Total No. of Questions: 22

Enrollment No : _____

MEDICAPS UNIVERSITY

Faculty of Management Studies

End Semester Examination May 2025

MS5EB04 Big Data Analytics & Hadoop

Programme	:	MBA	**Branch/Specialisation**	:	-
Duration	:	3 hours	**Maximum Marks**	:	60

Note: *All questions are compulsory. Internal choices, if any, are indicated. Assume suitable data if necessary. Notations and symbols have their usual meaning.*

Section 1 (Answer all question(s))

Marks CO BL

Q1. Which of the following is not considered one of the 4Vs of big data? **1 1 1**

- ○ Volume
- ○ Variety
- ○ Value
- ◉ Visibility

Q2. Which component of hadoop is responsible for resource management? **1 1 1**

- ○ HDFS
- ○ Hive
- ◉ YARN
- ○ Pig

Q3. Which of the following commands is used to list files in an HDFS directory? **1 2 1**

- ○ hadoop fs -list
- ○ hdfs ls
- ◉ hdfs dfs -ls
- ○ hadoop dfs -read

Q4. Which HDFS feature helps in processing tasks close to data location? **1 2 1**

- ○ Data sharding
- ◉ Rack awareness
- ○ Data mirroring
- ○ Data swapping

Q5. Which is not a transformation operation in spark? **1 3 1**

- ○ map()
- ○ filter()
- ◉ collect()
- ○ flatMap()

Q6. Which spark component is used for working with structured data? **1 3 1**

- ○ Spark streaming
- ○ MLlib
- ◉ Spark SQL
- ○ GraphX

Q7. Which phase in data architecture focuses on ingesting data from sources? **1 4 1**

- ◉ Collect
- ○ Store
- ○ Process
- ○ Consume

Q8. Which of the following is not a typical tool used for building data architecture? **1 4 1**

- ○ Apache NiFi
- ○ Apache cassandra
- ◉ Adobe Photoshop
- ○ Apache Kafka

Q9. What is meant by "Streaming Data"? **1 5 1**

- ○ Data that is stored in the cloud
- ◉ Data that is continuously generated and processed
- ○ Data stored in relational databases
- ○ Predefined static data sets

Q10. What distinguishes unstructured data? **1 5 1**

- ○ It is stored in a relational database
- ○ It has a defined schema
- ◉ It includes text, images, and videos
- ○ It is always numerical

Section 2 (Answer all question(s))

Marks CO BL

Q11. Define big data. 2 1 1

Rubric	Marks
Definition of Big Data.	2

Q12. Differentiate between structured and unstructured data in the context of big data. 3 1 2

Rubric	Marks
Each 1 Difference one mark(At least 3) .	3

Q13. (a) Explain the various aspects of big data along with real-world examples. 5 1 2

Rubric	Marks
Explain the various aspects	3
Big Data along with real-world examples	2

(OR)

(b) Explain the working of MapReduce with a suitable example.

Rubric	Marks
Explain the working of MapReduce	3
suitable example.	2

Section 3 (Answer all question(s))

Marks CO BL

Q14. What is the purpose of rack awareness in Hadoop? 2 2 1

Rubric	Marks
purpose of Rack Awareness in Hadoop?	2

Q15. (a) Explain the HDFS architecture with file storage, block concept, and replication in detail. 8 2 2

Rubric	Marks
Explain the **HDFS architecture** with file storage	4
block concept, and replication in detail	4

(OR)

(b) What is the role of combiner and partitioner in MapReduce? Explain with a sample MapReduce flow.

Rubric	Marks
What is the role of **Combiner and Partitioner** in MapReduce	4
Explain with a sample MapReduce flow	4

Section 4 (Answer all question(s))

Marks CO BL

Q16. Differentiate between Spark and MapReduce. 3 3 1

Rubric	Marks
Each differences 1 mark(At least 3)	3

Q17. (a) Explain Spark architecture with a neat diagram. Describe the role of driver and executors. **7 3 2**

Rubric	Marks
Explain Spark Architecture with a neat diagram	4
Describe the role of driver and executors.	3

(OR)

(b) Compare and contrast RDDs, dataframes and datasets in Apache Spark.

Rubric	Marks
Compare and contrast RDDs	4
DataFrames, and Datasets in Apache Spark	3

Section 5 (Answer all question(s))

Marks CO BL

Q18. Define data architecture. Explain its significance in AI/ML initiatives. **4 4 2**

Rubric	Marks
Define Data Architecture	2
its significance in AI/ML initiatives	2

Q19. (a) Discuss the principles of good data architecture and their importance in scalable system design. **6 4 2**

Rubric	Marks
Discuss the principles of good Data Architecture .	3
their importance in scalable system design.	3

(OR)

(b) What is data context? How does it enhance data understanding in analytics systems?

Rubric	Marks
What is Data Context?	3
How does it enhance data understanding in analytics systems.	3

Section 6 (Answer any 2 question(s))

Marks CO BL

Q20. Explain the importance of a well-defined data strategy in the success of an AI or ML initiative. **5 5 1**

Rubric	Marks
Explain the importance of a well-defined data strategy	3
the success of an AI or ML initiative	2

Q21. Differentiate between various types of AI initiatives with suitable real-world examples. **5 5 2**

Rubric	Marks
Differentiate between various types of AI initiatives	3
uitable real-world examples	2

Q22. What is unstructured data? How is it used in AI/ML applications? Give two examples. **5 5 2**

Rubric	Marks
What is unstructured data?	3
How is it used in AI/ML applications? Give two examples?	2

******