VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

SCHOOL OF INFORMATION TECHNOLOGY AND ENGINEERING

Mid-Term Examination Summer Semester MCA - 2022-2023

Course Name: BigData Analytics **Duration: 1hr 30Mins.**

Course Code: ITA6008 **Max. Marks : 50 marks**

Slot : C1+TC1+C2+TC2 Faculty : Prof. Pounambal M, Prof. Balasubramani and Prof. Rajiv

---

Q.No

Answer ALL Questions 5 * 10 = 50 Marks

1 Specify the Industry vertical under which the following are categorized. Write one challenge and application for each.

a. Spotify [0.5+1+1]

b. Carnegie Learning [0.5+1+1]

c. City of Dublin [0.5+1+1]

d. Smart Grid [0.5+1+1]

2 Derive the data analytics life cycle for recommending an item in AMAZON website.

3 a. A file named moviedb.txt of size 500KB is stored in HDFS. A client wants to read the file. Explain the components involved in the above scenario. [5]

b. Write the features of the following type of data, classify the type of data also guide the user to store it in HADOOP environment. [2.5+2.5]

i)

```
<CATALOG>
<PLANT>
<COMMON>Bloodroot</COMMON>
<BOTANICAL>Sanguinaria canadensis</BOTANICAL>
<ZONE>4</ZONE>
<LIGHT>Mostly Shady</LIGHT>
<PRICE>$2.44</PRICE>
<AVAILABILITY>031599</AVAILABILITY>
</PLANT>
<PLANT>

<COMMON>Jack-In-The-Pulpit</COMMON>
<BOTANICAL>Arisaema triphyllum</BOTANICAL>
<ZONE>4</ZONE>
<LIGHT>Mostly Shady</LIGHT>
<PRICE>$3.23</PRICE>
<AVAILABILITY>020199</AVAILABILITY>
</PLANT>
```

<PLANT>

<COMMON>Cardinal Flower</COMMON>
<BOTANICAL>Lobelia cardinalis</BOTANICAL>
<ZONE>2</ZONE>
<LIGHT>Shade</LIGHT>
<PRICE>$3.02</PRICE>
<AVAILABILITY>022299</AVAILABILITY>
</PLANT>
</CATALOG>

ii)

Persons			
PersonID	Name	Age	Salary
P1	John	32	$400000
P2	Johnny	33	$410000
P3	Janet	31	$400000
P4	Jeremy	32	$450000
P5	Justin	33	$600000
P6	Jazmyn	35	$250000
P7	Judy	30	$900000
P8	Jolly	33	$100000
P9	Jack	31	$120000

4 If you have rights to access VTOP access login permission also you need to collect the 2nd and 3rd year student information data of all VIT branches from VTOP and need to be stored in Hadoop echo system, to do so identify the suitable tools available in Hadoop ecosystem and brief on it.

5 Explain the Map reduce process supports to count the total number of occurrences of each single word present in text document