# Basic Tools for NLP

Rishu Kumar, Iuliia Zaitova, Chris Hyek

Saarland University

20.07.2023

## Outline: cluster computing

- Computer clusters at coli
- File system
- CPU / GPU / CUDA
- Requirements of your job
- Running jobs at the coli cluster

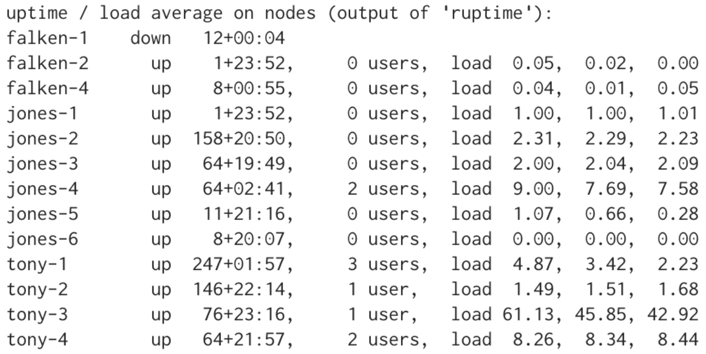
## Coli account

Do you have it?

If not, *please* follow the link:
https://wiki.coli.uni-saarland.de/public/AccountApplication

## Coli cluster

- Login to the cluster:
  `ssh your_coli_username@login.coli.uni-saarland.de`
- `See the output of the "rwho" and "ruptime" commands`



```
uptime / load average on nodes (output of 'ruptime'):
falken-1    down    12+00:04
falken-2      up     1+23:52,     0 users,  load  0.05,  0.02,  0.00
falken-4      up     8+00:55,     0 users,  load  0.04,  0.01,  0.05
jones-1       up     1+23:52,     0 users,  load  1.00,  1.00,  1.01
jones-2       up   158+20:50,     0 users,  load  2.31,  2.29,  2.23
jones-3       up    64+19:49,     0 users,  load  2.00,  2.04,  2.09
jones-4       up    64+02:41,     2 users,  load  9.00,  7.69,  7.58
jones-5       up    11+21:16,     0 users,  load  1.07,  0.66,  0.28
jones-6       up     8+20:07,     0 users,  load  0.00,  0.00,  0.00
tony-1        up   247+01:57,     3 users,  load  4.87,  3.42,  2.23
tony-2        up   146+22:14,     1 user,   load  1.49,  1.51,  1.68
tony-3        up    76+23:16,     1 user,   load 61.13, 45.85, 42.92
tony-4        up    64+21:57,     2 users,  load  8.26,  8.34,  8.44
```

## What do these mean?

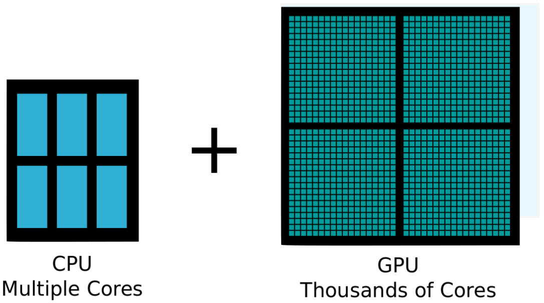
- **falken-1 to falken-4**: Large memory compute servers Dell R815 featuring 4 AMD Opteron 6380 16-core processors (64 core in total) at 1TB RAM, each.
- **fjones-1 to jones-6**: GPU compute servers Dell T630 featuring 4 Intel Xeon E5-2687W v3 10-core processors (40 cores in total) at 256GB RAM and 4 MSI nvidia Titan-X (2015) GPU cards (12GB), each.
- **ftony-1 (aka volta-1) and tony-2 (aka volta-2)**: GPU compute servers Supermicro featuring 2 Intel Xeon 6128 6-core processors (12 cores in total, 24 cores with hyperthreading) at 768GB RAM and 8 nvidia Tesla V100 (2018) GPU cards (32GB), each.
- **ftony-3 (aka ampere-1)**: GPU compute servers Supermicro featuring 2 AMD EPYC 7713 64-core processors (128 cores in total, 256 cores with hyperthreading) at 512GB RAM and 8 nvidia Tesla A100 (2022) GPU cards (80GB).
- **ftony-4 (aka ampere-2)**: GPU compute servers Supermicro featuring 2 AMD EPYC 7662 64-core processors (128 cores in total, 256 cores with hyperthreading) at 512GB RAM and 8 nvidia Tesla A100 (2022) GPU cards (40GB), each.

## CPU (Central Processing Unit)

- General-purpose processor designed to handle sequential tasks.
- Fewer but more powerful cores with complex instruction pipelines.

## GPU (Graphics Processing Unit)

- Specialized for parallel processing of graphics and computational tasks.
- Thousands of cores designed to handle multiple threads simultaneously.



## CPU (Central Processing Unit)

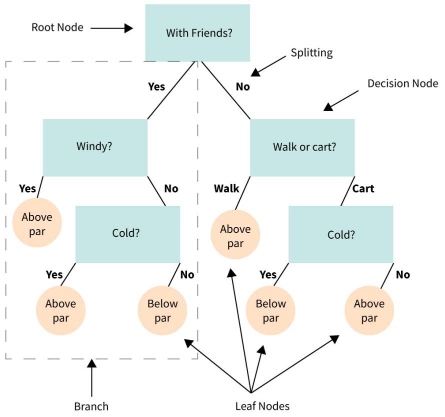
- General-purpose processor designed to handle sequential tasks.
- Fewer but more powerful cores with complex instruction pipelines.
- **Used in general computing tasks, operating systems, and software applications.**

## GPU (Graphics Processing Unit)

- Specialized for parallel processing of graphics and computational tasks.
- Thousands of cores designed to handle multiple threads simultaneously.
- **Optimized for data parallelism.**
- Often used in ML and AI (and gaming) — efficiently handles large matrices, vector operations.

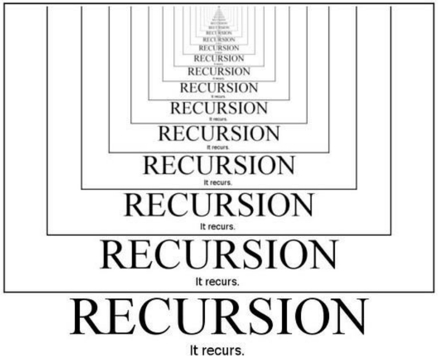
## 3D graphics rendering?

## **Programs that contain a lot of conditional statements?** (if, switch, while and such)



## Applying filter to a photo?

## Recursive algorithms?



UNIVERSITÄT DES SAARLANDES

When doing machine learning

1. Run serial workload on CPU
2. Offload parallel computation to GPUs

## Instructions by Noon Pokaratsiri Goldstein

iuliiazaitova.github.io/basic-tools-nlp-2023

Download the instructions and follow them

pip install torch, numpy

run numpy_matrices.py on the server

run cpu_task.py on the server

experiment with `nvidia-smi`

```
export CUDA_VISIBLE_DEVICES=0,1
```