Run flume-ng version command and take a screen shot showing that flume is installed

![Text

Description automatically generated]()

flume-ng agent --conf ./conf/ -f ../conf/flume.conf -n agent1 -Xmx512m -Dflume.root.logger=DEBUG

Take a screenshot of the successful output.

![Text

Description automatically generated]()

Use a 2nd SSH connection, use the hdfs cat command to output the content copied to the /flume01 folder. Take a screen shot of this.

![Text

Description automatically generated]()

Questions:

1. Describe the purpose of Apache Flume ( 4 marks)
2. What is a Flume agent? ( 2 marks)
3. What is a Flume source? ( 2 marks) Give an example ( 2 marks)
4. What is a Flume sink? ( 2 marks) Give an example ( 2 marks)
5. Describe the purpose of Apache Flume.

Apache Flume is a tool for collecting and moving very large amounts of stream data from various sources. The main purpose of Apache Flume is to update the collection and intake of large datasets into HDFS or other data stores.

1. What is a Flume agent?

A Flume agent is a Java virtual machine process that hosts the components through event flows. The agent is responsible for collecting and filtering data to a destination. There are three main components of a flume agent sources, channels and sinks.

1. What is a flume source . Give an example.

A flume source is a component that is responsible for taking in data from various sources and makes it ready for processing by the flume agent. The flume source is the input factor of an Apache Flume job. An example of a flume source would be like the step we have done in our lab

agent1.sources.tail.type = exec  
agent1.sources.tail.command = cat /home/barryadam2018/access.log  
agent1.sources.tail.channels = Channel-2

1. What is a flume sink. Give an example.

A flume sink is a component of the flume agent it is used for storing data into HDFS. Sinks are the end point of the data within the flume agent. An example of a flume sink would be like the step we had done in step 4 here is snippet of the code.

agent1.sinks.sink-1.channel = Channel-2  
agent1.sinks.sink-1.type = hdfs  
agent1.sinks.sink-1.hdfs.path = /flume01  
agent1.sinks.sink-1.hdfs.fileType = DataStream  
agent1.sinks.sink-1.hdfs.rollInterval = 60  
agent1.sinks.sink-1.hdfs.rollSize = 0  
agent1.sinks.sink-1.hdfs.rollCount = 0