Exercise „Data Science Infrastructures“

Exercise 01 (DSI E01 ST24) – Related Lecture DSI L02

18.04.2024, Philipp Wieder, GWDG

Content

I. Compute Architectures
II. Get familiar with HPC access
III. Cost of Cloud Computing

- Solutions can be both submitted in German and English
- Please submit the solutions to the “Exercise Solutions” folder in Stud.IP

Assignment I
Compute Architectures

a) Research on the characteristics of CPUs, GPUs, and TPUs. What purpose have they been designed for? Which data science problems benefit most?
b) Choose a current model each (one CPU, one GPU, and one TPU) and compare
   - Performance
   - Power consumption
   - Cost

Provide references regarding the sources you used. Also in case you use a language model-based chat system, refer it.

Assignment II
HPC Access

a) We are going to use the HPC systems in Göttingen to get experience with a few data science-related infrastructure components.
   - You need a GWDG account for that (course accounts will be generated!) and it has to be activated for HPC usage.
   - Please make yourself familiar how to activate it as this includes generating SSH keys
   - URL: https://info.gwdg.de/dokuwiki/doku.php?id=en:services:application_services:high_performance_computing:connect_with_ssh

Assignment III
Cost of Cloud Computing

For a data science project you need the following IaaS ressources

- Virtual machine with at least 4 vCPUs & 8 GB RAM (no special instance needed)
- Object storage for up to 100TB of data (standard)
- Hosted in Europe

a) How much will this setting cost you per year? Choose one of the providers presented in the lecture.

b) You put your project in operation and monitor a mean value of 50 TB/month data transfer (outgoing). How much will that cost you per month?