NIRMA UNIVERSITY

Institute of Technology

Class Test-1, August 2021

Big Data Analytics (2CS702), B.Tech in CSE– Semester VII

Time: 1.15 Hours Max Marks: 35

Q: 1 **[08]**

Agriculture is the backbone of the Indian economy and contributes to up to 23% to Indian Gross Domestic Product (GDP) and in recent years Data analytics is widely used in the agricultural domain to improve yield. Viticulture (the cultivation of grapes) is one of the most lucrative farming in India.. It is a subdivision of horticulture and is the study of wine growing. The demand for Indian Wine is increasing at about 27% each year, worldwide since the 21st century. Thus more and more ways are being developed to improve the quality and quantity of Viticulture. We focus on a specific agricultural practice as viticulture. Weather forecasting and leaf disease detection are the two main research areas in precision viticulture considering following main objectives:

- Precision Sensors: These sensors are deployed between the grapevines, and they gather the acute weather details of the grape plant such as leaf moisture and humidity.
- Microclimate Monitoring: Precision monitoring of climatic conditions per plant measuring its humidity so that monitoring the intensity of fungal diseases becomes possible.
- Grape Disease Prediction: Monitoring the leaf images of the grape plant can help to identify the early onset of diseases which can be treated or the bad plant may be removed to protect the rest.

Study the above case study and add your assumption if required to answer the following questions.

A) How can Big Data Analytics be useful in the development of Advanced Viticulture? Discuss challenges under Big Data based solution. [4]

B) Write down all the type of data generated in above case study and classify according to "Variety" dimension of Big Data, Also discuss the role of volume and velocity. [4]

Q: 2 State whether the following statements are **true, partially true or false** with **proper reason and justification**. **[08]**

a. Peer-to-Peer network performs poor for real time processing based applications.

b. Input data file kept on HDFS by client will be splitted into packets by DFSOutputStream.

c. In Hadoop distributed file system cp command is used to copy the data from local file system to HDFS.
d. Role of secondary namenode is to periodically merge FSimage and editlogs from namenode.

Q: 3 Answer the following questions **[19]**

a. Write down the brief functioning of following terms **[5]**

1. Shared Nothing Architecture
2. Vertical Scaling
3. Dataqueue in HDFS

b. How does the namenode choose which datanodes to store replica on? Explain the replica management in HDFS in detail. **[4]**

c. Draw and explain the approach adopted by Hadoop to find the distance of two nodes in a Network. **[3]**

d. What do you mean by "Fault Tolerance"? Compare and justify the fault tolerance performance of horizontal scaling platforms and vertical scaling platforms. **[3]**

e. Discuss the role of a client in HDFS architecture. **[2]**

f. Write down the ways and methodology to deal with unstructured data. **[2]**