-- LAB 8 ---

-- QUESTION 4

Scaling up is making a component bigger or faster so that it can handle more load. Scaling out is adding more components in parallel to spread out a load.

-- QUESTION 9

The Hadoop Distributed File System is a distributed file system designed to run on hardware based on open standards. This means the system can run different operating systems such as Windows or Linux without requiring special drivers. It is not designed for user interaction. It is used for batch processing of applications that need streaming access to their datasets. Hardware failure is the norm rather than the exception. A Hadoop Distributed File System instance may consist of hundreds or thousands of server machines, each storing part of the file system’s data.

-- QUESTION 14

Document databases organize documents into groups called collections, which are analogous to the tables in relational databases. By contrast, key-value databases store all key-value pairs together in a single namespace, which is analogous to a relational schema.

-- QUESTION 19

The data mining process is heavily based on algorithms to analyze and extract information that automatically discovers hidden patterns and relationships within the data. Within predictive analytics, the process uses data patterns to make predictions with machine learning.

-- QUESTION 20

The data mining process is classified in two stages: Data preparation/data preprocessing and data mining. The data preparation process includes data cleaning, data integration, data selection, and data transformation. The second phase includes data mining, pattern evaluation, and knowledge representation.