Code No: **RT4105B**   **R13** | **Set No. 1** |

IV B.Tech I Semester Supplementary Examinations, March - 2017
HADOOP AND BIG DATA
(Common to Computer Science & Engineering and Information Technology)

Time: 3 hours   **Max. Marks: 70**

Question paper consists of Part-A and Part-B
Answer ALL sub questions from Part-A
Answer any THREE questions from Part-B
*****

PART−A *(22 Marks)*

1. a) Write Difference between Wildcard(?) argument and Normal Type argument in Generic Programming in Java   [4]
    b) Write the reasons for Why Hadoop won't be using JAVA serialization.   [4]
    c) What is structured, semi-structured, unstructured data? Write examples?   [4]
    d) Why Key Type need be both Writable and Comparable in Map-Reduce Programs.   [3]
    e) Write a PIG script for word count.   [4]
    f) Why HIVE is relevant in Hadoop Eco system.   [3]

PART−B *(3x16 = 48 Marks)*

2. a) Explain the difference between Array List and Linked List class functionalities with examples.   [8]
    b) Write Java Program to Implement Generic Single-linked list?   [8]

3.   Write a Map-Reduce Algorithm to get the Dot Product of two Large Vectors and Implement it using Java , Assuming Only non-zero elements of those vectors are given in input files and output file should show only non-zero entries(assuming two vectors are same size)

    ex: v1=[5 4 0 1 2] v2=[4 2 1 0 6]

file1:	file2:	output:
(0,5)	(0,4)	(0,20)
(1,4)	(1,2)	(1,8)
(3,1)	(2,1)	(4,12)
(4,2)		(4,6)

[16]

Code No: **RT4105B**   # R13   Set No. 1

4. a) What are the different modes in which Hadoop can be installed and what is the use of each mode from application and developer point of view? [6]

   b) Explain the uses of Name node, Data node and Secondary Name node in Hadoop Distributed File system. [6]

   c) What is replication factor in HDFS and what is the default value. [4]

5. a) In Map-Reduce Implementation of Word count, Write Types of keys and Values mapper, reducer, combiner(if any) and which of the types need to Writable. [8]

   b) Differentiate "Scale up and Scale out"
   Explain with an example How Hadoop uses Scale out feature to improve the Performance. [8]

6. a) Write the general PIG Latin program/flow organization. [6]

   b) Consider The student data File (st.txt), Data in the following format

   Name, District, age, gender

   i) Write a PIG script to Display Names of all female students

   ii) Write a PIG script to find the number of Students form Prakasham District

   iii) Write a PIG script to Display District wise count of all male students. [10]

7. a) Explain in detail how hive is different form PIG [6]

   b) Write Example Hive Queries for Natural Join and outer-Join [10]