Discuss the dashboard that monitors your cluster in GCP and include a screenshot of the dashboard.

A screenshot of a graph

Description automatically generated

The dashboard monitors the cluster in GCP (Google Cloud Platform). It shows the centralized view of the cluster’s performance.

The cluster dashboard includes the following metrics and graphs some of the KPI include:

1. CPU Utilization: This graph shows the CPU usage percentage at the run time, averaging below 1%.

2. Network Bytes: This tracks the amount of network traffic, both incoming and outgoing, measured in bytes per second. Share spike happens when adding files and moving files.

3. Network Packets: Similar to the Network Bytes graph but focusing on the number of packets sent and received per second.

4. Disk Bytes: This graph shows the read and write operations to the disk in bytes per second. There is a notable spike in write operations without a corresponding increase in read operations.

5. Disk Operations: This measures the read and write operations to the disk in operations per second. The pattern is similar to the Disk Bytes graph, with write operations spiking at specific points, indicating bursts of write activity to the disk.

* Steps to create the new subfolder "data" in your directory (with screenshots) within the Hadoop framework.
  1. Go Navigator > DataProc > Start Cluster and VM instance
  2. Go to the master node through “SSH” > Open in Browse
  3. Type Clear > Enter
  4. Used the following command to create my subdirectory under user
  5. Create a directory in hdfs
     + **hdfs dfs -mkdir /user/biniamabebe**
     + **hdfs dfs -mkdir /user/biniamabebe/data**

A screenshot of a computer program

Description automatically generated

* Steps I took to create two subfolders (userdata & weblog) within the newly created "data" folder and include screenshots.
  1. Check the successful creation of data dir
     + hdfs dfs -ls /user/biniamabebe/data
  2. Create the userdata & weblog subfolders
     + hdfs dfs - **mkdir** /user/biniamabebe/data/userdata
     + hdfs dfs - **mkdir** /user/biniamabebe/data/weblog

A black screen with white text

Description automatically generated

* I took steps to copy the two .csv files (userdata & weblog) from the GCP bucket into HDFS and included screenshots.
  1. Create a dir to move the file from the GCP bucket into the HDFS cluster environment
     + Command > mkdir DATA  
       A screenshot of a computer

       Description automatically generated
  2. Move to Data dir
     + CD DATA
  3. Copy GCP Bucket into the Master Node
     + gsutil cp gs:// bucket\_adta5240bin/data/userdata.csv userdata.csv
     + gsutil cp gs:// bucket\_adta5240bin/data/weblog.csv weblog.csv
  4. Move from Master node to the HDFS ecosystem
     + userdata > hdfs dfs -put userdata.csv /user/leannboyce19/data/ userdata /
     + weblog > hdfs dfs -put weblog.csv /user/leannboyce19/data/weblog/

A screenshot of a computer program

Description automatically generated

* Steps I took to check if the .csv files were copied into HDFS and included screenshots.
  1. I run > hdfs dfs -ls /user/biniamabebe/data/weblog to check that the .csv file is moved to the userdata subdirectory
  2. Run > hdfs dfs -ls /user/biniamabebe/data/weblog to check that the .csv file is moved to the weblog subdirectory  
     A black and white screen

     Description automatically generated
* Screenshot that illustrates your three nodes are turned off in GCP. A screenshot of a computer

  Description automatically generatedA screenshot of a computer

  Description automatically generated

Reference and definition

1 Google Cloud. (n.d.). Monitoring Bigtable instances. Retrieved January 28, 2024, from https://cloud.google.com/bigtable/docs/monitoring-instance