GUJARAT TECHNOLOGICAL UNIVERSITY

Syllabus for PG Diploma in Bioinformatics, Semester - 1

Subject Name: Cloud Computing, Data Mining & Visualization
Subject Code: 1310204 W.E.F 2021-22

1. Learning Outcomes

Learning Outcome Component	Learning Outcome (Learner will be able to)
Domain Knowledge	• Understand the fundamentals of Cloud Computing, Data Mining & Visualization.
Critical thinking, Logic Building, Problem Solving	• Resolving real-world Cloud Computing problems, Practicing Data Mining. • Assess different problem-solving approaches. • Building logical step-by-step solutions to solve Cloud Computing problems and analyze large volumes of data.
Exposure and Cross-discipline Understanding	• Understanding the application of Cloud Computing and Data Mining in Computational Biology.
Effective Communication	• Communicate ideas clearly and effectively
Professional & Ethical Behaviour	• Transparency, honesty and ethical practices in Cloud Computing, data handling and processing.

LO – PO Mapping: Correlation Levels:

1 = Slight (Low); 2 = Moderate (Medium); 3 = Substantial (High), "-"= no correlation

Sub Code: 1310204	PO1	PO2	PO3	PO4	PO5	PO6
LO1: Domain Knowledge	3	3	3	3	2	3
LO2:Critical thinking, Logic Building, Problem Solving	3	3	3	3	2	3
LO3:Exposure and Cross-discipline Understanding	3	3	3	3	2	3
LO4: Effective Communication	2	2	3	3	2	3
LO5: Professional & Ethical Behaviour	2	2	2	3	3	3

2. Course Duration: The course duration is 45 sessions of 60 minutes each.

3. Course Contents:

Module No:	Module Content	No. of Sessions	70 Marks (External Evaluation)
I	**Introduction to Cloud Computing** Overview, Roots of Cloud Computing, Layers and Types of Cloud, Desired Features of a Cloud, Benefits and Disadvantages of Cloud Computing, Cloud Infrastructure Management, Infrastructure as	10	20

	a Service Providers, Platform as a Service Providers, Challenges and Risks.		
2	**Cloud Architecture, Services and Applications** Exploring the Cloud Computing Stack, Connecting to the Cloud, Infrastructure as a Service, Platform as a Service, Saas Vs. Paas, Using PaaS Application Frameworks, Software as a Service Cloud Deployment Models, Public vs Private Cloud.	10	10
3	**Abstraction and Virtualization** Introduction to Virtualization Technologies, Load Balancing and Virtualization, Understanding Hyper visors, Understanding Machine Imaging, Porting Applications, Virtual Machines Provisioning and Manageability Virtual Machine Migration Services, Virtual Machine Provisioning and Migration in Action, Cloud Based Case-Studies specifically for Bioinformatics: Overview of Cloud services, Designing Solutions for the Cloud, Implement & Integrate Solutions, Emerging Markets and the Cloud, Tools for Building Private Cloud: IaaS using Eucalyptus, PaaS on IaaS – AppScale.	10	10
4	**Data Mining** Data Mining – Data ware housing-OLAP-Data Pre-processing, Data mining classification-Basic concepts of classification, Decision Tree introduction, K-nearest neighbour classifier, Naïve Bayes classification, frequent pattern mining, association rules, Apriori Algorithm, study and applications with biological data. Study of AI and machine learning algorithms in computational biology, combining theory with practice. Techniques and advances in the field of computational biology, analysing sets of biological data generated by high-throughput technologies.	10	20
5	**Data Visualization** Overview of data visualization – Definition, concepts different types of chart and graphs, Emergence of data visualization, overview of PowerBi.	5	10

6	**Cloud Computing** Virtual Machine configuration, launch and customization Architecture & Use of Aneka/Eucalyptus/KVM/ Microsoft Azure/Amazon Web Service Pull in data set from cloud repositories Sharing and scale up VM **Hadoop** Introduction to Apache Hadoop, Hadoop Distributed File System (HDFS) and associated open source software projects like cloud Aligner, cloudburst, cross bow, Myrna, BioDoop etc	---	(30 marks CEC)

4. Pedagogy:

- ICT enabled Classroom teaching
- Case study
- Practical/live assignment
- Interactive class room discussions

5. Evaluation:
Students shall be evaluated on the following components:

	Internal Evaluation	**(Internal Assessment – 20 Marks)**
A	• Assignments	10 marks
	• Class Presence	5 marks
	• Record maintenance	5 marks
B	**Mid-Semester Examination**	**(Internal assessment-30 Marks)**
C	**End-Semester Examination**	**(External assessment-70 Marks)**

6. Reference Books:

No	**Author**	**Name of the Book**	**Publisher**	**Year of Publication / Edition**
1	Rajkumar Buyya, James Broberg and Andrzej Goscinski	Cloud Computing: Principles and Paradigms	Wiley India	Latest edition

2	Barrie Sosinsky	Cloud Computing Bible	Wiley Publishing Inc.	Latest Edition
3	Miller Michael Pearson Velte T., Velte A., Elsenpeter R.	Cloud Computing: Web Based Applications that Change the Way You Work and Collaborate Online	Tata McGraw Hill	Latest Edition

Note: Wherever the standard books are not available for the topic appropriate print and online resources, journals and books published by different authors may be prescribed.

7. List of Journals/Periodicals/Magazines/Newspapers / Web resources, etc

- https://pubmed.ncbi.nlm.nih.gov/21210976/
- https://github.com/palsingh104/Data-Mining-and-Bioinformatics
- https://omicstutorials.com/introduction-to-machine-learning-bioinformatics/

Course Outcomes:

On completion of this course, students should be able to:

- Understand the basic concepts, terminology of cloud computing, data mining and visualization.
- Gain working knowledge of different cloud computing, data mining and visualization tools.
- Develop skills to implement tools of cloud computing, data mining and visualization tools in different bioinformatics work-flows.