

Reg. No.

--	--	--	--	--	--	--	--	--	--	--	--

M.E. / M.TECH. DEGREE EXAMINATIONS, MAY 2019

Second Semester

CP18204 – BIG DATA ANALYTICS

(Computer Science and Engineering)
(Regulation 2018)

Time: Three Hours**Maximum : 100 Marks**

Answer ALL questions

PART A - (10 X 2 = 20 Marks)

1. Specify the dimension of big data.
2. Mention the real time applications of big data.
3. Describe coherency model for HDFS.
4. What do you mean by shuffle and sort?
5. Specify the advantages of K-means clustering algorithm.
6. State the characteristics of grid based methods.
7. Define data streaming.
8. What do you mean by moment?
9. List the drawbacks of Hive.
10. How does Pig differ from MapReduce?

PART B - (5 X16 = 80 Marks)

11. (a) (i) In what ways is big data different from traditional data? Explain in detail. (8)

(ii) Describe the structure of Web data. (8)

(OR)

- (b) (i) Explain attention modeling and response modeling in detail. (8)

(ii) Explain the working principle of page ranking in web. (8)

12. (a) (i) Explain about the anatomy of a MapReduce job. (8)

- (ii) Explain the working of Word Count example using MapReduce for the given input (8)

**The quick brown fox
the fox ate the mouse
how now
brown cow**

(OR)

- (b) (i) Explain the basic FileSystem operations in Hadoop. (8)

- (ii) Explain YARN architecture in detail. (8)

13. (a) Explain the maximum margin classifier for the non-linearly separable case. (16)

(OR)

- (b) Explain BIRCH algorithm in detail. (16)

14. (a) Explain Bloom filtering algorithm in detail. (16)

(OR)

- (b) Explain the Flajolet-Martin algorithm in detail. (16)

15. (a) (i) Explain about HBase Architecture. (8)

- (ii) Explain the working of Word Count example using PigLatin script for any input. (8)

(OR)

- (b) (i) Explain about Cassandra data model. (8)

- (ii) Explain the different types of NoSQL systems with examples. (8)