Syllabus

First Semester Courses in M.Sc. Big Data Analytics 2023-2024

Contents:

- **Syllabus for Elective courses:**
  - PSBDA6001EL1: Computing for Data Science
  - PSBDA6002EL1: Data Structures

- Evaluation and Assessment guidelines

ST. XAVIER'S COLLEGE-AUTONOMOUS MUMBAI-400 001

PRINCIPAL
ST. XAVIER'S COLLEGE
AUTONOMOUS
MUMBAI - 400 001.

APPROVED SYLLABUS

M.Sc-I (Big Data Analytics)

Course Code: PSBDA6001EL1

Title: Computing for Data Science
Credits: 4 (Theory -3 -Total 45 hr and Practical- 1-Total -30hr)

Course Objective:
1. Understand the various data structures.
2. Get versed with R.
3. Understand the concept of data science

Course Outcomes:
On completing the course, the student will be able to:
1. Acquire proficiency in at least one programming language commonly used in data science, such as Python or R.
2. Acquire skills for cleaning, transforming, and preprocessing data to make it suitable for analysis
3. Understand the basic concept of data structure
4. Develop the required skills in Python programming.

Unit 1	**15**
Computer Packages – R and Python: Usage of R and Python – data handling, data analysis, statistical modeling with illustration in python and R	
Unit 2	15
Data Structure & Concepts of Computation using Java Algorithms, Convergence, Complexity with illustrations, some sorting & searching algorithms, some numerical methods e.g. Newton-Raphson, Steepest ascent using Java	
Unit 3	**15**
Computing Methodologies: Monte-Carlo simulations of random numbers and various statistical methods, memory handling strategies for big data.	

List of Recommended Reference Books

1. Irizarry, R. A. (2019). Introduction to Data Science: Data Analysis and Prediction Algorithms with R. United States: CRC Press.
2. Grolemund, G. (2014). Hands-On Programming with R: Write Your Own Functions and Simulations. United States: O'Reilly Media.
3. Goodrich, M. T., Tamassia, R., Goldwasser, M. H. (2014). Data Structures and Algorithms in Java. United Kingdom: Wiley.

ASSESSMENT:
THEORY:
CIA I: Written test **for 20 marks**
CIA II: Assignments / Project / Presentation / Case Study/ Written Test **for 20 marks**
End semester Exam: 60 marks
Template for the Elective course End Semester examination in Semester I.

UNITS	KNOWLEDGE	UNDERSTANDING	APPLICATION and ANALYSES	TOTAL MARKS- Per unit
1	5	5	10	20
2	5	10	5	20
3	5	5	10	20
-TOTAL - Per objective	15	20	25	60
% WEIGHTAGE	25%	33%	42%	100%

M.Sc-I (Big Data Analytics)

Course Code: PSBDA6002EL1

Title: Data Structures
Credits: 4 (Theory -3 -Total 45 hr and Practical- 1-Total -30hr)

Course Objective:

1. Understand the role of complexity in developing application.
2. Understand various sorting algorithms
3. Learn linked list and trees, and its application

Course Outcome:

On completing the course, the student will be able to:

1. Acquire strong understanding of fundamental data structures such as arrays, linked lists, stacks, and queues.
2. Analyze the time and space complexity of algorithms related to data structures.

Unit 1 15

Introduction to Data Structure, Stacks, Queues and Recursion: Introduction and Complexity Data Types, Data Structure, Abstract Data Types, what is an algorithm, Rate of growth and its graph with analysis. Time Complexity (Big Oh and Big Omega, Theta Notation,), Master Theorem for divide and conquer, Problems on complexity for divide and conquer, Master Theorem for subtract and conquer and problems on it.

Stacks: Introduction to Stack, Array Representation of Stack, Notations (infix, prefix and post fix notation), understanding stack operations push, pop, peek, algorithm for converting infix to postfix and infix to prefix, algorithm to separate operator and operand from given string, Queue concept of queue, inserting deleting data in queue, concept of circular queue, inserting deleting data in circular queue.

Unit 2 15

Linked List

What is a Linked List, Comparing Linked List with Arrays, advantage and disadvantage of Linked List? Singly Linked List, traversing, insertion node at beginning, ending and at middle, deleting node from beginning, ending and at middle for singly linked list, Doubly Linked List, Insertion node at beginning, ending and at middle for doubly linked list, deleting node from beginning, ending and at middle for doubly linked list, circular linked list.

Unit 3 15

Trees

What is a Tree, Binary Tree and Binary search Tree, properties of Binary Tree, Structure of Binary Tree, Types of Binary Trees (Strict Binary Tree, Full Binary Tree, complete Binary Tree, Almost complete Binary Tree),

inorder, preorder and post order traversal with recursion and without recursion, searching element in Binary Search Tree.

List of Recommended Reference Books

1. Karumanchi, N. (2011). Data Structures and Algorithms Made Easy in Java: Data Structure and Algorithmic Puzzles, Second Edition. India: CareerMonk Publications.
2. Hubbard, J. R. (2009). Schaum's Outline of Data Structures with Java, 2ed. United Kingdom: McGraw Hill LLC.
3. Horowitz, E., Sahni, S. (1976). Fundamentals of data structures. United Kingdom: Computer Science Press.

ASSESSMENT:
THEORY:
CIA I: Written test **for 20 marks**
CIA II: Assignments / Project / Presentation / Case Study/ Written Test **for 20 marks**
End semester Exam: 60 marks
Template for the Elective course End Semester examination in Semester I.

UNITS	KNOWLEDGE	UNDERSTANDING	APPLICATION and ANALYSES	TOTAL MARKS- Per unit
1	5	5	10	20
2	5	10	5	20
3	5	5	10	20
-TOTAL - Per objective	15	20	25	60
% WEIGHTAGE	25%	33%	42%	100%

Syllabus

Second Semester Courses in M.Sc. Big Data Analytics 2023-2024

Contents:

APPROVED SYLLABUS

M.Sc-I(Big Data Analytics)

Course Code: PSBDA6003EL1

Title: OPERATIONS RESEARCH

Credits :4 (Theory -4-Total 60 hr)

Course Objective:

1. Solve linear programming problems using graphical and simplex methods.
2. Apply assignment models to allocation problems.
3. Understand the differences between linear and non-linear programming.
4. Formulate non-linear programming problems.
5. Describe the characteristics of queuing processes and their relevance in real-world systems.

Course outcomes:

On completing the course, the student will be able to:

1. Understand and apply fundamental concepts of linear programming and assignment models to solve real-world optimization problems.
2.
3. Analyze and optimize non-linear programming problems using appropriate techniques and algorithms.
4. Formulate and solve transportation models to address logistics and supply chain optimization challenges.
5. Comprehend the characteristics of queuing processes and their applications in various service systems.
6. Utilize queuing models to analyze and design efficient queuing systems in practical scenarios.

Unit 1 Review of Linear Programming, Assignment Models	**15**
Unit 2 Non-Linear Programming	**15**
Unit 3 Transportation Models.	**15**
Unit 4 Queuing Models: Characteristics of Queuing Process, Poisson Process, Birth Death Process, Single-Server Queues, Multi-Server Queues, Queues with Truncation, Finite-Source Queues, Numerical Techniques & Simulation.	**15**

SUGGESTED BOOKS:

1. Gupta, P. K., Hira, D. S. (1995). Problems in Operations Research: Principles and Solutions. India: S. Chand.

2. Shortle, J. F., Thompson, J. M., Gross, D., Harris, C. M. (2018). Fundamentals of Queueing Theory. Germany: Wiley.

ASSESSMENT:
THEORY:
CIA I: Written test **for 20 marks**
CIA II: Assignments / Project / Presentation / Case Study/ Written Test **for 20 marks**
End semester Exam: 60 marks
Template for the Elective course End Semester examination in Semester II.

UNITS	KNOWLEDGE	UNDERSTANDING	APPLICATION and ANALYSES	TOTAL MARKS- Per unit
1	5	5	5	15
2	5	5	5	15
3	5	5	5	15
4	5	5	5	15
-TOTAL - Per objective	20	20	20	60
% WEIGHTAGE	34%	33%	33%	100%

M.Sc-I (Big Data Analytics)

Course Code: PSBDA6004EL1

Title: Linux and Cloud Computing Fundamentals

Credits :4 (Theory -4 -Total 60 hr)

Learning Objective:
Students will be able to deploy applications on the cloud.

Course Outcome:
On completing the course, the student will be able to:

1. Understand the fundamental concepts of Linux and the cloud.
2. Learn how to manage resources and deploy applications on a server and a cloud instance.
3. Understand the fundamental concept of distributed systems and Virtualization.

Unit 1 **15**

Fundamentals of Linux

Introduction to Linux

- Brief History of Linux
- The Linux Kernel
- Flavors of Linux
- Installing Linux
- Basic Linux Commands: ls, mv, cp, vi, nano, awk, sed, cut, sort, grep, cat, etc
- Inode Structure and Recovery of Files
- Process Management (top, kill, pkill, etc)
- Package Manager (apt, yum, etc)

Shell and Shell Scripting

- Shell Types
- Global variables
- Basic syntax for writing scripts
- Modifying the .bashrc file

User Management

- The root user
- Creating Users and groups
- Adding Home Directories
- Giving users sudo privileges
- Removing Users and Groups

Permission Management

- Read, Write and Execute privileges
- Granting and revoking permissions to access directories and files
- Umask,Chown, Immutable bit(chattr), Sticky Bit
- SUID, SGID and Access Control Lists

Services

- What are services in linux?
- Using systemctl to enable, start, stop and disable services

SSH

- How does SSH work?
- Setting up OpenSSH on Linux
- Using scp/rsync to copy files between hosts
- Making your host as an ssh server to allow multiple users to access the server simultaneously

VNC

- An overview of VNC
- Setting up a VNC Server on Linux
- Connecting to the server by using a VNC Client.

Unit 2 **15**

Servers, Virtualization, Containers, etc

Servers

- Overview of various Linux Server OS flavors
- Using Apache/Nginx for hosting web applications
- Webmin/Virtualmin/Cockpit for remotely administering the Server
- Deploying a database server such as MariaDB, MySQL, MongoDB, etc

Distributed Systems

- What are Distributed Systems?
- Types of Distributed Systems
- Examples
- High Throughput Systems vs High Performance Systems

Virtualization

- Introduction to Virtualization
- Server Consolidation
- Hypervisors and it types
- Characteristics of Virtualization
- Snapshot Management
- Having a look at various hypervisors and testing them out. (KVM, Xen, VMware, Hyper-V etc)

Storage Systems and their Protocols

- Types of Storage
- Using RAID in the datacenter.
- SAS vs SATA vs NVMe
- Fiber Channel
- NAS, SAN and DAS
- SMB, iSCSi, NFS/CIFS
- Implementing TrueNAS to understand the various protocols

Containers

- What are containers?
- Various Container Technologies (Docker, OpenVZ, LXC, etc)
- Persistent vs Transient Containers
- A brief introduction to Kubernetes (Container Orchestration)
- Setting up a Kubernetes Cluster
- Scaling with Kubernetes

Unit 3 **15**

Fundamentals of Cloud Computing

Introduction to Cloud Computing

- What is Cloud Computing?
- Building Blocks of the Cloud
- Cloud Deployment Models
- As a Service (*aaS) Offerings of the Cloud
- Introduction to Server-less Computing

Service Level Agreements

- Key components of an SLA
- Basic Terminologies: -MTBF, MTTR and MTTF
- Types of SLA's
- High Availability (Based on the number of nine's)
- Studying SLA's of various cloud providers.

Unit 4 **15**

Services offered by various cloud providers. (AWS, GCP, Azure, etc)

The following hands-on demonstrations will be done using a cloud service provider.

- User Management
- Bill Management
- Implementing virtual instances/containers
- Implementing microservices
- Implementing a cloud database
- Implementing High Availability
- Defining firewall rules
- Delegating rights to various users
- Creating a complete web service along with an SSL certificate
- Deploying a function on the server-less framework

Using Object Storage

Suggested Books

- Bresnahan, C., Blum, R. (2021). Mastering Linux System Administration. United States: Wiley.
- Cobbaut, P. (2016). Mastering Linux - Fundamentals. Hong Kong: Samurai Media Limited.
- Dean, A. K. (2018). Linux Administration Cookbook: Insightful Recipes to Work with System Administration Tasks on Linux. United Kingdom: Packt Publishing.
- Portnoy, M. (2023). Virtualization Essentials. United Kingdom: Wiley.
- Comer, D. (2021). The Cloud Computing Book: The Future of Computing Explained. United Kingdom: CRC Press.

Website References

https://docs.vmware.com/en/VMware-vSphere/index.html

https://www.linux-kvm.org/page/Documents

https://cloud.google.com/docs

https://docs.aws.amazon.com

https://docs.microsoft.com/en-us/azure/?product=featured

ASSESSMENT:
THEORY:
CIA I: Written test **for 20 marks**
CIA II: Assignments / Project / Presentation / Case Study/ Written Test **for 20 marks**
End semester Exam: 30 marks and Project: 30 marks
Template of Elective course End Semester examination in Semester II

UNITS	KNOWLEDGE	UNDERSTANDING	TOTAL MARKS- Per unit
1	5	5	10
2		5	05
3	5	5	10
4		5	05
-TOTAL - Per objective	10	20	30
% WEIGHTAGE	33.3%	66.6	100%

****