* ETL stands for extract transform load

* memory becomes a limiting factor for scaling if we keep using a single namenode HDFS "DATANODES" allows the cluster to scale by adding name nodes.each of which manages a portion of the filesystem namespace

* before YARN was added to Hadoop. we had to start the "JOBTRACKER" and the task tacker daemons to allow jobs to run in the cluster

* Hadoop is meant for batch processing rather than real time processing "False".Hadoop is cpapble of processing data in real time

* In contrast to mapreduce 1, YARN resource manager will keep trackof the available resource and the running jobs while each application will track its own process"TRUE'
* data locality in mapreduces applies only to the mappers rather than the mappers and reducers "TRUE"
* hadoop distcp is much more efficient than HDFS cp command for coping data to and from hadoop filessystems in parallel "True"
* the only programming language that can be used in mapreduce applications is java "false"
* In mapreduce 1 we used containers to manage and allocate resources resources among jobs."FALSE"
* To use HDFS serivices we have to start the "NAME NODE" and "DATA NODE" daemons on the master and worker machines respectively.
* Among YARN available schedulers, the FIFO scheduler is the one recommended for shared cluster."FALSE"
* In Hadoop the idea of running the computation where the data resides is called "data locality".
* In HDFS HA, graceful failover that is initiated manually by the administrator usually for routine maintenance."FALSE"
* What is the YARN command to submit and run the job in Hadoop?
  "yarn jar <jar_file> <class_name>"
* What is the HDFS command to show the whole output in the terminal?
  "hdfs dfs -cat".
*
* what is a dedicated HDFS implementation used in HDFS HA to provide a highly available shared edit log. It is recommended to choice for the most HDFS instrallation?
The recommended dedicated HDFS implementation for HDFS HA is the Quorum Journal Manager (QJM). QJM provides a highly available shared edit log that is used by the Namenode to store changes to the file system.

* When we submit an application to YARN, the application will use the first allocated container to launch and run its "APPLICATION MASTER" process.

* command used to start YARN in the VM is?
  "yarn start". Yarn start will start the
  resource manager and the node managers

* HDFS is an efficient distributed file system for reading and writing large amount of data in parallel. “FALSE”

* The HDFS command to display the content of the root directory in HDFS is "hdfs dfs -ls/"
* Setting the mappers and reducers classes for mapReduce application in the driver code is optional."False"
* The mapper and reducer write their output to HDFS."true"