**Q1.**

For all OSS jobs that require big data processing, including ETL and machine learning, Dataproc is a managed service. The most well-liked open source software is supported right out of the box. To increase productivity and enable scale, you may use Dataproc to relocate your on-premise OSS clusters to the cloud. You can also use it in conjunction with Cloud AI Notebook or BigQuery to create a complete data science environment. With Dataproc, an IT-managed cluster can be auto scaled in about 90 seconds. It oversees job orchestration, monitoring, and cluster building. To begin, a cluster must be created via the web UI, the Clod SDK, the REST APIs, or SSH access. Once the cluster is set up, jobs can be submitted using any open-source framework. Even when jobs are running, the cluster can be scaled up or down at any time and only pay for the resources utilised up to that precise moment.

**Q2.**

It is true that the scientific domains, such as computational biology, meteorology, bioinformatics, and astronomy, are highly data-driven approach for various reasons:

Data availability: Because of the huge volumes of data that have been generated in various domains as a result of technological advancements, it is crucial to apply data-driven methodologies to analyse and glean insights from this wealth of knowledge.

Computing strength: Increased computer capacity makes it possible to handle and analyse massive datasets, allowing researchers to take on challenging simulations and issues that were previously impracticable.

Modelling and simulation: Because scientific modelling and simulation frequently produce enormous datasets, effective understanding and validation of these models depend on data-driven methodologies.

Discovering: Discovering new information, anticipating future developments, and spotting hidden patterns and connections in complicated systems have all been made possible by data-driven methodologies.

Interdisciplinary Research: In order to conduct meaningful analysis, many scientific disciplines increasingly demand the integration of data from diverse sources and domains.

Data Collaboration: Open access data repositories and collaboration tools make it simpler for researchers to access and share data, promoting collaboration and data-driven research.

Financial Priorities: Because data-driven research has the potential to spur innovation and tackle pressing scientific problems, funding organizations and academic institutions are placing more and more emphasis on it.

To conclude, technical developments, processing power, and the realization that data holds the key to furthering our understanding, generating predictions, and solving challenging problems in these areas are what are driving the increased dependence on data-driven techniques in these scientific subjects.

**Q3.**

For their genetic variant analysis, the researchers in the study utilised two cloud service providers:

Google Cloud Platform (GCP): They used Google Dataproc on the Google Cloud Platform (GCP) to construct Spark clusters and carry out genomic variant analysis.

Amazon Web Services (AWS): For the same goals, they employed Amazon Elastic MapReduce (EMR) on AWS as the second cloud provider.

Regarding the outcomes on the platforms from both providers:

Yes, their outcomes on the platforms of both providers were largely the same. On GCP and AWS, the researchers compared the effectiveness and cost-effectiveness of their method for analysing genetic variants. They discovered that when they set up both cloud services in the same way, they got essentially equal execution times for all operations involving genetic variation analysis. This indicates that their approach and procedure were successful on both GCP and AWS, and the outcomes were comparable in terms of computing performance between the two services. They did, however, point out that while there might be some changes in hardware features, cluster topologies, and network connections between the two providers, these variances did not have a substantial impact on the overall outcomes.

**Q4.**

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**References and Citations:**

 Chaudhry, A. F., Aniol, H., & Shegos, C. J. (2020). https://www.consultant360.com/article/consultant360/congenital-hypothyroidism-due-thyroid-agenesis. Consultant. https://doi.org/10.25270/con.2020.05.00011

In-Text Citation: (Chaudhry et al., 2020)

https://ieeexplore-ieee-org.libproxy.library.unt.edu/stamp/stamp.jsp?arnumber=8987848&casa\_token=8Dz7wOyCfLQAAAAA:UfckNwscsqsli0EMtH-sYTgjlYvXGdZWv2gY2shIMCKUW9Fs6edeSZMYVg\_\_mgUS-q9bV0A&tag=1