Section I

MCQ

Q1 to 5 - 1 point each

1. RDBMS is a perfect match for

a. Transaction Processing.

b. Customer Churn Rate Calculation.

c. Serial Returner Detection.

d. Potential Loan Defaulter Identification.

e. Market Basket Analysis.

Answer: a

2. Which one of the below is a limitation of NoSQL?

a. Bound to adhere CAP theorem and primarily supports BASE

b. Handles semi-structured data

c. Range of solutions with various types

d. Horizontal scalability

e. Handles Big Data Characteristics

Answer: a

3. Which one of below is not correct about Data Science?

a. Data Science is replacement of OLTP and OLAP.

b. Data Science is amalgamation of Business Acumen, Mathematics and Technology.

c. Data Science has evolved primarily because of Big Data.

d. Data Science has to still overcome challenges i.e. ask right question with right data.

e. Data Science is about gaining insight from data and improvise decision making.

Answer: a

4. Which one of below shall not be a requirement for cluster setup?

a. Unified Operating System Platform across every nodes

b. Establishing User Equivalence

c. Networking Channels

d. Naming and Resolution - IP/Hostname

e. Storage and Processing

Answer: a

5. Which one of below most concerns the client when adapting to Big Data Solutions? (next to budget)

a. Information Security

b. Data Streaming

c. Volume of Data

d. Infrastructure

e. Talent

Answer: a

Q6 to 10 - 2 points each

6. Provided below probable statements about evolution of analytics, which choice is incorrect?

- Diagnostic Analytics has to do with why did it happen.

- Prescriptive Analytics has to do with how can we make it happen.

- Predictive Analytics has to do with what will happen.

- Descriptive Analytics has to do with what happened.

a. Hadoop Eco System and related components aim to achieve at most predictive analytics

b. Hindsight to insight to Foresight

c. Industry 4.0

d. Big Data and Analytics is important as it aims for prescriptive analytics.

e. IoT, Cloud and ML - all contribute to data handling and analytics

Answer: a

7. Which one of below quotes is not completely self-describing w.r.to Big Data and Analytics?

a. Data is new oil.

b. Without data you are just another person with an opinion.

c. Without big data analytics, companies are blind and deaf, wandering out onto the web like deer on a freeway.

d. Data is the new science. Big Data holds the answers.

e. In God we trust. All others must bring data.

Answer: a

8. In accordance with Hadoop Eco System various stacks, what might be correct ordering to achieve one type of big data analytics solution.

a. HDFS -> Map Reduce -> Hive

b. MongoDB -> Scoop -> Map Reduce

c. Pig -> Ambari -> Hive

d. HDFS -> Map Reduce -> HBase

e. HBase -> MongoDB -> HDFS

Answer: a

9. Which one of below is a open source data format and serialization system?

a. Avro

b. Azure Databricks

c. JasperSoft

d. JSON

e. Ambari

Answer: a

10. Which one of the below is not a valid phase in build lifecycle using apache maven?

a. build

b. test

c. verify

d. install

e. deploy

Answer: a

Section II

2 marks question

1. What are different types of NoSQL databases? Mention with example.

2. Differentiate tumbling window and sliding window diagrammatically with respect to windowing protocols for stream data analytics.

3. How is JasperSoft tool utilized? Provide steps.

5 marks question

1. Provide HDFS Commands for below independent requirements :

(i) Create directory named hive\_warehouse in hdfs at root /

(ii) Copy all .txt files from host file system location /home/user1/data to hdfs location /user/hadoop/input

(iii) Display the content of all files within hdfs location /user/hadoop/outut

(iv) Move file test.txt from hdfs location /user/hadoop/datalc1 to /user/hadoop/datalc2

(v) Receive all files from hdfs location /user/hadoop/project to local.

2. Provide MongoDB solutions to below independent topics:

(i) Start mongoDB and confirm that it is up and running via showing all existing databases.

(ii) In the selected database create a collection named 'userlog'.

(iii) Display the content of collection named 'serviceHistory' in JSON formatted manner.

(iv) Provided mapp and reducee named js function as per api requirement, execute them on 'people' collection and record output in collection named 'population'.

(v) Display avg salary for male employees department wise in ascending order.

3.

Provide suitable brief explanation to support each below:

(i) Data Silos Vs Data Lakes

(ii) Client Server Vs Master Slave architecture

(iii) Loosely coupled Vs Tightly coupled

(iv) Data at rest Vs Data in motion

(v) Schema on read Vs Schema on write