**Exercise 1:**

**Task 1:**

1. Research one recent HPDA use case from industry in the literature, e.g., using Google or Google Scholar. Describe the use case briefly, put an emphasis on performance relevant characteristics, e.g., how much data do they process, how long does the workflow run? Address how the Big Data Challenges (5V) apply.

Make sure you cite the used literature properly! Your notes should cover about 1/4-page.

**Answer**:

A notable use case in High Performance Data Analytics (HPDA) is genome sequencing, which has become increasingly essential for both medical research and healthcare innovation. This process leverages HPDA to analyse massive datasets produced by sequencers-each generating up to 6 terabytes of data daily. Using HPDA allows researchers to significantly reduce the time required for analysis, with workflows that used to take years now producing results within a day. This increase in efficiency is vital for timely insights, such as identifying genetic markers or developing targeted therapies for diseases.

In this use case, several of the 5V Big Data challenges are present.

* Volume: The vast data produced daily (in petabytes) must be processed, requiring scalable storage and efficient data access systems.
* Velocity: Real-time data throughput is crucial, especially when processing high bandwidth data, which demands optimized storage and retrieval speeds.
* Variety: The data spans structured genomic sequences and unstructured metadata, making unified management systems essential.
* Veracity: Reliable, accurate data handling is critical for clinical relevance, demanding low error rates and data integrity.
* Value: Timely genomic delivers immense values, transforming healthcare through personalized treatments and rapid diagnostics.

Rerence: Wang H, Wang X, Ren H, Wang X, Lu Z.2020.3-Hydroxypyridine Dehydrogenase HpdA Is Encoded by a Novel Four-Component Gene Cluster and Catalyzes the First Step of 3-Hydroxypyridine Catabolism in Ensifer adhaerens HP1. Appl Environ Microbiol86:e01313-20.https://doi.org/10.1128/AEM.01313-20

2. Research a scientific HPDA use case following the same instructions as above.

**Answer:**

One prominent scientific use case for High Performance Data Analytics (HPDA) is in earth system modeling, where HPDA techniques are used to simulate complex environmental and climate processes. This application involves running extensive simulations that demand high-resolution data from multiple sources, such as satellite imagery and environmental sensors. HPDA tools enable researchers to analyze petabytes of data to model weather patterns, predict natural disasters, and study climate change with improved accuracy. By integrating deep learning models as “surrogates,” the time required for these simulations has been reduced significantly, allowing researchers to explore various scenarios rapidly and efficiently.

This application directly addresses the Big Data Challenges (5V):

* Volume: The system processes petabytes of data generated from various global sources, which must be integrated and analyzed continuously.
* Velocity: Real-time data processing is crucial to predict rapid changes in weather and environmental conditions.
* Variety: Data sources include structured meteorological data, unstructured satellite imagery, and heterogeneous environmental datasets.
* Veracity: Data accuracy is essential to ensure reliable climate predictions, as errors can lead to significant predictive inaccuracies.
* Value: Timely and accurate climate models provide substantial value in environmental protection, disaster preparedness, and policymaking.

Reference: Lewandowski, Natalie and Koller, Bastian. ‘Transforming Medical Sciences with High-performance Computing, High-performance Data Analytics and AI’. 1 Jan. 2023 : 1505 – 1507.

**Task 2:**

1. $ echo "Hello World" 🡪 Prints out hello world. Everything within semicolon is treated as one string.

2. $ echo Hello, World 🡪 Also prints hello world. However, echo treats “Hello,” and “World” as two different strings. By default, the string for echo is separated by space.

3. $ echo Hello, world; Foo bar 🡪 The bash prints “Hello,” and “world”. However, it could not identify “foo”.

4. $ echo Hello, world! 🡪 Prints Hello, world!

5. $ echo "line one";echo "line two" 🡪 Prints following. “;” separates the line.

line one

line two

6. $ echo "Hello, world > readme" 🡪

7. $ echo "Hello, world" > readme

8. $ cat readme

9. $ example="Hello, World"

10. $ echo $example

11. $ echo ’$example’

12. $ echo "$example"

13. $ echo "Please enter your name."; read example

14. $ echo "Hello $example"

15. $ three=1+1+1;echo $three

16. $ bc

HPDA – Exercise Week 1 2/717. $ echo 1+1+1 | bc

18. $ let three=1+1+1;echo $three

19. $ echo date

20. $ cal

21. $ which cal

22. $ /bin/cal

23. $ $(which cal)

24. $ ‘which cal‘

25. $ echo "The date is $(date)"

26. $ seq 0 9

27. $ seq 0 9 | wc -l

28. $ seq 0 9 > sequence

29. $ wc -l < sequence

30. $ for I in $(seq 1 9) ; do echo $I ; done

31. $ (echo -n 0 ; for I in $(seq 1 9) ; do echo -n +$I ; done ; echo) | bc