple streams** of operations
     - **Work distribution**, **coordination** (synchronization, etc), ...
  3. Distributed memory: **MPI**, ...
     - Multiple streams of operations
     - Work distribution, coordination (synchronization, etc), ...
     - **Data distribution and communication**

High Performance Computing (programming models)

- The programming model changes when we aim for extra performance and/or memory:
  1. Single-core: Matlab, Python, C, Fortran, ...
     - Single stream of operations
  2. Multi-core: Vectorized Matlab, pthreads, **OpenMP**
     - **Multiple streams** of operations
     - **Work distribution**, **coordination** (synchronization, etc), ...
  3. Distributed memory: **MPI**, ...
     - Multiple streams of operations
     - Work distribution, coordination (synchronization, etc), ...
     - **Data distribution and communication**
- GPUs: **CUDA**, OpenCL, OpenACC, OpenMP, ...

High Performance Computing (programming models)

- The programming model changes when we aim for extra performance and/or memory:
  1. Single-core: Matlab, Python, C, Fortran, ...
     - Single stream of operations
  2. Multi-core: Vectorized Matlab, pthreads, **OpenMP**
     - **Multiple streams** of operations
     - **Work distribution**, **coordination** (synchronization, etc), ...
  3. Distributed memory: **MPI**, ...
     - Multiple streams of operations
     - Work distribution, coordination (synchronization, etc), ...
     - **Data distribution and communication**
- GPUs: **CUDA**, OpenCL, OpenACC, OpenMP, ...
  - **Many lightweight** streams of operations

High Performance Computing (programming models)

- The programming model changes when we aim for extra performance and/or memory:
  1. Single-core: Matlab, Python, C, Fortran, ...
     - Single stream of operations
  2. Multi-core: Vectorized Matlab, pthreads, **OpenMP**
     - **Multiple streams** of operations
     - **Work distribution**, **coordination** (synchronization, etc), ...
  3. Distributed memory: **MPI**, ...
     - Multiple streams of operations
     - Work distribution, coordination (synchronization, etc), ...
     - **Data distribution and communication**
- GPUs: **CUDA**, OpenCL, OpenACC, OpenMP, ...
  - **Many lightweight** streams of operations
  - Work distribution, coordination (synchronization, etc), ...

High Performance Computing (programming models)

- The programming model changes when we aim for extra performance and/or memory:
  1. Single-core: Matlab, Python, C, Fortran, ...
     - Single stream of operations
  2. Multi-core: Vectorized Matlab, pthreads, **OpenMP**
     - **Multiple streams** of operations
     - **Work distribution**, **coordination** (synchronization, etc), ...
  3. Distributed memory: **MPI**, ...
     - Multiple streams of operations
     - Work distribution, coordination (synchronization, etc), ...
     - **Data distribution and communication**
- GPUs: **CUDA**, OpenCL, OpenACC, OpenMP, ...
  - **Many lightweight** streams of operations
  - Work distribution, coordination (synchronization, etc), ...
  - **Data distribution across memory spaces and movement**

High Performance Computing (software)

- Complexity grows when we aim for extra performance and/or memory/storage:

High Performance Computing (software)

- Complexity grows when we aim for extra performance and/or memory/storage:
  1. Single-core: LAPACK, ...
     - Load correct toolchain etc

High Performance Computing (software)

- Complexity grows when we aim for extra performance and/or memory/storage:
  1. Single-core: LAPACK, ...
     - Load correct toolchain etc
  2. Multi-core: LAPACK + parallel BLAS, ...
     - Load correct toolchain etc

High Performance Computing (software)

- Complexity grows when we aim for extra performance and/or memory/storage:
  1. Single-core: LAPACK, ...
     - Load correct toolchain etc
  2. Multi-core: LAPACK + parallel BLAS, ...
     - Load correct toolchain etc
     - **Allocate** correct number of cores, **configure** software to use correct number of cores, ...

High Performance Computing (software)

- Complexity grows when we aim for extra performance and/or memory/storage:
  1. Single-core: LAPACK, ...
     - Load correct toolchain etc
  2. Multi-core: LAPACK + parallel BLAS, ...
     - Load correct toolchain etc
     - **Allocate** correct number of cores, **configure** software to use correct number of cores, ...
  3. Distributed memory: ScaLAPACK, ...
     - Load correct toolchain etc

High Performance Computing (software)

- Complexity grows when we aim for extra performance and/or memory/storage:
  1. Single-core: LAPACK, ...
     - Load correct toolchain etc
  2. Multi-core: LAPACK + parallel BLAS, ...
     - Load correct toolchain etc
     - **Allocate** correct number of cores, **configure** software to use correct number of cores, ...
  3. Distributed memory: ScaLAPACK, ...
     - Load correct toolchain etc
     - Allocate correct number of **nodes and cores**, configure software to use correct number of **nodes and cores**, ...

High Performance Computing (software)

- Complexity grows when we aim for extra performance and/or memory/storage:
  1. Single-core: LAPACK, ...
     - Load correct toolchain etc
  2. Multi-core: LAPACK + parallel BLAS, ...
     - Load correct toolchain etc
     - **Allocate** correct number of cores, **configure** software to use correct number of cores, ...
  3. Distributed memory: ScaLAPACK, ...
     - Load correct toolchain etc
     - Allocate correct number of **nodes and cores**, configure software to use correct number of **nodes and cores**, ...
     - Data distribution, storage, ...

High Performance Computing (software)

- Complexity grows when we aim for extra performance and/or memory/storage:
  1. Single-core: LAPACK, ...
     - Load correct toolchain etc
  2. Multi-core: LAPACK + parallel BLAS, ...
     - Load correct toolchain etc
     - **Allocate** correct number of cores, **configure** software to use correct number of cores, ...
  3. Distributed memory: ScaLAPACK, ...
     - Load correct toolchain etc
     - Allocate correct number of **nodes and cores**, configure software to use correct number of **nodes and cores**, ...
     - Data distribution, storage, ...
- GPUs: MAGMA, TensorFlow, ...
  - Load correct toolchain etc
  - Allocate correct number of **cores and GPUs**, configure software to use correct number of **cores and GPUs**, ...

End (questions?)

Questions?