

University Institute of Engineering & Technology

(Recognised Under Section 2(f) and 12B of UGC)

Kurukshetra University, Kurukshetra

TIME – 3 Hrs 15Min

THEORY EXAMINATION – JAN 2021

B.TECH – CSE

SEMESTER –VII

M.M. - 56

PAPER - CSE - 423

SUBJECT- BIG DATA AND ANALYTICS

INSTRUCTIONS TO BE FOLLOWED

- Allotted time for examination is 3 hours 15 minutes that includes time for downloading the question paper, writing answers, scanning of answer sheets and E-mailing the PDF files to the designated Email ID.
- For CSE-A Regular Students, the Email ID is:- btech7thcsea@kuk.ac.in
- For CSE-B Regular and All Reappear Students, the Email ID is:- btech7thcseb@kuk.ac.in
- The candidates will be required to attempt 75% of the question paper (maximum) by choosing to their any best questions accumulating 56 marks.
- The PDF files should be saved as Roll No. and Subject Code. Proper attention should be given while sending the email and in the subject line, the Roll Number and Subject Code should be mentioned.
- Maximum Page Limit should be 20 (Twenty) for attempting the question paper on A4 sheets which could be downloaded and printed from the sample sheets given in the Kurukshetra University Examination guidelines.
- Over-attemptation should be avoided.
- Handwriting should be neat and clean and diagrams should be clear and contrasted.
- The candidate should not write their Mobile No. otherwise Unfair Means Case will be made.
- While attempting the paper, the candidate will use blue/black pen only.
- Before attempting the paper, the candidate will ensure that he/she has downloaded the correct question paper. No complaint for attempting wrong question paper by the candidate will be entertained.
- Candidate must ensure that he/she has put his/her signature on each page of the answer sheet used by him/her. Answer sheet without the signature of the candidate will not be evaluated.

PART-A

Q. No. – 1 Answer the following questions.

15x1=15

(i)	List the 4 V's of Big Data.
(ii)	Data generated as web logs is classified as data. (structured/semi-structured/unstructured)
(iii) deal with the growing number of data sources. (Web designer/Data Scientist/Business developer)
(iv)	What is data deduplication ?
(v)	What are the major technologies used for processing big data ?
(vi)	What are the schema less databases ?
(vii)	What is shuffling and sorting during task execution in Hadoop ?
(viii)	What is the relationship between IoT and big data ?
(ix)	List the differences between parallel and distributed computing.
(x)	Pig was developed by (Yahoo/Twitter/Facebook/Gmail)
(xi)	What is Content-Defined Chunking ?
(xii)	What is delta compression ?
(xiii)	Message Digest MD5 consists of (64-bits/128-bits/160-bits/256-bits)
(xiv)	Secure Hash Algorithm SHA-1 consists of (64-bits/128-bits/160-bits/256-bits)
(xv)	What is the significant role of YARN architecture ?

PART-B

2	How are decision-making processes carried out with the help of Big Data Analytics ?	5
3	What is MongoDB ?	5
4	What is Rabin Fingerprinting technique ?	5
5	What is HIVE ? What are the HiveQL queries ?	5

PART-C

6	Discuss the role cloud services play in handling Big Data.	10
7	Describe the features of cloud computing that can be used to handle Big Data.	10
8	Explain about Frequency Based Chunking Algorithm.	10
9	What is HBase ? Discuss the concept of regions in HBase.	10
10	Explain in detail about Tiger Hashing Technique.	10
11	Discuss in detail the Gear Hashing approach.	10
12	Discuss the Local Mode and Map Reduce Mode to run the Pig Scripts. What are the main reasons for developing Pig Latin ?	10
13	What is Cassandra data model ? How is Cassandra technology used in big data analytics ?	10