INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous) Dundigal, Hyderabad - 500 043

COMPUTER SCIENCE AND ENGINEERING

TUTORIAL QUESTION BANK

Course Name	Big data analytics
Course Code	BCS212
Class	I M. Tech
Branch	Common for CSE / SE
Year	2016 - 2017
Team of Instructors	

OBJECTIVES

To meet the challenge of ensuring excellence in engineering education, the issue of quality needs to be addressed, debated and taken forward in a systematic manner. Accreditation is the principal means of quality assurance in higher education. The major emphasis of accreditation process is to measure the outcomes of the program that is being accredited.

In line with this, faculty of Institute of Aeronautical Engineering, Hyderabad has taken a lead in incorporating philosophy of outcome based education in the process of problem solving and career development. So, all students of the institute should understand the depth and approach of course to be taught through this question bank, which will enhance learner's learning process.

PART – A (SHORT ANSWER QUESTIONS)

S. No	Question	Blooms Taxonomy Level	Course Outcome
	UNIT – III		
1.	When is the reducers are started in a MapReduce job?	Understand	3
2.	What is the difference between HDFS and NAS ?	Understand	3
3.	What is HDFS ? How it is different from traditional file systems?	Understand	3
4.	What is Writable & WritableComparable interface?	Understand	3
5.	What is a Task Tracker in Hadoop? How many instances of TaskTracker run on a Hadoop Cluster	Understand	3
6.	Can I set the number of reducers to zero?	Understand	3
7. 17 	Where is the Mapper Output (intermediate kay-value data) stored ?	Understand	3

8.	What are combiners? When should a combiner in my MapReduce Job be	Remember	3
	used?		
9.	How the Client communicates with HDFS?	Remember	3
10.	How the HDFS Blocks are replicated?	Understand	3
11.	What is HDFS?	Understand	3
12.	What is MAP REDUCE? What Mapper does?		
13.	What is the InputSplit in map reduce software?	Understand	3
14.	What is the default replication factor in HDFS?	Understand	3
15.	How does Mappers run() method works?	Understand	3

PART – B (LONG ANSWER QUESTIONS)

S. No	Question	Blooms Taxonomy Level	Course Outcome
	UNIT – III		
1.	Does MapReduce programming model provide a way for reducers to communicate with each other? In a MapReduce job can a reducer communicate with another reducer?	Understand	3
2.	What is Writable & WritableComparable interface?	Apply	3
3.	What is the Hadoop MapReduce API contract for a key and value Class?	Understand	3
4.	What is a IdentityMapper and IdentityReducer in MapReduce ?	Apply	3
5.	Which interface needs to be implemented to create Mapper and Reducer for the Hadoop?	Apply	3
6.	What is the Reducer used for?	Remember	
7.	What are the primary phases of the Reducer?	Apply	
8.	Explain the Reducer's reduce phase?How many Reducers should be configured?		
9.	Explain the Reducer?s Sort phase?		
10.	Explain the core methods of the Reducer?		
11.	What is next step after Mapper or MapTask?		
12.	How can you add the arbitrary key-value pairs in your mapper?		
13.	What happens if you don?t override the Mapper methods and keep them as it is?		
18	Page	I	

14.	What alternate way does HDFS provides to recover data in case a Namenode, without backup, fails and cannot be recovered?	
15.	Explain the shuffle?	

PART – C (PROBLEM SOLVING AND CRITICAL THINKING QUESTIONS)

S. No	Question	Blooms Taxonomy Level	Course Outcome
	UNIT – III		
1.	If reducers do not start before all mappers finish then why does the progress on MapReduce job shows something like Map(50%) Reduce(10%)? Why reducers progress percentage is displayed when mapper is not finished yet?	Understand	3
2.	What are the main components of HBase?	Apply	3
3.	What should be an upper limit for counters of a Map Reduce job?	Understand	3
4.	Which class is responsible for converting inputs to key-value Pairs of Map Reduce?	Apply	3
5.	Which writables can be used to know value from a mapper/reducer?	Apply	3
6.	Which type of joins can be performed in Reduce side join operation?	Remember	
7.	What are the reasons for the usage of Hbase?	Apply	
8.	What mechanism is used to create replica in HDFS?		
9.	Where is HDFS replication factor controlled?		
10.	Write the correct flow of sequence of MapReduce.		
11.	What things are to be used to control the number of part files in a map reduce program output directory?		
12.	Which operations can't use Reducer as combiner also?		
13.	For which type of data sets, reduce side join is useful?		
14.	In which type of phase (mapper / reducer) distributed cache can be used?		
15.	MapReduce was devised by which company?		

PART – A (SHORT ANSWER QUESTIONS)

S. No	Question	Blooms Taxonomy Level	Course Outcome
	UNIT – IV		
1.	How does Hadoop process large volumes of data?	Understand	3
2.	When using HDFS, what occurs when a file is deleted from the command line?	Understand	3
3.	Which components retrieves the input splits directly from HDFS to determine the number of map tasks?	Understand	3
4.	What mechanisms Hadoop uses to make namenode resilient to failure.	Understand	3
5.	What is the property used to set the default filesystem for Hadoop in core- site.xml ?	Understand	3
6.	Which inter process communication, Hadoop uses between different nodes ?	Understand	3
7.	What does commodity Hardware in Hadoop world mean?	Understand	3
8.	How many instances of Job tracker can run on Hadoop cluster ?	Remember	3
9.	What is the difference between namenode and datanode in Hadoop?	Remember	3
		Understand	3
10	What is the difference between a Hadoop database and Relational Database?	Understand	3
11	What is Hadoop framework?		
12	On What concept the Hadoop framework works?	Understand	3
13	How many Daemon processes run on a Hadoop system?	Understand	3
14	What is configuration of a typical slave node on Hadoop cluster? How many JVMs run on a slave node?	Understand	3
15	What is the difference between HDFS and NAS ?		

PART – B (LONG ANSWER QUESTIONS)

S. No	Question	Blooms Taxonomy Level	Course Outcome
	UNIT – IV		
1.	What is the meaning of speculative execution in Hadoop? Why is it important?	Understand	3
2.	What is a NameNode? How many instances of NameNode run on a Hadoop Cluster?	Apply	3
20	Page		•

3.	What is a DataNode? How many instances of DataNode run on a Hadoop Cluster?	
4.	Discuss in detail how to read and write data in HADOOP Using Java Interface.	
5.	Discuss the design of Hadoop distributed file system and concept in detail.	
6.	Discuss YARN in detail with failures in classic Map-reduce.	
7.	Explain the HDFS Architecture and list the various HDFS daemons in HDFS cluster?	
8.	What is the difference between NAS (Network Attached Storage) and HDFS?	
9.	What is the difference between traditional RDBMS and Hadoop?	
10.	What do you mean by meta data in HDFS? List the files associated with metadata.	
11.	How to copy a file into HDFS with a different block size to that of existing block size configuration?	
12.	Define Hadoop Archives? What is the command for archiving a group of files in HDFS.	
13.	What do you mean by the High Availability of a NameNode? How is it achieved?	
14.	Define Data Integrity? How does HDFS ensure data integrity of data blocks stored in HDFS?	

PART – C (PROBLEM SOLVING AND CRITICAL THINKING QUESTIONS)

S. No	Question	Blooms Taxonomy Level	Course Outcome
	UNIT – IV		
1.	Suppose there is file of size 514 MB stored in HDFS (Hadoop 2.x) using default block size configuration and default replication factor. Then, how many blocks will be created in total and what will be the size of each block?	Understand	3
2.	HDFS stores data using commodity hardware which has higher chances of failures. So, How HDFS ensures the Fault Tolerance capability of the system?	Apply	3
3.	Replication causes data redundancy and consume a lot of space, then why is it pursued in HDFS?	Understand	3
4.	What is a rack awareness algorithm and why is it used in Hadoop?	Apply	3
5.	How data or a file is written into HDFS?	Apply	3

6.	Does HDFS allow a client to read a file which is already opened for writing?	Remember	
7.	How to perform the inter cluster data copying work in HDFS?	Apply	

PART – A (SHORT ANSWER QUESTIONS)

S. No	Question	Blooms Taxonomy Level	Course Outcome
	UNIT – V		
1	What is opinion mining?		
2	What are the sub components of opinion mining?		
3	What are the approaches for finding opinion?		
4	What are sentiment finding components?		
5	List the challenges imposed by social media.	Understand	3
6	What is sentiment analysis?	Understand	3

PART – B (LONG ANSWER QUESTIONS)

S. No	Question	Blooms Taxonomy Level	Course Outcome	
UNIT – V				
1.	Explain the anatomy of subjectivity analysis.	Understand	3	
2.	Define opinion.	Apply	3	
3.	List the components of sentiment analysis.			
4.	What are the most challenging issues in Sentiment Analysis?			
5.	What are the applications of opinion mining?			

PART – C (PROBLEM SOLVING AND CRITICAL THINKING QUESTIONS)

S. No	Question	Blooms Taxonomy Level	Course Outcome	
UNIT – V				
1.	Why do opinion analysis?	Understand	3	
2.	What are the levels of opinion analysis?	Apply	3	
3.	What tool Should be used for Mobile App Analytics?	Understand	3	
4.	How to implement Mobile SDKs?	Apply	3	
5.	What is difference between mobile analytics and web analytics?	Apply	3	