Faculty of Science
End Sem (Even) Examination May-2019
CA3EL14 Big Data Analytics

Programme: BCA   Branch/Specialisation: Computer Application

Duration: 3 Hrs.   **Maximum Marks: 60**

Note: All questions are compulsory. Internal choices, if any, are indicated. Answers of Q.1 (MCQs) should be written in full instead of only a, b, c or d.

Q.1 i. According to analysts, for what can traditional IT systems provide a foundation when they are integrated with big data technologies like Hadoop?   **1**
(a) Big data management and data mining
(b) Data warehousing and business intelligence
(c) Management of Hadoop clusters
(d) Collecting and storing unstructured data

ii. What are the main components of Big Data?   **1**
(a) MapReduce   (b) HDFS
(c) YARN   (d) All of these

iii. Which of the following is the first layer in the 'Big data stack'?   **1**
(a) Processing layer   (b) Data layer
(c) Analytics layer   (c) Integration layer

iv. Which of the following characteristics should you consider when deciding whether to deploy an application or services to a virtual machine?   **1**
(a) Hardware requirement
(b) Software support and compatibility
(c) Licensing
(d) All of these

v. Which of the following is not the phase of data analytics life cycle?   **1**
(a) The data preparation   (b) Model planning
(c) Data removal   (d) Operationalization

vi. Data analytics life cycle is not cover   **1**
(a) Extraction of data   (b) Analysis of data
(c) Visualization of data   (d) Removal of data

vii. MapReduce was devised by _____.   **1**
(a) Apple   (b) Google   (c) Microsoft (d) Samsung

viii. Which of the following is not a phase of Reducer?   **1**
(a) Map   (b) Reduce   (c) Shuffle   (d) Sort

ix. Which of the following is the data types in Hive?   **1**
(a) ARRAY   (b) STRUCT (c) MAP   (d) All of these

x. Which of the following is the Key components of Hive Architecture?   **1**
(a) User Interface   (b) Metastore
(c) Driver   (d) All of these

Q.2 i. Explain the steps to be followed to deploy a Big Data solution.   **2**
ii. What do you know about the term "Big Data"?   **3**
iii. What are the three V's of Big Data?   **5**
OR iv. Describe decision support system and data warehousing.   **5**

Q.3 i. What is Big Data Virtualization?   **2**
ii. Compare Emerging analytics and Advance analytics with suitable example.   **8**
OR iii. Compare business intelligence and data science.   **8**

Q.4 i. Compare data discovery, data aggregation in brief?   **3**
ii. What is data analytics life cycle?   **7**
OR iii. What are the different features of big data analytics?   **7**

Q.5 i. Explain the difference between Hadoop and RDBMS.   **4**
ii. How does HDFS ensure data integrity in a Hadoop cluster?   **6**
OR iii. What is Shuffling and Sorting in MapReduce?   **6**

Q.6   Attempt any two:
i. Explain Hive Data Units with appropriate uses.   **5**
ii. Define the difference between Hive and HBase?   **5**
iii. What are the key components of HBase?   **5**

P.T.O.

*****

Marking Scheme
CA3EL14 Big Data Analytics

Q.1　i.　According to analysts, for what can traditional IT systems provide a foundation when they are integrated with big data technologies like Hadoop?　**1**
(a) Big data management and data mining

　ii.　What are the main components of Big Data?　**1**
(d) All of these

　iii.　Which of the following is the first layer in the 'Big data stack'?　**1**
(b) Data layer

　iv.　Which of the following characteristics should you consider when deciding whether to deploy an application or services to a virtual machine?　**1**
(d) All of these

　v.　Which of the following is not the phase of data analytics life cycle?　**1**
(c) Data removal

　vi.　Data analytics life cycle is not cover　**1**
(d) Removal of data

　vii.　MapReduce was devised by _____.　**1**
(b) Google

　viii.　Which of the following is not a phase of Reducer?　**1**
(a) Map

　ix.　Which of the following is the data types in Hive?　**1**
(d) All of these

　x.　Which of the following is the Key components of Hive Architecture?　**1**
(d) All of these

Q.2　i.　Steps to be followed to deploy a Big Data solution.　**2**
Description of steps

　ii.　Big Data　**3**
Definition　2 marks
Characteristic　1 mark

　iii.　Three V's of Big Data　**5**
1 mark for each (1 mark * 3)　3 marks
For diagram　2 marks

OR　iv.　Decision support system　2.5 marks　**5**
Data warehousing.　2.5 marks

Q.3　i.　Definition of Big Data Virtualization　**2**

　ii.　Comparison b/w Emerging analytics and Advance analytics　**8**
1 mark for each comparison　(1 mark * 8)

OR　iii.　Comparison b/w business intelligence and data science.　**8**
1 mark for each comparison　(1 mark * 8)

Q.4　i.　Comparison b/w data discovery, data aggregation　**3**
For each point 0.5 mark　(0.5 mark * 6)

　ii.　Data analytics life cycle　**7**
Description　4 marks
Diagram　3 marks

OR　iii.　Different features of big data analytics　**7**
At least seven features 1 mark each　(1 mark * 7)

Q.5　i.　Difference b/w Hadoop and RDBMS.　**4**
At least eight differences 0.5 mark for each　(0.5 mark * 8)

　ii.　HDFS ensure data integrity in a Hadoop cluster　**6**
Description of HDFS & Hadoop cluster　3 marks
Description of ensurance of data integrity　3 marks

OR　iii.　Description of Shuffling　3 marks　**6**
Description of Sorting in MapReduce　3 marks

Q.6　　Attempt any two:
　i.　Hive Data Units　3 marks　**5**
Uses.　2 marks

　ii.　Difference b/w Hive and HBase　**5**
1 mark for each point　(1 mark * 5)_

　iii.　Components of HBase　3 marks　**5**
Diagram　2 marks

******