RAMAIAH UNIVERSITY OF APPLIED SCIENCES

Faculty of Engineering and Technology			
Ramaiah University of Applied Sciences			
Department	Computer Science and Engineering	Programme	B. Tech
Semester/Batch	8th /2017		
Course Code	CSC409A	Course Title	Data Analytics
Course Leader	Mohan Kumar K N/ E. Ami Rai		

Assignment - 01			
Register No		Name of Student	

Sections		Marking Scheme	Marks		
			Max Marks	First Examiner Marks	Moderator Marks
Part-A					
	A 1.1	Data Analytics and its applications	04		
	A 1.2	Real world use cases –industries, technologies	04		
	A 1.3	Barriers to adoption – technical and non technical	04		
	A 1.4	future of analytics – best guess	04		
	A1.5	Justification and stance taken	04		
		Part-A Max Marks	**20**		
Part B 1					
	B 1.1	Phase 2 tools (any two)	8		
	B 1.2	Phase 4 tools (any two)	8		
	B 1.3	kinds of use scenarios	4		
		Part-B 1 Max Marks	**20**		
Part B 2					
	B2.1	Introduction to the recommended method(s)	06		
	B2.2	Suggestion and selection the attributes	08		
	B2.3	Justification	06		
		Part-B 2 Max Marks	**20**		
Part B 3					
	B3.1	Recommend relevant solution	08		
	B3.2	Issues with information retrieval	04		
	B3.3	justification	08		

		Part-B 3 Max Marks	**20**		
Part B 4					
	B4.1	Introduction to big data platform	04		
	B4.2	Problem solving approach	04		
	B4.3	Design and implementation	04		
	B4.4	Performance analysis	04		
		Part-B 4 Max Marks	**20**		
		Total Assignment Marks	**100**		

Course Marks Tabulation				
Component- CET B Assignment	**First Examiner**	**Remarks**	**Second Examiner**	**Remarks**
A				
B.1				
B.2				
B.3				
B.4				
Marks (Max 1000)				
Marks (out of 25)				

Signature of First Examiner Signature of Second Examiner

Please note:

1. Documental evidence for all the components/parts of the assessment such as the reports, photographs, laboratory exam / tool tests are required to be attached to the assignment report in a proper order.
2. The First Examiner is required to mark the comments in RED ink and the Second Examiner's comments should be in GREEN ink.
3. The marks for all the questions of the assignment have to be written only in the **Component – CET B: Assignment** table.
4. If the variation between the marks awarded by the first examiner and the second examiner lies within +/- 3 marks, then the marks allotted by the first examiner is considered to be final. If the variation is more than +/- 3 marks then both the examiners should resolve the issue in consultation with the Chairman BoE.

RAMAIAH UNIVERSITY OF APPLIED SCIENCES

Assignment 1

Term - 1

Instructions to students:

1. The assignment consists of **5** questions: Part A – **1** Question, Part B- **4** Questions.
2. A maximum mark is **100**.
3. The assignment has to be neatly word processed as per the prescribed format.
4. The maximum number of pages should be restricted to **25**.
5. Restrict your report for Part-A to 5 pages only.
6. Restrict your report for Part-B to a maximum of 20 pages.
7. The printed assignment must be submitted to the course leader.
8. **Submission Date: XX / 0X/2021**
9. **Submission after the due date is not permitted.**
10. **IMPORTANT**: It is essential that all the sources used in preparation of the assignment must be suitably referenced in the text.
11. Marks will be awarded only to the sections and subsections clearly indicated as per the problem statement/exercise/question

Preamble:

The course is intended to teach the design, development, analysis and evaluation of Data Analytics applications. Employing appropriate techniques, methods and technology in various domains of computing is discussed. Data mining algorithms, tuning them for a given application and actionable interpretations are emphasized. It helps to solve practical applications with data analysis, turning business intelligence into real-world outcomes. Students are trained to analyses, visualize and interpret the data and associated implicit insights.

PART – A **20 Marks**

Data Analytics is the science of analyzing data to convert information into useful knowledge. This helps to understand the world better and in many contexts enable to make better decisions. Technological advances and associated changes in daily life have produced a rapidly expanding new content/data/information sources. Although many opportunities exist, big data and data analytic technologies also present many challenges such as understanding data, quality of data, security and real time integration. Most of the organizations are facing the imbalance on data analysts and the amount of data being produced.

Debate on the statement "**Data deluge in information and starving for knowledge after Data Analytics operation"**

Your debate should include:

A1.1 Introduction to Data Analytics and its applications
A1.2 Illustration with real world examples
A1.3 Discussion on the barriers for adoption
A1.4 Discussion of the future of analytics
A1.5 Stance taken and justification

PART – B **80 Marks**

B.1 **20 Marks**

Data analytics lifecycle defines analytics process and best practices spanning from discovery to project completion. Consider data preparation and model building phases of data analytics lifecycle and select relevant tools for each phase and defend with suitable example. Perform the following:

B1.1 Discuss data preparation phase tools
B1.2 Discuss model building phase tools
B1.3 Justify with suitable scenarios

B.2 **20 Marks**

A data science team is working on a book recommendation problem. The books are available in different categories. If a customer buys a book, he or she should be recommended other books and categories of books of preference:

B2.1 Model different method(s) to address the above issue.
B2.2 Identify suitable attributes.
B2.3 Justify your solution by comparison.

B.3 **20 Marks**

A certain company 'A' wants to market its new product. Manual marketing is time consuming and a costly process. The model should spread the product information like virus (viral marketing).

B3.1 Recommend a solution
B3.2 Discuss issues
B3.3 Justification

B.4 **20 Marks**

Inverted index: Inverted Index is mapping of text in the document. It is mainly used in search engines and provides faster lookup on text searches. The output file must contain a list of all words with frequencies of their occurrences. The Map method should read the input file and output occurrences of words as the key-value pair. Reducer method can use a hash map to count the occurrences for a particular word key. Solve the problem using Big data (Hadoop).

RAMAIAH UNIVERSITY
OF APPLIED SCIENCES

Your report should include:

B4.1 Introduction to Big data platform
B4.2 Problem solving approach
B4.3 Design and implementation
B4.4 Performance analysis

ఌఌఌ

RAMAIAH UNIVERSITY OF APPLIED SCIENCES

Detailed Marking Scheme

Question No.	Tasks –Steps involved	Marks Allotted for steps	Instructor's Expected Solution	Total Allotted Marks for the question
A	Data Analytics and its applications Illustrate with Real world use cases Discuss the barriers to adoption Discuss the future of analytics Justification and stance taken	4 4 4 4 4	• Definition and any of its applications • Real life examples • Technical and non technical • With respect to best guess • Based on facts and figures	**20**
B1	Discussion of Phase 2 tools Discussion of Phase 4 tools Discuss which kinds of use scenarios	8 8 4	• Name any 2-3 examples • Name any 2-3 examples • With example explain it	**20**
B2	Introduction to the recommended method(s) Why you are suggesting it? How does it select the attributes? Justification	6 8 6	• Decision making • Based on problem statement and its requirements	**20**

B3	B3.1 Recommend relevant analysis B3.2 Describe the issues B3.3 Discuss it with example B3.4 Justification	4 6 6 4	• Model a Decision Support system • Based on problem statement and its requirements	**20**
B4	4.1 Introduction to big data platform. 4.2 Problem solving approach. 4.3 Design and implementation. B4.4 Discuss its performance.	4 4 6 6	• HDFS(Hadoop Distributed File System) big data platform • Problem solving approach MapReduce is used • Design and implementation using Hadoop/R Using Java platform • Performance	20