APJ ABDULKALAM TECHNOLOGICAL UNIVERSITY
08 PALAKKAD CLUSTER

Q. P. Code : 08CSE19-6042B-1 **(Pages: 3)** **Name: .....**

Reg. No:.....

SECOND SEMESTER M.TECH. DEGREE EXAMINATION MAY 2019

Branch: Computer Science and Engineering Specialization: Computer Science and Engineering

08CS6042(B): BIG DATA ESSENTIALS

(Common to CSE)

Time:3 hours **Max. marks: 60**

Answer all six questions.

Modules 1 to 6: Part 'a' of each question is compulsory and answer either part 'b' or part 'c' of each question.

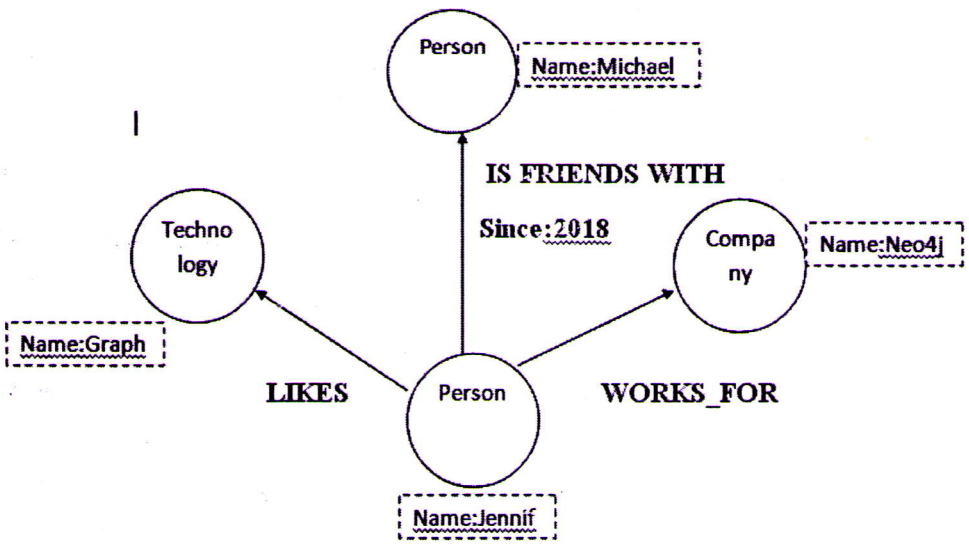
Q.no.		Module 1	Marks
1.a		List any two real world big data sources where the "velocity" characteristic dominates.	3
		Answer b or c	
b		Write about three big data use cases in detail.	6
c	i.	Identify and list key roles in a data analytics project.	2
	ii.	List different steps in a data analytics life cycle in sequence.	2
	iii.	Define analytical sandbox.	2

Q.no.		Module 2	Marks
2.a		How Hadoop makes the Namenode resilient to failures?	3
		Answer b or c	
b		Describe MapReduce Framework in detail. What are the limitations of MapReduce framework and how it is solved in YARN?	6
c		Justify the need of MultipleInputs and MultipleOutputs formats provided by Hadoop with proper examples.	6

Q.no.		Module 3	Marks
3.a		Differentiate Partitioners and Combiners used in a MapReduce program.	3
		Answer b or c	

b	List the general sequence of steps to be followed in development of a MapReduce Application	**6**
c	Suppose you have a large tabular data, which stores your internal marks for each subjects in the current semester. Structure of the tab separated data is given below: <SubjectCode RollNo Name Marks> Design a MapReduce algorithm to find total marks of each student. Also draw MapReduce data flow diagram for the same.	**6**

Q.no.	**Module 4**	**Marks**
4.a	How data gets organized in HBase table?	**3**
	Answer b or c	
b	Distinguish different types of NoSQL datastores with respect to its characteristics and use cases along with an example for each type.	**6**
c	Write about Neo4j and write sequence of statements to create the following graph using Neo4j.	**6**



Q.no.	**Module 5**	**Marks**
5.a	List various limitation of MapReduce Framework.	**4**
	Answer b or c	
b	State and explain HACE theorem in detail.	**8**
c	Write any one big data clustering algorithm with an example.	**8**

Q.no.		Module 6	Marks
6.a		Draw and explain Real Time Big Data Analytics Stack.	4
		Answer b or c	
b	i.	How in memory models works faster than MapReduce? Explain with Apache Spark as example.	4
	ii.	List any two Transformations and Actions that can be performed on a Spark RDD (Resilient Distributed Dataset).	4
c	i.	List various components and working principle of a Bulk Synchronous Parallel model computer.	4
	ii.	Write short notes on Apache Giraph.	4