Notes

(i) Make a group of 2-4:.

(ii) Exercises & slides are enough for good marks.

⇒ jonathan.decker @ cscs-goettingen.de

Note: Use a virtual machine for exercise.

Note: This exercise is super useful for your CV as well.

(iii) Most of the things here are learning by doing, no memorization.

Lecture-1:.

(i) D/P/S ⇒ Distributed Parallel scientific computing.

Note: Learning outcome is the all you need to learn & confronted.

(ii) Main problem of Distributed system ⇒ concurrency introduces new types of programming mistakes.

(iii) HPC is a kind of formula 1 for the super computing.

Note: You cannot do the testing if you don't know the results.

(iii) Scientific method example: Falling of an apple. Why?

(iv) Raw data = low value (fine grained)

Group Work:. (Example).

(i) Questions to be solved: getting insights from scientific methods.

(ii) Hypothesis ⇒ Rainfall is highly correlated to thunder.

(iii) How to test ⇒ By. getting correlation matrix.

(iv) What data ⇒ Weather

Another example: Did Quarantine made less COVID cases?
Possible solution ⇒ By the help of simulations.

V = Volume

Never changing data are not big data.

V = Variety.
V = Veracity (Trustworthiness) → 5 Vs → challenges of HPDA

Analytics and theories increases the value of the data.

⇒ A simple tool like Hadoop has 50.5 B worth.

Note: Thesis option is also a good option for these course.

→ Big data analytics is a sub-aspect of HPDA.

→ HPDA is required for big data analytics.

solve the exercises of HPDA

Note: Linux crash course is available on the website of the professor. So you can have a look at it.

→ DFS is the python system.
→ Yarn is is a cluster manager.

→ real time → user time + system time.

→ When using multiple sources, real time is even less than the system and user time.

Exercises-I:

→ formative assessment ⇒ development assessment. → imp for exam as well.

Contents:- (i) HPCA Use cases understanding
(ii) Bash Exercise
Bash Research.
(iii) Set up software environment in a Virtual Machine
(iv) Reading CSV Data (Python).

Tip: Since this course is very flexible and you should just attempt to solve questions. It can be super helpful for just trying & using chatGPT as much as you can.

⇒ -p ⇒ Creates a parent directories if needed. Don't throw the error if the directory already exists.

⇒ && ⇒ After creating the directory, other functions are implemented. However, they are only implemented if the directories are created.

⇒ CTRL + C ⇒ to terminate the running command early.

⇒ command + right arrow → to point the cursor of the terminal to the left / right. (at the beginning / end)

⇒ ; ! <> → cause error in terminal when included outside of quotes in a command.