This exam serves as the assessment for those students who cannot utilize the Hadoop system and/or Ambari GUI.

1. Compare and contrast the Hadoop and AWS big data experiences.

HAdoop only works with big data, while AWS does work with big data it can also do things unrelaed to big data too. AWS is more up to date & hadoop is not really supported anymore , it is an older program. AWS requires money or school credit & hadoop is free. AWS allows the use of existing programs, hadoop usesnovel programs. Both can be used for big data though.

1. Which Hadoop tool do you think is most useful, and why?

HDFS because it can write and read files, store data, and distribute data.

1. What are the Vs of big data?

volume, varitey, velocity, veracity, value

1. Why is Hadoop important to big data? What functions does it provide?

Because of map reduce, the foundation of big data processing in hadoop. Also because of hive, allowing sql queries to interact with the data

1. Explain the concept and overall process of data streaming.

data streaming is real time data coming in by the second and being processed for analyzation by a machine or program

1. Compare and contrast Pig and Hive.

both are very flexible.

pig is harder to use compared to hive.

both are for data sets but have different names for thier datasets, pig calls them relations and hive calls them tables. Hive is quicker,Pig is slower. They do not use the same language.

1. Why is MapReduce important?

its the foundation of big data, the original big data system

it helps you find the specif data needed towork with and it aggregates large data.

1. What area of big data would you like to learn more about and why?
2. Map reduce and dask. map reduce becuase its an original one and dask because its newer so i feel like it may be introduced and used in more settings over time.