

Seat No. _____

Set P

**Fourth Y. (B. Tech) (Sem - II) (New) (CBCS) Examination:
March/April-2023**

**COMPUTER SCIENCE & ENGINEERING
Bigdata Analytics (197043805)**

Day & Date: Friday 23-06-2023

Max. Marks: 70

Time: 03:00 PM To 06:00 PM

- Instructions:** 1) Q. No. 1 is compulsory. It should be solved in the first 30 minutes in answer book. Page no 03 (Starting page of the Answer Book). Each question carries one mark.
 2) Don't forget to Mention question paper set (P/Q/R/S) on top of page.
 3) Figures to the right indicates full marks.

MCQ/Objective Type Questions

Duraction: 30 Minutes

Marks: 14

Q.1 Choose the correct alternatives from the given options.

14

- 1) _____ is a collection of data that is used in volume, yet growing exponentially with time.
 - a) Big Database
 - b) Big DBMS
 - c) Big Datafile
 - d) Big Data
- 2) The term NoSQL was first coined by _____.
 - a) Doug Laney
 - b) Carlo Strozzi
 - c) Brewer
 - d) Gartner
- 3) All of the following accurately describe Hadoop, except
 - a) Open Source
 - b) Java Based
 - c) Real Time
 - d) Distributed computing approach
- 4) In Big Data environment, Veracity of data refers _____.
 - a) Quality or fidelity of data
 - b) Large size of the data that cannot be process
 - c) Small size of the data that can easily process
 - d) All of the mentioned above
- 5) Amongst which of the following is/are not Big Data Technologies?
 - a) Apache Hadoop
 - b) Apache Spark
 - c) Apache Kafka
 - d) Apache Pytarch
- 6) Which of the following is a tool to transfer data between Hadoop and Relational Databases?
 - a) Sqoop
 - b) HBase
 - c) Hive
 - d) Pig
- 7) Data that does not conform to a data model or data schema is known as _____.
 - a) Structured data
 - b) Unstructured data
 - c) Semi-structured data
 - d) All of the mentioned above

23/06/2023
23/06/2023

- 8) Data in _____ bytes size is called Big Data.

 - a) Tera
 - b) Giga
 - c) Peta
 - d) Meta

9) CCTV footage is which type of Data?

 - a) structured
 - b) semi structured
 - c) Unstructured
 - d) None of above

10) MongoDB is a _____ database.

 - a) SQL
 - b) DBMS
 - c) NoSQL
 - d) RDBMS

11) Cassandra is a _____ database.

 - a) Document-oriented
 - b) Graph-oriented
 - c) Column-oriented
 - d) SQL

12) ETL processing in Pig stand for _____.

 - a) Extract, transform and load
 - b) Extend transfer and load
 - c) Extract, transform and local
 - d) None of the above

13) Hive provides _____ kinds of partitions.

 - a) Static
 - b) Dynamic
 - c) Both Static and dynamic
 - d) Neither static nor dynamic

14) PIG is _____.

 - a) dataflow language
 - b) NoSQL database
 - c) import export tool
 - d) scheduling engine

Fourth Y. (B. Tech) (Sem - II) (New) (CBCS) Examination:

March/April-2023

**COMPUTER SCIENCE & ENGINEERING
Bigdata Analytics (197043805)**

Day & Date: Friday 23-06-2023

Time: 03:00 PM To 06:00 PM

Max. Marks: 56

Instructions: 1) All questions are compulsory.

2) Figures to the right indicate full marks.

3) Assume suitable data if necessary.

Section – I

Q.2 Attempt any four of the following questions.

- a) What do you mean by Unstructured Data? Explain sources of Unstructured data.
- b) List different characteristics of Big Data & Explain different challenges of Big Data.
- c) Describe the anatomy of File Read in HDFS.
- d) Explain YARN Architecture in detail.
- e) What is Big Data Analytics? Explain second schools of thought of Big Data Analytics.

16

Q.3 Attempt any two of the following questions.

- a) Explain various terminologies used in Big Data environments.
- b) Explain High Level Architecture of Hadoop.
- c) How does MapReduce work? Explain with suitable example.

12

Section – II

Q.4 Attempt any four of the following questions.

- a) List four important differences between SQL and MongoDB.
- b) Write word count example using PIG. Explain Pig Latin Statement.
- c) What is static and dynamic partitioning in HIVE? Illustrate with example.
- d) Write features of Cassandra.
- e) Explain pig philosophy also Explain ETL processing in pig.

16

Q.5 Attempt any two of the following questions.

- a) Draw the architecture of Hive and explain its components.
- b) Explain Map reduce Programming in Mongo DB with suitable example.
- c) Explain different CRUDE operation in Cassandra.

12

823331

PUNYASHLOK AHILYADEVI HOLKAR, SOLAPUR UNIVERSITY, SOLAPUR
 University Examination - March /April 2022
B.TECH Part- Sem-ii
Subject -- Big Data Analytics

Paper Code-197043805

SLR- FI-186 Set-A

Time: _____

Max. Marks: _____

Q1.**Choose the machine generated structured data**

A. Scientific data	B. Social media data
C. point-of-scale	D. postgresSql

Q2.

Arrange the following in accordance with their transformation (lower to higher)

- i. Information
- ii. Data
- iii. Knowledge
- iv. Actionable Insights

A. i, ii, iii, iv	B. ii, i, iii, iv
C. ii, iii, i, iv	D. iii,iv,ii,i

Q3.**System will continue to function even when network partition occurs is ----- in CAP Theorem**

A. Consistency	B. Availability
C. Partition Tolerant	D. None of the above

Q4.**In Hadoop, ----- manages file related operations and keeps track of blocks?**

A. Data Node	B. Name Node
C. Secondary Name Node	D. Client node

Q5.**Which of the following is true about Apache Cassandra?**

A. Apache Cassandra is a free and open-source	B. Apache Cassandra is a distributed
C. Apache Cassandra has wide column store	D. All of the above

Q6.

In Cassandra, what do the following query does? DESCRIBE TABLE student_info;

A.

Look up the names of all tables

B.

Displays the contents of the student_info table

C.

Describes all the tables in the database

D.

Describes the table student_info with its details

Q7.

"Heartbeat" message is sent by ----- node to ----- node

A.

Data, Name

B.

Name, Data

C.

Secondary Name, Name

D.

Name, Secondary Name

Q8.

Which of the following is an Analytics Tool?

A.

SAS

B.

IBM SPSS Modeler

C.

Statistica

D.

All of the above

Q9.

What is Big Data is not?

A.

Only used by huge online companies like Google or Amazon

B.

Better and faster decisions in real time

C.

It is collaboration with business users and data scientists

D.

Richer deeper insights

Q10.

What are the Hadoop core components?

A.

Data and code

B.

HBase and Flume

C.

HDFS and MapReduce

D.

Data storage and Daemons

Q11.

Which of the following is equivalent to SQL command-
where designation = "Manager" ?

Select * from employees

A.

employees.find({"designation": "Manager"})

B.

db.employees.find({"designation": "Manager"})

C.

db.employees.find({"designation": "Manager"})

D.

db.employees.find({Designation="Manager"})

Q12.

You can run Pig in batch mode using _____

A. Pig shell command

B. Pig scripts

C. Pig options

D. All of the mentioned

Q13.

Choose the human generated unstructured data

A. Scientific data

B. Social media data

C. point-of-sale

D. Mobile data

Q14.

Cassandra uses a protocol called _____ to discover location and state information.

A. gossip

B. intergos

C. goss

D. all of the mentioned

Q15.

Analytics 3.0 provides -----

A. Descriptive Statistics

B. Predictive Statistics

C. Prescriptive Statistics

D. All

Q16.

Apache Cassandra is a massively scalable open source _____ database.

A. SQL

B. NoSQL

C. NewSQL

D. All of the mentioned

Q17.

Which configuration file is used to control the HDFS replication factor?

A. mapred-site.xml

B. hdfs-site.xml

C. core-site.xml

D. yarn-site.xml

Q18.

- Consider the mongodb database inventory collection containing an array called ratings to the post by various users in the following format:

```
{
  _id: 1,
  post_text: "This is my first post",
  ratings: [5, 4, 2, 5],
  //other elements of document
}
```

Choose a query for all the documents where the array ratings contains at least one element between 3 and 6?

A.

```
db.inventory.find( { ratings: { $elemMatch: { $gt: 3, $lt: 6 } } } )
```

B.

```
db.inventory.find( { ratings: { $elemMatch: { $gt: 3, $lt: 6 } } } )
```

C.

```
db.inventory.find( { ratings: { ratings.$: { $gt: 3, $lt: 6 } } } )
```

D.

```
db.inventory.find( { ratings: { $elemMatch: { $gte: 3, $lte: 6 } } } )
```

Q19.

Identify the nodes which are called daemon of Hadoop?

A.

Name Node

B.

Node manager

C.

Data Node

D.

All of the above

Q20.

In MongoDB, choose the command for displaying the list of all databases

A.

use myDB;

B.

show dbs;

C.

db.show();

D.

display dbs;

Q21.

Which of the following command is correct when we want to fetch documents from collection for "only those employees whose salary is either 8500 or 10000"?

A.

```
db.employees.find.sort({ "salary": {$in:[8500,1000]} })
```

B.

```
db.employees.find({ "salary": {$in:[8500,1000]} })
```

C.

```
db.employees.find({ "salary": {"$in:[8500,1000]} })
```

D.

```
db.employees.find({ "salary": {"$in:{8500,1000}} })
```

Q22.

Which of the following line skips the first 10 documents in the bios collection and returns all remaining documents?

A. db.bios.find().limit(10)	B. db.bios.find().skip(-10)
C. db.bios.find().skip(10)	D. db.bios.find().sort(10)

Q23. Which of the following is used by Hive?

A. HDFS for storage	B. Map Reduce for execution
C. RDBMS for metadata storage	D. All of the above

Q24. In MongoDB ----- is called as Table in RDBMS.

A. Database	B. Document
C. Collection	D. Key-Value pair

Q25. Choose the relational operators in Pig for removing duplicate tuples?

A. GROUP	B. DISTINCT
C. LIMIT	D. ORDERBY

Q26. Command to check existence of collection is _____

A. show collection	B. show collections
C. show collect	D. Define Collection

Q27. Choose the external data sources for the Big Data ?

A. Public Web	B. Data storage
C. Archives	D. All

Q28. What category will you place a CCTV footage in to?

A. Structured	B. Semistructured
C. Unstructured	D. All of the above

Q29.

What is predictive analysis?

A.

What will happen

B.

What should be the action of what will happen

C.

Where it is happened

D.

What is the time of happening

Q30.

In Cassandra to add a list column "language" in a table student_info, which of the command is used?

A.

ALTER TABLE student_info INSERT language
list<text>

B.

ALTER TABLE student_info ADD language
list<text>

C.

UPDATE TABLE student_info ADD language
list<text>

D.

UPDATE TABLE student_info ADD langua
as list<text>

Q31.

----- is the characteristic of data explains spikes in data?

A.

Volatility

B:

Variability

C.

Velocity

D.

Volume

Q32.

In MongoDB, choose the query for: "to find those documents from the food collection which have the fruits array having grapes in the first index positon"

A.

db.find.food({fruits.1:grapes});

B.

db.food.find({fruits.1:grapes});

C.

'db.food.find({fruits.1:'grapes'})';

D.

All of the above statements

Q33.

----- are the collections of key-value pairs in Cassandra

A.

Lists

B.

Maps

C.

Tuples

D.

Set

Q34.

Choose the NoSQL database?

A.

CouchDB

B.

MySQL

C.

IBM DB2

D.

ACID

Q35.

Which of the followings are the components of the Hadoop Ecosystem?

- | | |
|-------------|------------------------|
| A.
Sqoop | B.
ZooKeeper |
| C.
Oozie | D.
All of the above |

Q36.

According to CAP theorem, Consistency and Availability (CA) is applied to -----

- | | |
|-----------------|---------------|
| A.
Cassandra | B.
MongoDB |
| C.
MySQL | D.
HBase |

Q37.

The ----- technology helps query data that resides in a computer's RAM rather than data stored on physical disks

- | | |
|--|---------------------------|
| A.
Massively Parallel Processing | B.
In-Memory analytics |
| C.
Symmetric Multiprocessing System | D.
Distributed Systems |

Q38.

Point out the correct statement

- | | |
|--|---|
| A.
Cassandra delivers continuous availability, linear scalability, and operational simplicity across many commodity servers | B.
Cassandra has a "masterless" architecture, meaning all nodes are the same |
| C.
Cassandra also provides customizable replication, storing redundant copies of data across nodes that participate in a Cassandra ring | D.
All of the mentioned |

Q39.

MongoDB is ----- and -----

- | | |
|--|--|
| A.
Partition Tolerant and Available | B.
Consistent and Partition Tolerant |
| C.
Available and Consistent | D.
Available, Consistent and Partition Tolerant |

Q40.

Which of the following command is correct when we want to fetch documents from a collection for "only those employees whose salary is either 8500 or 10000"?

- | | |
|---|--|
| A.
<code>db.employees.find.sort({"salary": {\$in:[8500,1000]}})</code> | B.
<code>db.employees.find({"salary": {\$in:[8500,1000]}})</code> |
|---|--|

C. db.employees.find("salary":
{"\$in:[8500,1000]}))

D. db.employees.find("salary":
{\$in:{8500,1000}}))

Q41. Which of the following is not true about Pig?

- A. Apache Pig is an abstraction over MapReduce
C. Pig is a tool/platform which is used to analyze larger sets of data representing them as data flows.

- B. Pig can not perform all the data manipulation operations in Hadoop
D. None of the above

Q42. Hadoop focuses on moving ----- to -----.

- A. Code, Code
C. Code, Data

- B. Data, Data
D. Data, Code

Q43. Choose the responsibilities of the Data Scientist?

- A. Data Management
C. Applies Analytical techniques for interpretation of data

- B. Business Analytics
D. All of the above

Q44. Which of the following is not a Features of HiveQL?

- A. Supports joins
C. Support views

- B. Supports indexes
D. Support Transactions

Q45. Select the characteristics of Semi-structured data

- A. Inconsistent Structure
C. Schema blended with value

- B. Self describing
D. All

Q46.

In Hive, what is the use of a following command? LOAD DATA LOCAL INPATH '/root/hivedemos/student.csv' OVERWRITE INTO TABLE EXT STUDENT;

Seat
No.

B.E. (Computer Science & Engineering) (Part – II) (CGPA) Examination

April/May-2019

BIG DATA ANALYTICS (Elective – IV)

Day & Date: Monday, 20-05-2019

Max. Marks: 70

Time: 02:30 PM To 05:30 PM

Instructions: 1) Figures to the right indicate full marks.

- 2) First question is objective type.
- 3) All questions are compulsory.
- 4) Q. No. 1 is compulsory. It should be solved in first 30 minutes in Answer book page no. 3. Each question carries one mark.

MCQ/Objective Type Questions

Duration: 30 Minutes

Marks: 14

14

Q.1 Choose the correct answer

- 1) In _____ architecture, the central memory is shared by multiple processors.

a) Shared Memory	b) Shared Disk
c) Both a and b	d) None of the above
- 2) The term NoSQL was first coined by _____.

a) Doug Laney	b) Carlo Strozzi
c) Brewer	d) Gartner
- 3) The real time processing deals with which of the following characteristics?

a) Volume	b) Volatility
c) Variability	d) Velocity
- 4) In which of the following analysis, Data is descriptive, predictive and prescriptive?

a) Analytics 1.0	b) Analytics 2.0
c) Analytics 3.0	d) None
- 5) _____ does not support ACID properties of transaction.

a) SQL	b) NoSQL
c) NewSQL	d) All
- 6) HDFS breaks a large file into smaller pieces called _____.

a) Blocks	b) Rack
c) Map	d) Disk
- 7) In anatomy of file write, as the client writes data, data is split into packets by DFSOutputStream, which is then written to an internal queue called _____.

a) Input Queue	b) Data Queue
c) Ack Queue	d) Output Queue
- 8) Which of the following statement is true?

a) Cassandra does not have master-slave architecture
b) Cassandra was born at Facebook
c) Cassandra is highly scalable, high performance distributed database
d) All

- 9) In Cassandra, which type of collection maintains the order of elements?

 - a) Set Collection
 - b) List Collection
 - c) Map Collection
 - d) All

10) Cassandra is a column-oriented database designed to support _____ symmetric node architecture.

 - a) Peer-to-peer
 - b) Master-slave
 - c) Both a & b
 - d) None of these

11) Pig is _____.

 - a) Data Flow language
 - b) NoSQL database
 - c) Import Export tool
 - d) Scheduling engine

12) Hive is a _____ tool.

 - a) Data Flow
 - b) Data transfer
 - c) Column-oriented
 - d) Data warehousing

13) _____ analytics is used to drive direct business revenue.

 - a) Basic Analytics
 - b) Operationalized Analytics
 - c) Advanced Analytics
 - d) Monetized Analytics

14) The most commonly used interface to interact with Hive is _____.

 - a) Command Line Interface
 - b) Graphical User Interface
 - c) Both a and b
 - d) None

Seat
No.

(i.e.) (Computer Science & Engineering) (Part – II) (CGPA) Examination
April/May-2019

BIG DATA ANALYTICS (Elective – IV)

Day & Date: Monday, 20-05-2019
Time: 02.30 PM to 05.30 PM

Max. Marks: 70

Instructions: 1) All questions are compulsory.

2) Figure to the right indicate full marks.

Section – I

Q.2 Attempt any Three

- a) What is Big Data Analytics? Explain second schools of thought of Big Data Analytics.
- b) What are the Sources of Structured data? Why it is easy to work with structured data?
- c) Explain Hadoop Ecosystem.
- d) Describe the anatomy of File Read in HDFS.
- e) Explain Hadoop 1.0 and Hadoop 2.0

12

Q.3 Attempt any Two

- a) How data is processed with Hadoop? Explain with suitable example.
- b) Write a short note on Data Science, and CAP Theorem.
- c) What are the limitations of Hadoop 1.0 architecture? Explain YARN Architecture in detail.

16

Section – II

Q.4 Attempt any Three

- a) Explain Map Reduce programming in MongoDB with suitable example.
- b) Explain Hinted Handoffs feature of Cassandra with diagram.
- c) Write difference between SQL and MongoDB.
- d) Write MongoDB query for an objective given below:
Create a collection "food" and insert 5 documents into the food collection by "_id" and "fruits" array. Check these documents are present in food collection. Find those document from food collection where "grapes" is present in 2nd index position of the fruit array.
- e) What is Metastore in Hive? Explain 3 kinds of Metastore.

12

Q.5 Attempt any Two

- a) Write a short note on JasperReports, ETL processing in Pig.
- b) Explain the features of Cassandra. What are the collections in GQLSH?
- c) What is Hive? Explain Hive Architecture in detail.

16



SLR-TC – 308

Set F

P

B.E. (C.S.E.) (New CGPA) (Part – II) Examination, 2018
Elective – IV
BIG DATA ANALYTICS

Day and Date : Tuesday, 22-5-2018
Time : 2.30 p.m. to 5.30 p.m.

Total Marks : 70

Instructions :

- 1) Figures to the right indicate full marks.
- 2) All questions are compulsory.
- 3) Q. No. 1 is compulsory. It should be solved in first 30 minutes in Answer Book Page No. 3. Each question carries one mark.
- 4) Answer MCQ/Objective type questions on Page No. 3 only. Don't forget to mention, Q.P. Set (P/Q/R/S) on Top of Page.

MCQ/Objective Type Questions

Duration : 30 Minutes

Marks : 14

1. Choose the correct answer : (1x14=14)

 - 1) CCTV footage is which type of data ?
 - a) Structured data
 - b) Unstructured data
 - c) Semi-structured data
 - d) All of these
 - 2) Big volume of data like 1 Yottabyte is equal to
 - a) 1024^4 bytes
 - b) 1024^6 bytes
 - c) 1024^8 bytes
 - d) 1024^9 bytes
 - 3) System will continue to function even when network partition occurs is
 - a) Partition tolerant
 - b) Consistency
 - c) Availability
 - d) None
 - 4) _____ is robust database that support ACID properties of transaction and has the scalability of NoSQL.
 - a) SQL
 - b) NewSQL
 - c) MySQL
 - d) NoSQL

P.T.O.

- 5) A typical block size is used by HDFS is
a) 64 MB b) 128 MB c) 32 MB d) 256 MB
- 6) _____ is the book-keeper of HDFS.
a) Name node b) Data node c) Job Tracker d) Task tracker
- 7) Name node uses _____ to record every transaction.
a) FslImage b) EditLog c) Data node d) Map Reduce
- 8) MongoDB has been adopted as _____ software by a number of major websites and services.
a) Frontend b) Backend
c) Proprietary d) All of the mentioned
- 9) Which one of the following is equivalent to ?
Select * from employee order by salary desc;
a) db.employee.find.sort({"salary":1})
b) db.employee.find.sort({"salary":-1})
c) db.employee.sort({"salary":1})
d) db.employee.sort({"salary":-1})
- 10) Hive is _____ tool.
a) Data Flow b) Data Warehouse
c) Import Export Tool d) Data Transfer
- 11) Cassandra is column oriented database designed to support _____ symmetric node architecture.
a) Peer to Peer b) Master Slave
c) Both a) and b) d) None
- 12) The 3 types of collection used in Cassandra are
a) Array, Set, List b) Set, List, Struct
c) Set, MAP, Array d) Set, List, MAP
- 13) MetaStore consist of _____ and a _____
a) Metaservices, database b) Metatable, WebUI
c) Metaservices, drivers d) CLI, Server
- 14) Pig is used in _____ process.
a) ETL b) Scripting c) Database d) None



B.E. (C.S.E.) (New CGPA) (Part – II) Examination, 2018
Elective – IV
BIG DATA ANALYTICS

Day and Date : Tuesday, 22-5-2018
Time : 2.30 p.m. to 5.30 p.m.

Marks : 56

Instructions : 1) Figures to the right indicate full marks.
2) All questions are compulsory.

SECTION – I

2. Answer any four : (4x4=16)
- Explain issue in handling unstructured data.
 - Describe big data stack.
 - Explain classification of big data analytics.
 - How traditional BI environment different from Big data environment ?
 - Explain role of Job Tracker and Task Tracker.

3. Answer any two : (6x2=12)
- Describe the anatomy of File write in HDFS.
 - What is HDFS ? Explain with help of block diagram.
 - Define Map Reduce. Give examples of word count using Map Reduce.

SECTION – II

4. Answer any four : (4x4=16)
- Write MongoDB query for an objective given below :

Create a collection “food” and insert 5 documents into the food collection by “_id” and “fruits” array. Check these documents are present in food collection. Find that document from food collection where “grapes” is present in 2nd index position of the fruit array.

- 2) What is Cassandra ? Explain features of Cassandra.
- 3) Explain replication and sharding in MongoDB.
- 4) Give different pig Latin statement with example.
- 5) Explain Map Reduce programming in MongoDB with example.
5. Answer **any two** :
- 1) Write different CRUD operation of MongoDB with example query.
- 2) Describe pig philosophy also explain ETL processing in pig.
- 3) Explain Hive architecture with suitable diagram.

(6x2=12)

**B.E. (C.S.E.) (Part – II) Examination, 2017
BIG DATA ANALYTICS (Elective – IV)**

Day and Date : Tuesday, 23-5-2017
Time : 3.00 p.m. to 6.00 p.m.

Max. Marks : 100

- Instructions :**
- 1) Figures to the right indicate full marks.
 - 2) All questions are compulsory.
 - 3) Q. No. 1 is compulsory. It should be solved in first 30 minutes in Answer book Page No. 3. Each question carries one mark.
 - 4) Answer MCQ/Objective type questions on Page No. 3 only.
Don't forget to mention, Q.P. Set (P/Q/R/S) on Top of Page.

MCQ/Objective Type Questions

Duration : 30 Minutes

Marks : 20

1. Choose the correct answer : **(20×1=20)**
- 1) Databases are under the category of _____ data.

a) Structured data	b) Semi-structured data
c) Unstructured data	d) None of above
 - 2) _____ is used to transmit data between a web server and a web application.

a) XML	b) JSON	c) Both a) and b)	d) None
--------	---------	-------------------	---------
 - 3) Cassandra based on _____ and _____ properties of CAP theorem.

a) Consistency and availability	b) Availability and partition tolerance
c) Consistency and partition tolerance	d) All of the above
 - 4) Big volume of data like 1 petabytes is equal to _____

a) 1024^4 Bytes	b) 1024^6 Bytes	c) 1024^5 Bytes	d) 1024^7 Bytes
-------------------	-------------------	-------------------	-------------------
 - 5) A coordinated processing of a program by multiple processors, each working on different part of the program and using its own operating system and memory is called _____

a) In-memory analysis	b) Distributed system
c) Massively parallel processing	d) None of above
 - 6) The 3V's terms of Big data was first introduced by _____

a) Doug Laney	b) Brewer	c) Carlo Strozzi	d) Doug cutting
---------------	-----------	------------------	-----------------
 - 7) Cap theorem is also called as _____ theorem.

a) Doug Laney	b) Brewer	c) Carlo Strozzi	d) Doug cutting
---------------	-----------	------------------	-----------------

P.T.O.

- 8) _____ is a robust database that supports ACID properties of transactions and has the scalability of NoSQL.
 a) NoSQL b) NewSQL c) MySQL d) SQL
- 9) Hadoop supports _____ data formats.
 a) Structured b) Semi-structured
 c) Unstructured d) All above
- 10) Mongo DB is _____
 a) RDBMS b) Document-oriented DBMS
 c) Object oriented DBMS d) Key-value store
- 11) Which command in Mongo DB is equivalent to SQL select ?
 a) Find() b) Search() c) Document() d) None of above
- 12) What does the following command do ?
`db.sample.find().limit(10)`
 a) Show 10 documents randomly from the collection sample
 b) Show only first 10 documents from the collection sample
 c) Repeats the first document 10 times
 d) None of above
- 13) Cassandra is a column-oriented database designed to support _____ symmetric node architecture.
 a) Peer-to-peer b) Master-slave c) Both a) and b) d) None of above
- 14) The 3 types of collections used in Cassandra are
 a) Array, List, Map b) List, Map, Struct
 c) List, Set, Map d) Set, Map, Array
- 15) Hive is a _____ tool.
 a) Data flow b) Data transfer
 c) Column-oriented d) Data warehousing
- 16) The metastore in Hive consists of _____ and _____
 a) Driver, services b) Metaservices, database
 c) Driver, database d) Metaservices, driver
- 17) Pig is _____
 a) Data Flow language b) NoSQL database
 c) Import Export tool d) Scheduling engine
- 18) In Cassandra, _____ is called peer-to-peer communication protocol used for intra-ring communication.
 a) Anti-Entropy b) Hinted Handoffs
 c) Gossip Protocol d) None of above
- 19) ETL processing in Pig stands for
 a) Extract, Transform and Load b) Extend, Transfer and Load
 c) Extract, Transform and Local d) None of above
- 20) Core MongoDB operations are
 a) Create, Select, Update, Delete b) Create, Read, Update, Delete
 c) Create, Read, Update, Drop d) Create, Remove, Update, Drop

B.E. (C.S.E.) (Part – II) Examination, 2017
BIG DATA ANALYTICS (Elective – IV)

Day and Date : Tuesday, 23-5-2017
Time : 3.00 p.m. to 6.00 p.m.

Marks : 80

Instructions: 1) Figures to the right indicate full marks.
2) All questions are compulsory.

SECTION – I

2. Attempt any four (each carries 5 marks) : (4×5=20)

- a) What is Massively Parallel processing terminology used in Big Data ?
- b) Explain Sources of Unstructured Data.
- c) What is Big data ? Explain different characteristics of Big data.
- d) Explain features of Data Node.
- e) Explain Hadoop 1.0 and Hadoop 2.0.

3. Attempt any two (each carries 10 marks) : (2×10=20)

- a) What is Big Data analytics ? Explain classification of Big Data Analytics.
- b) Describe the anatomy of file write in HDFS.
- c) How MapReduce works ? Explain with example.

SECTION – II

4. Attempt any four (each carries 5 marks) : (4×5=20)

- a) Write difference between SQL and MongoDB.
- b) Explain CRUD operations in Cassandra.
- c) Explain Map Reduce programming in MongoDB with suitable example.
- d) Write Word Count example using Pig.
- e) Explain the process of replication and Sharding in MongoDB.

5. Attempt any two (each carries 10 marks) : (2×10=20)

- a) Explain Hive architecture with diagram. Also explain Hive file format.
- b) What is Cassandra ? Explain features of Cassandra.
- c) Describe Pig philosophy also explain ETL processing in Pig.

B.E. (C.S.E.) (Part - II) (New) Examination, 2016
Elective - IV : BIG DATA AND ANALYTICS

Set R

Day and Date : Thursday, 12-5-2016
Time : 3.00 p.m. to 6.00 p.m.

Total Marks : 100

Instructions : 1) Q. No. 1 is compulsory. It should be solved in first 30 minutes in Answer Book Page No. 3. Each question carries one mark.
2) Answer MCQ/Objective type questions on Page No. 3 only.
Don't forget to mention, Q.P. Set (P/Q/R/S) on Top of Page.

MCQ/Objective Type Questions

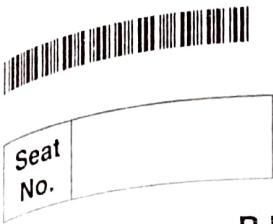
Duration : 30 Minutes

Marks : 20

(20x1=20)

1. Choose the correct answer :
 - 1) Which command in MongoDB is equivalent to SQL select ?
a) find() b) search() c) document() d) none of above
 - 2) What does the following command do ?
`db.sample.find().limit(10)`
a) Show 10 documents randomly from the collection sample
b) Show only first 10 documents from the collection sample
c) Repeats the first document 10 times
d) None of above
 - 3) Cassandra is a column-oriented database designed to supports _____
symmetric node architecture.
a) Peer-to-peer b) Master-slave
c) Both a and b d) None of above
 - 4) The 3 types of collections used in cassandra are
a) Array, List, Map b) List, Map, Struct
c) List, Set, Map d) Set, Map, Array
 - 5) Hive is a _____ tool.
a) Data flow b) Data transfer
c) Column-oriented d) Data warehousing
 - 6) The metastore in Hive consist of _____ and _____.
a) driver, services b) metaservices, database
c) driver, database d) metaservices, driver
 - 7) Pig is
a) Data flow language b) NoSQL database
c) Import export tool d) Scheduling engine

- 8) In Cassandra, _____ is called peer-to-peer communication protocol used for intra-ring communication.
 a) Anti-entropy
 b) Hinted Handoffs
 c) Gossip protocol
 d) None of above
- 9) ETL processing in Pig stands for
 a) Extract, Transform and Load
 b) Extend, Transfer and Load
 c) Extract, Transform and Local
 d) None of above
- 10) Core MongoDB operations are
 a) Create, Select, Update, Delete
 b) Create, Read, Update, Delete
 c) Create, Read, Update, Drop
 d) Create, Remove, Update, Drop
- 11) E-mails are under the category of _____ data.
 a) Structured data
 b) Semi-structured data
 c) Unstructured data
 d) None of above
- 12) _____ is used to transmit data between a web server and a web application.
 a) XML
 b) JSON
 c) Both a and b
 d) None
- 13) MongoDB based on _____ and _____ properties of CAP theorem.
 a) consistency and availability
 b) availability and partition tolerance
 c) consistency and partition tolerance
 d) all of the above
- 14) Big volume of data like 1 Zettabytes is equal to
 a) 1024^4 Bytes
 b) 1024^6 Bytes
 c) 1024^5 Bytes
 d) 1024^7 Bytes
- 15) A coordinated processing of a program by multiple processors, each working on different part of the program and using its own operating system and memory is called
 a) In-memory analysis
 b) Distributed system
 c) Massively parallel processing
 d) None of above
- 16) The 3V's terms of big data was first introduced by
 a) Doug Laney
 b) Brewer
 c) Carlo Strozzi
 d) Doug cutting
- 17) Cap theorem is also called as _____ theorem.
 a) Doug Laney
 b) Brewer
 c) Carlo Strozzi
 d) Doug cutting
- 18) _____ is a robust database that supports ACID properties of transactions and has the scalability of NoSQL.
 a) NoSQL
 b) NewSQL
 c) MySQL
 d) SQL
- 19) Hadoop supports _____ data formats.
 a) structured
 b) semi-structured
 c) unstructured
 d) all above
- 20) MongoDB is
 a) RDBMS
 b) Document-oriented DBMS
 c) Object Oriented DBMS
 d) Key-value store



B.E. (C.S.E.) (Part – II) (New) Examination, 2016
Elective – IV : BIG DATA AND ANALYTICS

Day and Date : Thursday, 12-5-2016

Marks : 80

Time : 3.00 p.m. to 6.00 p.m.

SECTION – I

2. Attempt **any four** (each carries 5 marks).

20

- a) How traditional BI environment different from the Big Data environment ?
- b) Explain sources of structural data.
- c) What is data science ? Explain responsibilities of data scientist.
- d) Explain features of NameNode.
- e) How MapReduce works ? Explain with example.

3. Attempt **any two** (each carries 10 marks).

20

- a) What is Big Data Analytics ? Explain classification of Big Data Analytics.
- b) Describe the anatomy of File Write in HDFS.
- c) Describe YARN architecture in detail.

SECTION – II

4. Attempt **any four** (each carries 5 marks).

20

- a) Write difference between SQL and MongoDB.
- b) Explain CRUD operations in Cassandra.
- c) Describe Pig philosophy also explain ETL processing in Pig.
- d) Write Word Count example using Pig.
- e) Explain the process of Replication and Sharding in MongoDB.

Set R



5. Attempt **any two** (each carries **10** marks).

20

- a) Explain Hive architecture with diagram. Also explain Hive file format.
 - b) What is Cassandra ? Explain features of Cassandra.
 - c) Explain Map Reduce programming in MongoDB with suitable example.
-