

[07 BENG – 4104]

IV/IV B.Tech. DEGREE EXAMINATION.

First Semester

Computer Science Engineering

BIG DATA ANALYTICS

(Common with Information Technology)

(Effective from the admitted batch of 2015–2016)

Time : Three hours

Maximum : 70 marks

First question is compulsory.

Answer any FOUR from the remaining questions.

All questions carry equal marks.

Answer all parts of any question at one place.

1. Answer the following in brief :

- (a) What is MapReduce?**
- (b) Features of HDFS.**
- (c) Secondary Name Node.**
- (d) Hadoop pipes.**
- (e) MapReduce workflows.**
- (f) Hadoop Archives.**
- (g) Bloom filters.**

~~Q.~~ 2. (a) What is Big data? Explain the characteristics of Big Data.

(b) Define the terms, structured data, semi structured data and unstructured data w.r. to Big data.

~~Q.~~ 3. (a) Explain the components of Hadoop with a neat diagram.

(b) Describe various phases of MapReduce.

~~Q.~~ 4. (a) Define HDFS and explain HDFS architecture with neat diagram.

(b) What are the functions of Job Tracker and Task Tracker? Discuss.

5. (a) Write a MapReduce code for word count.

(b) How do you run MapReduce and Message Passing Interface (MPI) on YARN architecture?

6. (a) What are the problems related to MapReduce? Discuss.

(b) Explain Page Rank Algorithm in detail.

7. (a) Explain the process of creating parallelized bloom filters in MapReduce.
- (b) Discuss friends-of-friends algorithm.

~~8.~~ Write short notes on the following.

- ~~(a)~~ Chaining MapReduce jobs.
- ~~(b)~~ Shortest path algorithm.
- ~~(c)~~ Hadoop file system.
-