

DBDA

**PG-DBDA Aug 2024
Syllabus and Marks Distribution**

Duration: 28 Classroom hours + 22 Lab hours

Objective: To introduce Linux environment and hands on Linux commands.

Prerequisites: Knowledge of Computer Fundamentals

Evaluation method: Theory exam – 40% weightage
 Lab exam – 40% weightage
 Internal exam – 20% weightage

List of Books / Other training material

Reference:

1. Linux: The Complete Reference – Petersen/ TMH 6th Edition
2. The Linux Programming Interface: Linux and UNIX System Programming Handbook
3. Pro Bash Programming: Scripting the GNU/Linux Shell, Second Edition
4. Beginning Unix – Joe Marilino (Wrox Publication)
5. Linux Command Line and Shell Scripting Bible – Blum (Wiley – India)

Linux Programming

Session 1 & 2:

Lecture:

Linux History and Operation

- o The Evolution of Linux
- o The GNU Movement and the GPL
- o Linux Operations as a Server
- o The Architecture and Structure of Linux

Installing and Configuring Linux (Ubuntu and CentOS)

- o Introduction to Installation and Media Types
- o Performing a Custom Linux Server Installation
- o Run Levels and the Startup/Shutdown Sequence
- o Logging In and Out of a Linux System

Basic Commands

(ls, cp, mv, sort, grep, cat, head, tail, man, locate, find, diff, file, rm, mkdir, rmdir, cd, pwd, ln and ln -s, gzip and gunzip, zip and unzip, tar and its variants, touch, echo, who, whoami, ps, kill, makefile, etc.)

Assignment – Lab:

1. Getting Acquainted with the Linux Environment
2. Use various commands in the Linux system.
3. As root, create a directory dbda and under it create a directory named test and create 100 files under it with names file1, file2, ..., file100 - all this using a single command.

Session 3:**Lecture:****Gaining Confidence with Linux**

- o Access control list and chmod command, chown and chgrp commands
- o Commands like telnet, ftp, ssh, and sftp
- o Basics of I/O system with mount and umount.

Vi/vim/gedit Editor

- o Features and different modes of vi editor
- o Editing using vi editor
- o Find and replace commands
- o Cut-copy-paste commands
- o The set command
- o Other related commands of vi

Assignment – Lab:

- Create the file /tmp/acl_file.
 1. Allow Larry and Curly to rwx the file.
 2. Don't allow Moe to access the file (rwx).
 3. All members of the group stooges (except Moe) should be able to access the file (rw).

Session 4, 5, & 6:**Lecture:****Linux Shell Programming**

- o Introduction to Shells
 - a.shell
 - b.Different types of Linux shells
 - c.Bourne Again Shell (BASH)
 - d.Shell variables (environment and user-defined)
 - e.Shell files (.bashrc, .profile, .bash_profile, .bash_logout)
 - f. Positional parameters
- o Get started with simple scripts (User variable, expr, multiple commands)
- o Wildcards (*) and (?)
- o Command line arguments
- o Arithmetic in shell scripts
- o read and echo commands in shell scripts
- o The tput command
- o Taking decisions:
 - if-then-fi
 - if-then-else-fi
 - The test command (file tests, string tests)
 - Nested if-else
 - The case control structure
- o The loop control structure
 - The while, until, and for loop structures
 - The break and continue statements
- o Shell metacharacters

PG-DBDA August 2024

-
- o Command line expansion
 - o Directory stacks manipulation
 - o Job control, history, and processes
 - o Built-ins and functions
 - o Shell Files

Assignment – Lab:

- Change the shell of user3 to nologin. Now login as user4 and try to switch to user3. Observe the result.
- Login as root, create a file filewithacl and apply ACL on it in such a way that only user5 is able to read and write to it. Note: root will do all its work under the dbda folder.

Session 7:

Lecture:

Git / GitHub

- o Introduction to Version Control Systems
- o Creating GitHub repository
- o Using Git – Introduction to Git commands.

Cloud Computing Reference Books:

Reference:

1. Cloud Computing Black Book by Kailash Jayaswal, Dreamtech
2. Mastering Cloud Computing by Rajkumar/ McGraw Hill Education
3. Cloud Computing a Practical Approach by Anthony T. Velte/ McGraw Hill Education
4. Architecting the Cloud: Design Decisions for Cloud Computing Service Models (SaaS, PaaS, and IaaS)
5. Cloud Computing
6. An Introduction to Parallel Computing: Design and Analysis of Algorithms (Authors: Vipin Kumar, Ananth Grama, Anshul Gupta, George Karypis)
7. High Performance Cluster Computing: Architectures & Systems (Volume-1) by Rajkumar Buyya, Pearson
8. Parallel Programming in C with MPI and Open MPI, Michael, TMH
9. High-Performance Computing on Complex Environments

Session 8:

Lecture:

- o Introduction to Cloud
- o paradigms
- o Characteristics and benefits
- o Understanding Cloud Vendors (AWS/Azure/GCP)
- o Definition
- o Characteristics
- o Components

Lab Assignments:

- o Study about cloud and other similar configurations.
- o Explore available solutions.
- o Cloud Architecture.

Session 9 & 10:**Lecture:**

- o Introduction to SaaS
- o Pros and Cons of SaaS Model
- o Traditional Packaged Software vs. SaaS
- o SaaS Examples
- o Introduction to IaaS
- o Examples
- o Introduction to Virtualization
- o Types and Uses of Virtualization
- o Virtual Machine Provisioning
- o Virtual Machine Migration Services
- o Private Cloud Computing Deployment
- o Introduction to PaaS
- o Challenges of Cloud Environment
- o Hypervisor
- o Comparisons of Web Services
- o Organizational Scenarios of Clouds

Lab Assignments:

- o Provide a solution on cloud as SaaS using available systems.

Session 11 & 12:**Lecture:**

- o Administering & Monitoring Cloud Services
- o Benefits and limitations
- o Deploy application over cloud.
- o Comparison among SaaS, PaaS, IaaS
- o Cloud Computing Basics
- o Cloud Products and Solutions
- o Cloud Pricing
- o Compute Products and Services

Session 13 & 14:**Lecture:**

- o Elastic Cloud Compute
- o Dashboard
- o Launching Linux VM
- o Accessing Linux VM

PG-DBDA August 2024

-
- o Launching & Accessing Windows Server VM
 - o Introduction to AWS
 - o Introduction to Virtual Private Cloud (VPC) Setup
 - o Services provided by AWS: EC2, Lambda, AWS Storage Services S3
 - o Introduction to Azure
 - Azure Platform (Azure Portal, Azure CLI, Azure PowerShell)
 - Azure Data Services (Azure SQL Database, Cosmos DB)
 - Azure Storage (Blob Storage, File Storage, Queue Storage)
 - Azure Functions

Lab:

- o Study about cloud and other similar configurations.
- o Exposure to big data technologies on cloud.
- o Create AWS EC2 instance.
- o Create AWS Lambda.
- o Create AWS Storage Services S3.
- o Create AWS VPC.
- o Deployment of application on AWS using GitHub.
- o Deployment of application on Azure using GitHub.
- o Deployment of application on Azure using Azure DevOps.
- o Deploy AI and analytics workloads on cloud environments with a sample mini project.

Suggested Teaching Guidelines for
Python & R Programming PG-DBDA
August 2024

Duration: 44 classroom hours + 36 lab hours

Objective: To introduce the student to Python programming & R programming concepts.

Prerequisites: Knowledge of programming in any language like C, C++ and some basic statistical knowledge.

Evaluation method: Theory exam – 40% weightage
 Lab exam – 40% weightage
 Internal exam – 20% weightage

List of Books / Other training material

Text Book:

1. Python for Everybody Exploring Data using Python, Charles R. Severance, Shroff Publishers & Distributors, 1st Edition.

Reference Book:

1. Introduction to Computer Science using Python, Charles/ Wiley
2. Python Power!: The Comprehensive Guide
3. Python Crash Course: A Hands-on, Project-Based Introduction to Programming
4. Beginning Programming with Python For DummiesLearning Python by: Fabrizio Romano
5. Python Projects by Laura Cassell , Alan Gauld / Wiley
6. Python Cookbook by David B. Brain K. Jones / Shroff / O'reilly Publisher
7. Head First Python by Paul Barry / Shroff / O'reilly Publisher
8. Professional Iron Python by John Paul Muller / Wiley India Pvt Ltd
9. Beginning Programming with Python for Dummies by John Paul Muller / Wiley India Pvt Ltd

Note: Each session mentioned is for theory and of 2 hours duration. Lab assignments are indicatives, faculty need to assign more assignments for better practice.

Session 1:

- Installing Python
- Introduction to Python
- Basic Syntax, Data Types, Variables, Operators, Input/Output
- Declaring Variables, Data Types in Programs
- Your First Python Program
- Flow of Control (Modules, Branching)
- If, If-else, Nested if-else
- Looping, For, While, Nested Loops
- Control Structure
- Uses of Break, Continue, and Pass

Suggested Teaching Guidelines for
Python & R Programming PG-DBDA
August 2024

Lab Assignments:

Q.1. Using for loop, write and run a Python program for this algorithm.

Here is an algorithm to print out $n!$ (n factorial) from $0!$ to $10!$:

1. Set $f = 1$
2. Set $n = 0$
3. Repeat the following 10 times:
 - a. Output $n, "!" = ", f$
 - b. Add 1 to n
 - c. Multiply f by n

Q.2. Modify the program above using a while loop so it prints out all of the factorial values that are less than 2 billion. (You should be able to do this without looking at the output of the previous exercise.)

Session 2:

- Strings and Tuples
- Accessing Strings
- Basic Operations
- Assigning Multiple Values at Once
- Formatting Strings
- String Slices

Lab Assignments:

Q.1. Write a program that asks the user how many days are in a particular month, and what day of the week the month begins on (0 for Monday, 1 for Tuesday, etc), and then prints a calendar for that month. For example, here is the output for a 30-day month that begins on day 4 (Thursday):

| S | M | T | W | T | F | S |
|----|----|----|----|----|----|----|
| | | | 1 | 2 | 3 | |
| 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| 25 | 26 | 27 | 28 | 29 | 30 | |

Q. 2. Define a procedure histogram() that takes a list of integers and prints a histogram to the screen. For example, histogram([4, 9, 7]) should print the following:

```
*****
*****
*****
```

Q. 3. Write a version of a palindrome recognizer that also accepts phrase palindromes such as "Go hang a salami I'm a lasagna hog.", "Was it a rat I saw?", "Step on no pets", "Sit on a potato pan, Otis", "Lisa Bonet ate no basil", "Satan, oscillate my metallic

Suggested Teaching Guidelines for
Python & R Programming PG-DBDA
August 2024

sonatas", "I roamed under it as a tired nude Maori", "Rise to vote sir", or the exclamation "Dammit, I'm mad!". Note that punctuation, capitalization, and spacing are usually ignored.

Q. 4. A pangram is a sentence that contains all the letters of the English alphabet at least once, for example: The quick brown fox jumps over the lazy dog. Your task here is to write a function to check a sentence to see if it is a pangram or not.

Session 3:

- Dictionaries
- Introducing Dictionaries
- Defining Dictionaries
- Modifying Dictionaries
- Deleting Items from Dictionaries
- Dictionary Comprehension

Lab Assignments:

Q. 1. In cryptography, a Caesar cipher is a very simple encryption techniques in which each letter in the plain text is replaced by a letter some fixed number of positions down the alphabet. For example, with a shift of 3, A would be replaced by D, B would become E, and so on. The method is named after Julius Caesar, who used it to communicate with his generals. ROT-13 ("rotate by 13 places") is a widely used example of a Caesar cipher where the shift is 13. In Python, the key for ROT-13 may be represented by means of the following dictionary:

```
key = {'a':'n', 'b':'o', 'c':'p', 'd':'q', 'e':'r', 'f':'s', 'g':'t', 'h':'u', 'i':'v', 'j':'w', 'k':'x', 'l':'y', 'm':'z', 'n':'a', 'o':'b', 'p':'c', 'q':'d', 'r':'e', 's':'f', 't':'g', 'u':'h', 'v':'i', 'w':'j', 'x':'k', 'y':'l', 'z':'m', 'A':'N', 'B':'O', 'C':'P', 'D':'Q', 'E':'R', 'F':'S', 'G':'T', 'H':'U', 'I':'V', 'J':'W', 'K':'X', 'L':'Y', 'M':'Z', 'N':'A', 'O':'B', 'P':'C', 'Q':'D', 'R':'E', 'S':'F', 'T':'G', 'U':'H', 'V':'I', 'W':'J', 'X':'K', 'Y':'L', 'Z':'M'}
```

Your task in this exercise is to implement an encoder/decoder of ROT-13. Once you're done, you will be able to read the following secret message:

Pnrfne pvcure? V zhpu cersre Pnrfne fnynq!

Note that since English has 26 characters, your ROT-13 program will be able to both encode and decode texts written in English.

Session 4:

- Working with Lists
- Introducing Lists
- Defining Lists
- Declaring, Assigning, and Retrieving Values from Lists
- Accessing Lists

Suggested Teaching Guidelines for
Python & R Programming PG-DBDA
August 2024

- Operations on Lists
- Adding Elements to Lists
- Searching Lists
- Deleting List Elements
- Using List Operators
- Mapping Lists
- Joining Lists and Splitting Strings
- Historical Note on String Methods

Lab Assignments:

1. Find the largest and smallest number in the list which taken as input from user using list operations.
2. Write a Python program to create multiple lists

Session 5 & 6:

- Function and Methods
- Defining and Calling Functions
- Types of Functions
- Function Arguments
- Anonymous Functions (Lambda, Map, List Comprehension)
- Global and Local Variables
- Using Optional and Named Arguments
- Using type, str, dir, and Other Built-In Functions
- Concepts of Modules
- Pickling

Lab Assignments:

Q. 1. Given a dictionary of students and their favourite colours:
 people={'Arham':'Blue','Lisa':'Yellow','Vinod':'Purple','Jenny':'Pink'}

1. Find out how many students are in the list
2. Change Lisa's favourite colour
3. Remove 'Jenny' and her favourite colour
4. Sort and print students and their favourite colours alphabetically by name

Write a function translate() that will translate a text into "rövarspråket" (Swedish for "robber's language"). That is, double every consonant and place an occurrence of "o" in between. For example, translate("this is fun") should return the string "tothohisos isos fofoon".

Q. 2. Write a program that contains a function that has one parameter, n, representing an integer greater than 0. The function should return $n!$ (n factorial). Then write a main function that calls this function with the values 1 through 20, one at a time, printing the returned results. This is what your output should look like:

Suggested Teaching Guidelines for
Python & R Programming PG-DBDA
August 2024

| | |
|----|---------|
| 1 | 1 |
| 2 | 2 |
| 3 | 6 |
| 4 | 24 |
| 5 | 120 |
| 6 | 720 |
| 7 | 5040 |
| 8 | 40320 |
| 9 | 362880 |
| 10 | 3628800 |

Q. 3. We can define sum from 1 to x (i.e. $1 + 2 + \dots + x$) recursively as follows for integer $x \geq 1$:

1, if $x = 1$
 x + sum from 1 to x-1 if $x > 1$

Complete the following Python program to compute the sum $1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10$ recursively:

```
def main():
    # compute and print 1 + 2 + ... + 10
    print sum(10)
def sum(x):
    # you complete this function recursively main ()
```

Q. 4. Define a function overlapping () that takes two lists and returns True if they have at least one member in common, False otherwise.

Q. 5. Write a function find_longest_word() that takes a list of words and returns the length of the longest one.

Q. 6. Write a function filter_long_words() that takes a list of words and an integer n and returns the list of words that are longer than n

Q. 7. Define a simple "spelling correction" function correct () that takes a string and sees to it that

1) two or more occurrences of the space character is compressed into one, and
 2) inserts an extra space after a period if the period is directly followed by a letter.
 e.g. correct ("This is very funny and cool.Indeed!") should return "This is very funny and cool. Indeed!"

Q. 8. In English, present participle is formed by adding suffix -ing to infinite form: go -> going. A simple set of heuristic rules can be given as follows:

- If the verb ends in e, drop the e and add ing (if not exception be, see, flee, knee, etc.)
- If the verb ends in ie, change ie to y and add ing

Python & R Programming PG-DBDA
August 2024

- For words consisting of consonant-vowel-consonant, double the final letter before adding ing
- By default, just add ing

Your task in this exercise is to define a function `make_ing_form()` which given a verb in infinitive form returns its present participle form. Test your function with words such as lie, see, move and hug. However, you must not expect such simple rules to work for all cases.

Session 7:

- Working with Tuples
- Introducing Tuples
- Accessing tuples
- Operation

Lab Assignments:

1. Write a Python program to find the repeated items of a tuple.
2. Write a Python program to sort a tuple by its float element.
 Sample data: `[('item1', '12.20'), ('item2', '15.10'), ('item3', '24.5')]`
 Expected Output: `[('item3', '24.5'), ('item2', '15.10'), ('item1', '12.20')]`
3. Write a Python program to count the elements in a list until an element is a tuple.
 Sample input : `list = [10, 20, 30, (40,50), 60]`
 Sample output = 3
4. Write a Python program to compute element-wise sum of given tuples, using “zip()” function
 Original tuples:
`(1, 2, 3, 4)`
`(3, 5, 2, 1)`
`(2, 2, 3, 1)`
 Element-wise sum of the said tuples:
`(6, 9, 8, 6)`

Session 8 & 9:

Advanced Python:

- Object Oriented Python
- OOPs Concept
- Object
- Indenting Code
- Native Data Types
- Declaring Variables
- Referencing Variables
- Object References
- Class and Object
- Attributes, Inheritance
- Overloading & Overriding

Suggested Teaching Guidelines for

Python & R Programming PG-DBDA
August 2024

- Data Hiding
- Regular Expressions Using Python
- Object Oriented Linux Environment
- Generators
- Decorators

Lab Assignments:

1. Create a class 'Student' with rollno, studentName, course ,dictionary of marks(subjectName -> marks [5]). Provide following functionalities
 - A. initializer
 - B. override `__str__` method
 - C. accept student data
 - D. Print student data for given id.
 - E. Print Student who has failed in any subject. Write menu driven program to test above functionalities.(accept records of 5 students and store those in list)
2. Write a menu driven program to maintain student information. for every student store studentid, sname, and m1,m2,m3 marks for 3 subjects. also store gpa in student class, add a function in student class to return GPA of a student
 - Calculate GPA()
 - $\text{gpa}=(1/3)*\text{m1}+(1/2)*\text{m2}+(1/4)*\text{m3}$
 - Create an array to store Multiple students.
 1. Display All Student
 2. Search by id
 3. Search by name
 4. calculate GPA of a student
 5. Exit
3. **Generators:**
 Create a class NumberSeries that includes a generator method to generate a series of even numbers up to a given limit. Write a script to create an instance of this class and print all the even numbers generated.

Decorators:

Create a decorator `log_method_call` that logs the method name and arguments whenever a method is called. Apply this decorator to methods in the Student class to log the calls to display student information and calculate GPA. Write a script to test this functionality by creating instances of the Student class and calling these methods.

Session 10 & 11:

- Operations Exception
- Exception Handling
- Except Clause
- Try-Finally Clause

Suggested Teaching Guidelines for
Python & R Programming PG-DBDA
August 2024

- User Defined Exceptions
- Logging in Python

Session 12, 13 & 14:

- Working with Pandas
- Data Wrangling with Pandas
- Working with NumPy
- Data Cleaning with Python

Lab Assignments:

1. Write a NumPy program to read a CSV data file and store records in an array.
2. Write a NumPy program to convert a PIL Image into a NumPy array. Also convert a NumPy array to an image. Display the image.

Sample Output:

```
[[[255 255 255 0]]
```

.....

```
[255 255 255 0]]]
```

3. Write a NumPy program to convert Pandas dataframe to Numpy array with headers.
4. Write a Pandas program to read a dataset from diamonds
 - DataFrame and modify the default columns values and print the first 6 rows
 - calculate the mean of each numeric column of diamonds DataFrame.
 - calculate count, minimum, maximum price for each cut of diamonds DataFrame
 - print a concise summary of diamonds DataFrame.
 - count the duplicate rows of diamonds DataFrame.

Session 15 & 16:

- Working with beautiful soup
- Visualizing Using Matplotlib, Seaborn
- Working with ggplot, Plotly

Lab Assignments:

1. Extract any website data using Beautiful Soup.
2. Import csv file using pandas, find correlation and plot heatmap of correlation using seaborn, plot the scatter plot for any two highest correlated columns using matplotlib and plotly.

Session 17:

- Load Images using Pillow
- Load Audio Files using Scikit-learn (scipy.io)
- Creation of Python Virtual Environment

Suggested Teaching Guidelines for
Python & R Programming PG-DBDA
August 2024

Session 18:

- Connecting Databases with Python
- Working with Databases using Python
- Accessing and Manipulating Databases

R-Programming: (16 Hours = 8T + 8L)**Session 19:**

- The R project for Statistical Computing
- Why R
- Introduction & Installation of R
- R Basics, Finding Help,
- Code Editors for R,
- Exploring R Gui
- Exploring RStudio
- Basic Mathematical & Arithmetic operations in R

Lab Assignments:

1. Read Introduction to R, R basics and Scope of R.
2. Explore R GUI and Rstudio

Session 20:

- Data Objects- Data Types & Data Structures (e.g. lists, Arrays, matrices, data frames)
- Packages in R
- Working with Packages
- Handling Data in R Workspace
- Reading & Importing data from Text files, Excel files, Multiple databases
- Exporting Data from R

Lab Assignments:

1. Write R program to compute sum, mean and product of a given vector elements.
2. Write R program to call the (built-in) dataset airquality.
3. Write R program to create a list of dataframes and access each of those data frames from the list.
4. Write R Program to create dataframe,
 - Get the statistical Summary and nature of the data of dataframe
 - Add new column in dataframe.
 - Sort dataframe using multiple columns.
 - Export Dataframe to excel file using writexl package.

Session 21:

- Introduction to tidyverse (group of packages)
- Manipulating and Processing Data in R
- Creating, Accessing and Sorting data frames
- Extracting, Combining, Merging, reshaping data frames

Suggested Teaching Guidelines for
Python & R Programming PG-DBDA
August 2024

Lab Assignments:

1. Load one XML file and one Json file in R studio, Print the data in both files one by one and get statistical summary of data.
2. Extract Data from any website using R packages.
3. Call the built in dataset “diamonds” in R and plot Pie chart and Bar graph

Session 22:

- Functions
- Built in functions in R (numeric, character, statistical)
- Interactive reporting with R markdown
- Case study

Lab Assignments:

1. Examine the built in ChickWeight data (the help gives background about the data). The function split will prove useful to do the following (as will a script).
 - Construct a plot of weight against time for chick number 34.
 - For chicks in diet group 4, display box plots for each time point.
 - Compute the mean weight for chicks in group 4, for each time point. Plot this mean value against time.
 - Repeat the previous computation for group 2. Add the mean for group 2 to the existing plot.
 - Add a legend and a title.
 - Copy and paste the graph into Word.

*Suggested Teaching Guidelines for***Java Programming PG-DBDA August 2024****Duration:** 34 class room hours + 36 Lab hours**Objective:** To reinforce knowledge of Java Programming**Prerequisites:** Knowledge of Linux command, OOPS concepts and any programming language**Evaluation method:** Theory exam – 40% weightage
 Lab exam – 40% weightage
 Internal exam – 20% weightage**List of Books / Other training material****Text Book :**

Java for Dummies, Barry Burd, Wiley India, Seventh Edition.

Reference:

1. Java Server Programming (J2EE 1.7 Edition) Black Book by Dreamtech Software Team
2. Core Java : Fundamentals - Volume 1 Gary Cornell, Cay S. Horstmann/ Pearson
3. Advanced Java programming by Uttam K Roy / Oxford University press

Note: Each session having 2 Hours**Session 1 & 2:****Lecture**

- Java 8 Basics : Overview of Java, Features of Java, Scope of variables
- Object Oriented Concepts
- Java Virtual Machine – Overview
- JDK and its usage (Java Compiler, Java Runtime, Java Debugger, Java doc)
- Working with Data Types: Structure of a Java Class, Importing Packages, Difference between object reference variables and primitive variables, how to read or write to object fields)

Session 3:**Lecture**

- Object's lifecycle (creation, reassignment, garbage collection: new, finalize)
- Wrapper classes (Boolean, Double and Integer)
- Operators (Unary, Binary, Arithmetic, Assignment, Compound, Relational, Logical, Equality) and Control Statements (if, if-else, for, while, switch, do-while, break and continue, ternary constructs)

Assignment – Lab:

1. Create Java Program for simple calculator, compile & test it.

Suggested Teaching Guidelines for
Java Programming PG-DBDA August 2024

Session 4:**Lecture**

- Packages and classpath
- Arrays
- Understanding of String Class, StringBuilder Class, StringBuffer class
- Methods and Encapsulation: Methods, Access Modifiers, Method Overloading, Passing Data, Creating Constructors, Immutable Classes

Assignment – Lab:

Get yourself acquainted with java environment. Build a class Emp, which contains details about the employee and compile and run its instance

Session 5:**Lecture**

- Class Inheritance, Abstract Classes, Inner Classes, Interface and Implementation classes.
- Understanding Polymorphism: Object vs Reference, Object Casting, Virtual Methods, Method Overriding

Assignment – Lab:

Create an inner class for a manager, which contains information about the manager. Use the appropriate interfaces. Create an anonymous inner class for Tech. Members using the Session one assignment

Session 6:**Lecture**

- Exception-Handling: Basics, Role of Exceptions, Types
- Using try and catch, Multiple Catch, Nested try (throw, throws, finally)
- Built-in Exceptions, Runtime Exceptions Checked Exceptions, Errors
- Creating own Exception Subclasses

Assignment – Lab:

Create a user defined exception to check whether your employee exist in your data structure and using the catch and finally block. Redeem an appropriate solution

Session 7:**Lecture**

- Enumerations, Auto boxing, and Annotations

Assignment – Lab:

Create sample classes to understand boxing & unboxing. Use different methods of java defined wrapper classes

Session 8:**Lecture**

- java.util, java.lang

Suggested Teaching Guidelines for
Java Programming PG-DBDA August 2024

Assignment – Lab:

Create an appropriate data structures to store your employee object and use the java.util.package properties.

Session 9 & 10:**Lecture**

- Generics and Collections

Assignment – Lab:

1. Implement String class and util package
2. Using the collection framework define an appropriate interface to your above application

Session 11:**Lecture**

- Functional Programming Overview
- Functional Interfaces
- Explore java.util.function package : Predicate, Map, Consumer, Supplier
- Lambda Expressions
- Impact of Functional programming upon Collection Framework

Session 12:**Lecture**

- Introduction to Streams
- Streams vs. Collections
- java.util.stream.Stream API
- Types of Primitive Streams : IntStream,LongStream,DoubleStream & its API
- Different operations on streams : filter, map, reduce, sort, flatMap, anyMatch, count, boxing.
- Overview of Java 8/17 Date Time API

Assignment – Lab:

1. Process bank accounts collection using stream functions.

Session 13 & 14:**Lecture**

- Java Concurrency: Using threads in Java, Life cycle of thread
- Advantages and issues
- Thread class, thread groups
- The Runnable interface
- Synchronization

Suggested Teaching Guidelines for
Java Programming PG-DBDA August 2024

Session: 15:**Lecture**

- The java.io Package
- Files
- Byte Streams and Unicode Character Streams
- Persistence of objects
- Object Serialization Methods

Assignment – Lab:

- Make your above Employee, manager classes object persistent.

Session: 16:**Lecture: Reflection in Java & JVM Architecture**

- Why Java Reflection
- Basic Reflection API for finding out details of the class name, super classes & interfaces.

Assignment – Lab:

- Use the Java Reflection API to print the name of a given class, its superclass, and all implemented interfaces.
- Create a Java program that lists all the methods declared in an interface implemented by a class using reflection.

Session 17:**Lecture**

- Introduction of JDBC API
- JDBC Architecture
- JDBC Drivers
- Drivers, Connection, Statement, Prepared Statement and Result Set interfaces and their relationship to provider implementations

Assignment – Lab:

- Build an application to get student's details using database.

Suggested Teaching Guidelines for
Advanced Analytics using Statistics
PG-DBDA August 2024

Duration: 46 hours classroom and 44 hours Lab

Objective: To perform advanced analytics using Python & R skills and important mathematical concepts.

Prerequisites: Good Knowledge of Basic Mathematics

Evaluation method: Theory exam– 40% weightage
 Lab exam – 40% weightage
 Internal exam– 20% weightage

List of Books / Other training materials

Text Book:

1. Business Analytics, James R Evans, Pearson Education, 3rd Edition.

Reference:

1. Beginning R – The Statistical Programming Language by Dr. Mark Gardener
PUB: WILEY
2. Art of Programming in R, by Norman Matloff
3. Statistics for Management by Levin
4. Business Analytics: Methods, Models, and Decisions by James R Evans
5. Introductory Statistics with R (Statistics and Computing) by Peter Dalgaard
6. R in a Nutshell by Joseph Adler (O'REILLY)
7. R Cookbook by Paul Teator (O'REILLY)
8. The R Book, Second Edition
9. Statistics Using R, Shailaja Deshmukh, Sudha Purohit, Sharad Gore, Pub:
Narosa

Note:

- Each session mentioned is for theory of 2 hours' duration. Lab assignments are indicatives; faculty needs to assign more assignments for better practice.
- Trainer has to teach the statistical and probability concepts involved here in detail

Session 1 & 2:

- Introduction to Analytics
- Data analytics Life Cycle:
- Discovery,
- Data preparation
- Model planning
- Model building implementation
- Quality assurance
- Documentation
- Management approval
- Installation

Suggested Teaching Guidelines for
Advanced Analytics using Statistics
PG-DBDA August 2024

- Acceptance and operation
- Intelligent data analysis

Assignment –Lab: Import csv file using R and perform ETL operation using dplyr package.

Session 3 & 4

- Sample Spaces and Events
- Joint, Conditional and Marginal Probability
- Bayes' Theorem

Assignment –Lab: Load any dataset, apply Bayes' Theorem and predict the output

Session 5 & 6:

- Random Variable
- Concepts of Correlation
- Covariance
- Outliers

Assignment –Lab: Load any dataset and find out the covariance between two fields and also find the correlation and determine how two fields are correlated. Also handle the outliers in the data.

Session 7 & 8:

- Probability Distribution and Data
 - Continuous distribution – (Uniform, Exponential & Normal)
 - Discrete distribution – (Binomial, Poisson & Geometric distribution)

Assignment –Lab: generate random numbers and check if they are in normal distribution using scipy libraries.

Session 9 & 10:

- Descriptive Statistical Measures
- Summary Statistics - Central Tendency & Dispersion (Mean, Median, Mode, Quartiles, Percentiles, Range, Interquartile Range, Standard Deviation, Variance, and Coefficient of Variation)

Assignment –Lab: Load any dataset and find out the mean, median mode and other central tendencies of the dataset.

Session 11 & 12:

- Sampling and Estimation
- Sample & population, Uni-variate and bi-variate sampling, re-sampling
- Central Limit Theorem

Assignment –Lab: Load any dataset and Explore sampling techniques.

Suggested Teaching Guidelines for
Advanced Analytics using Statistics
PG-DBDA August 2024

Session 13 & 14:

- Statistical Inference Terminology (types of errors, tails of test, confidence intervals etc.)
- Hypothesis Testing
- Parametric Tests: ANOVA, t-test
- Non-parametric Tests- chi-Square, U-Test

Assignment –Lab: Load any dataset and Perform the hypothesis testing on correlated variables.

Session 15 & 16:

- Predictive Modelling (From Correlation to Supervised Segmentation):
 - Identifying Informative Attributes,
 - Segmenting Data by Progressive Attributive,
 - Models,
 - Induction and Prediction,
 - Supervised Segmentation,
 - Visualizing Segmentations,
 - Trees as Set of Rules,
 - Probability Estimation;

Assignment –Lab: Explore predictive modelling techniques.

Session 17:

- Simulation and Risk Analysis
- Optimization, Linear

Assignment –Lab: Explore Monte Carlo simulation.

Session 18 & 19:

- Decision Analytics:
 - Evaluating Classifiers,
 - Analytical Framework,
 - Evaluation,
 - Baseline,
 - Performance and Implications for Investments in Data;

Session 20 & 21:

- Evidence and Probabilities:
 - Explicit Evidence Combination with Bayes Rule,
 - Probabilistic Reasoning;

Session 22:

- Business Strategy:
 - Achieving Competitive Advantages,
 - Sustaining Competitive Advantages

Suggested Teaching Guidelines for
Advanced Analytics using Statistics
PG-DBDA August 2024

Session 23:

- Factor Analysis,
- Directional Data Analytics,

Assignment –Lab: Download dataset and perform factor analysis on it.

Duration: 44 classroom hours + 46 Lab hours

Objective: To reinforce knowledge of RDBMS and facilitate hands-on experience on SQL & NoSQL.

Prerequisites: Knowledge of Object-Oriented concepts.

Evaluation method: Theory exam– 40% weightage
 Lab exam – 40% weightage
 Internal exam– 20% weightage

List of Books / Other training material

Text Book:

1. Textbook of RDBMS, Vidya H Bankar, Techtree Educations, 1st edition.

Reference:

1. MongoDB in Action by DreamTech
2. MongoDB - The definitive guide - by Oreilly
3. The Definitive Guide –MongoDB by Kristina Chodorow
4. MongoDB Aggregation Framework Principles and Examples by John Lynn
5. Getting Started with NoSQL by Gaurav Vaish
6. Database System Concept by Henry Korth, S. Sudarshan & Abraham Silberschatz
7. Relational Database Design and Implementation: Clearly Explained, Third Edition
8. Beginning Database Design Solutions
9. Database Modeling and Design: Logical Design, Fifth Edition
10. Introduction to Database Management System

Note: Each session mentioned is for theory of 2 hours' duration. Lab assignments are indicatives; faculty needs to assign more assignments for better practice.

Session 1:

Lecture

- Database Concepts (File System and DBMS)
 - What is file system, its need?
 - What is DBMS, its need
 - Codd's 12 rules for RDBMS

Lab Assignment:

- Read and understand the concepts of File System, DBMS & RDBMS.

Session 2:**Lecture**

- Database Storage Structure
 - Table Space
 - Control File
 - Data file
- Structured and Unstructured Data
- Introduction to Data Collection like what is data collection.
- The tools and how data can be gathered in a systematic fashion

Lab Assignment:

- Read and understand the related chapters.

Session 3:**Lecture**

- Introduction to SQL
- DDL Commands
- DML & DCL Commands

Lab Assignment:

- DDL Commands: Create/Alter/Drop/Grant/Revoke
- DML Commands: Select/Insert/Update/Delete/Truncate
- DCL Commands: RollBack Commit
- Create new User named 'dbda' ,Grant all the privileges and Perform followingQueries.
- Create Table 'Books' using proper data types which contain columns(name, author, price, writer)

Session 4:**Lecture**

- Grouping Things Together (Group by, Having)
- Sorting Data (Order By)
- Advance Subqueries (Correlated Sub query, Outer Joins)

Lab Assignment:

- Queries containing Group By, Having Clause,
- Order by
- Correlated Queries, SubQueries, Outer Joins
- Find out number of employees in each department using employee table and department table
- Print the employee names who have 'A' as first letter and 'N' as last letter in theirname.
- using customers and product table, write sql query to find the salespersons and customers he handles, print customer name, city, salesman, commission.

Session 5 & 6:**Lecture**

- Constructs in SQL
- Data collection
- Designing Database Schema
- Normal Forms and ER Diagram
- Relational DB modelling
- Stored Procedures
- Gathering Data in Systematic fashion

Session 7:**Lecture**

- Views
- Triggers
- Window Function
- Case statement

Lab Assignment:

- Read and understand the related chapters.
- Create View to find employee Who have highest salary, Print name, salary, department number and department name.
- Create View to fine salesperson who handles a customers who make highest number of orders, return order date, salesperson ID, name.

Session 8 & 9:**Lecture**

- Data Ware Housing Concepts and Introduction to Tools
- Tools related to Data Warehousing
- Different algorithms related to Data Warehouse
- Importance and its Applications

Lab Assignment:

- Read and understand the related chapters.

Session 10:**Lecture**

- NOSQL
 - Introduction to NoSQL
 - Difference between a RDBMS and a NoSQL database
 - Understanding the Storage Architecture
 - Working with Column-Oriented Databases
 - Document Store Internals

Lab Assignment:

- Read and understand the related chapters.

Session 11:**Lecture**

- Practical Design of NoSQL
- NOSQL
 - Schema structure for Oracle NoSQL database
 - Changing Document Databases
 - Schema Evolution in Column-Oriented Databases
 - Data Evolution in Key/Value Stores

Lab Assignment:

- Practice Questions including Column-Oriented Databases

Session 12:**Lecture**

- Introduction to MongoDB (NoSQL)
 - Performing CRUD Operations
 - Creating Records
 - Accessing Data
 - Updating and Deleting Data
 - Working with Language Bindings
 - Querying NoSQL Stores
 - Similarities Between SQL and MongoDB Query Features
 - Accessing Data from Column-Oriented Databases Like HBase
 - Querying Redis Data Stores

Lab Assignment:

- Read and apply CRUD Operations.

Session 13 & 14:**Lecture**

- Introduction to MongoDB
 - What are MongoDB Internals
 - Essential Concepts behind a Database Index
 - Indexing and Ordering in MongoDB
 - Creating and Using Indexes in MongoDB

Lab Assignment:

- Practice to create and using Indexes in MongoDB

Session 15:**Lecture**

- MongoDB Queries
 - Create Operations
 - Read Operations
 - Data Aggregation Operations
 - Update Operations

Lab Assignment:

- Insert, Find, FindOne, logical Operators, Distinct, Group, Upsert, Update, Remove.
- Create database using MongoDB query.
- Create table books using MongoDB query.
- Write a MongoDB query to find the restaurants who achieved a score more than 90 using Restaurants collection.
- Write a MongoDB query to find the restaurants which do not prepare any cuisine of 'American' and achieved a score more than 70 and located in the longitude less than - 65.754168 using Restaurants collection.
- Write a MongoDB query to find the restaurants which do not prepare any cuisine of 'American ' and achieved a grade point 'A' not belongs to the borough Brooklyn. The document must be displayed according to the cuisine in descending order

Session 16 & 17:**Lecture**

- Data Model XML
- Querying and transformation
- Tools - OLTP and OLAP

Lab Assignment:

- Read and understand the related chapters

Session 18:**Lecture**

- Introduction to Cassandra
- Comparison between Cassandra and MongoDB
- Architecture
- Cqlsh
- Shell Commands

Lab Assignment:

- Read and understand the related chapter

Session 19:**Lecture**

- Table Operation (Create, Alter, Drop, Truncate, Index creation, Index deletion, Batch)

Lab Assignment:

- Read and understand the related chapters

Session 20 & 21:**Lecture**

- CRUD Operation
 - Create
 - Update
 - Read
 - Delete
- CQL Types
 - CQL Datatypes
 - CQL Collections
 - User Defined Datatypes

Lab Assignment:

- Read and understand the related chapters
- Create Table employees using CQL commands.
- Update employee's total salary to 20000 whose commission is '0'
- Create following tables using collections in Cassandra.
 1. Teachers and subjects.
 2. Books and Authors.
- Insert a value in employee table, update salary of employee whose id is 03 and change the names of employees into upper case whose name start with 'N'.(Perform all operations in single Query using Batch)
- Print all values present in Books table.

Session 22:**Lecture**

- Data Driven Decisions
- Enterprise Data Management
 - Data Preparation
 - Data Cleaning

Lab Assignment:

- Read and understand the related chapter

*Suggested Teaching Guidelines for***Big Data Technologies PG-DBDA August
2024**

Duration: 66 Classroom hours + 84 Lab hours

Objective: To reinforce knowledge of BigData Technologies such as Hadoop, Map reduce, HBase, PIG, Spark (PySpark)

Prerequisites: Knowledge of Linux command, SQL and Core Java

Evaluation method: Theory exam– 40% weightage
 Lab exam – 40% weightage
 Internal exam– 20% weightage

List of Books / Other training material**Textbook:**

1. Big Data, Black Book: Covers Hadoop 2, MapReduce, Hive, YARN, Pig, R and Data Visualization, DT Editorial Services , Wiley India,Latest.

Reference:

1. Big Data, Black Book by DreamTech
2. Programming Hive by O'Reilly (Author:- Edward Capriolo, Dean Wampler, and Jason Rutherglen)
3. Hadoop The Definitive Guide 4th Edition by O'Reilly (Author: - Tom White)
4. Hadoop with python
5. Hadoop Real-World Solutions Cookbook by Packt publication (Author: Jonathan R. Owens, Jon Lentz,Brian Femia)
6. Data Architecture: A Primer for the Data Scientist: Big Data, Data Warehouse and Data Vault
7. Big Data Analytics with Spark: A Practitioner's Guide to Using Spark for Large-Scale Data Processing, Machine Learning, and Graph Analytics, and High-Velocity Data Stream Processing

Session: 1, 2 & 3**Introduction to Big Data**

- Big Data - Beyond the Hype
- Big Data Skills and Sources of Big Data
- Big Data Adoption
- Research and Changing Nature of Data Repositories
- Data Sharing and Reuse Practices and Their Implications for Repository DataCuration
- Overlooked and Overrated Data Sharing
- Data Curation Services in Action
- Open Exit: Reaching the End of The Data Life Cycle
- The Current State of Meta-Repositories for Data
- Curation of Scientific Data at Risk of Loss: Data Rescue And Dissemination

Suggested Teaching Guidelines for

***Big Data Technologies PG-DBDA August
2024***

Introduction to Hadoop

- A Brief History of Hadoop
- Evolution of Hadoop
- Introduction to Hadoop and its components
- Comparison with Other Systems
- Hadoop Releases
- Hadoop Distributions and Vendors

Hadoop Distributed File System (HDFS)

Session: 4 & 5

Hadoop Distributed File System (HDFS)

- Distributed File System
- What is HDFS
- Where does HDFS fit in
- Core components of HDFS
- HDFS Daemons
- Hadoop Server Roles: Name Node, Secondary Name Node, and Data Node

HDFS Architecture

- HDFS Architecture
- Scaling and Rebalancing
- Replication
- Rack Awareness
- Data Pipelining,
- Node Failure Management.
- HDFS High Availability NameNode

Lab-Assignment:

- Run the HDFS commands, and add a one liner understanding for each of the command.
- Execute the provided code using HDFS, step run and understand

Session: 6

Getting Started: Hadoop Installation

- Hadoop Operation modes
- Setting up a Hadoop Cluster
- Cluster specification
- Single and Multi-Node Cluster Setup on Virtual & Physical Machines,
- Remote Login using Putty/Mac Terminal/Ubuntu Terminal.
- Hadoop Configuration, Security in Hadoop, Administering Hadoop,
- HDFS – Monitoring & Maintenance, Hadoop benchmarks,
- Hadoop in the cloud.

Session: 7

Hadoop Architecture

- Hadoop Architecture,
- Core components of Hadoop,
- Common Hadoop Shell commands.

Suggested Teaching Guidelines for

***Big Data Technologies PG-DBDA August
2024***

Session: 8**HDFS Data Storage Process**

- HDFS Data storage process,
- Anatomy of writing and reading file in HDFS,
- Handling Read/Write failures
- HDFS user and admin commands,
- HDFS Web Interface.

Session: 9**Getting in touch with Map Reduce Framework**

- Hadoop Map Reduce paradigm,
- Map and Reduce tasks,
- Map Reduce Execution Framework,
- Map Reduce Daemons
- Anatomy of a Map Reduce Job run

More Map Reduce Concepts

- Partitioners and Combiners,
- Input Formats (Input Splits and Records, Text Input, Binary Input, MultipleInputs),
- Output Formats (Text Output, Binary Output, Multiple Output).
- Distributed Cache

Session: 10**Basics of Map Reduce Programming**

- Hadoop Data Types,
- Java and Map Reduce,
- Map Reduce program structure,
- Map-only program, Reduce-only program,
- Use of combiner and partitioner,
- Counters, Schedulers (Job Scheduling),
- Custom Writables, Compression

Lab-Assignment:

- Execute the train data example.
- Execute the train data example using chained methods.

Session: 11**Map Reduce Streaming**

- Complex Map Reduce programming,
- Map Reduce streaming,
- Python and Map Reduce,
- Map Reduce on image dataset

Hadoop ETL**Session: 12**

- Hadoop ETL Development,
- ETL Process in Hadoop,
- Discussion of ETL functions,

*Suggested Teaching Guidelines for***Big Data Technologies PG-DBDA August
2024**

- Data Extractions,
- Need of ETL tools,
- Advantages of ETL tools.

Lab-Assignment:

- Understand the file formats and read the provided links

Session: 13**Introduction to HBase**

- Overview of HBase
- HBase architecture
- Installation

Session: 14 & 15**The HBaseAdmin and HBase Security**

- Various Operations on Tables
- HBase general command and shell,
- java client API for HBase
- Admin API
- CRUD operations
- Client API
- HBase – Scan, Count and Truncate
- HBase Security

Lab-Assignment:

- Run the Hbase shell commands
- Run the HBase using Java client

Session: 16**The Hive Data-ware House**

- Introduction to Hive,
- Hive architecture and Installation,
- Comparison with Traditional Database,
- Basics of Hive Query Language.

Session: 17**Working with Hive QL**

- Datatypes,
- Operators and Functions,
- Hive Tables (Managed Tables and Extended Tables),
- Partitions and Buckets,
- Storage Formats,
- Importing data,
- Altering and Dropping Tables

Suggested Teaching Guidelines for

***Big Data Technologies PG-DBDA August
2024***

Lab-Assessment:

- Create a hive DB and table (internal and external)
- Load the data into hive table (using local inpath and HSFS inpath)

Session: 18

Querying with Hive QL

- Querying Data-Sorting,
- Aggregating,
- Map Reduce Scripts,
- Joins and Sub queries,
- Views,
- Map and Reduce side joins to optimize query.

Lab-Assessment:

- Run all the types of joins in Hive
- Execute the data to be partitioned

Session: 19

More on Hive QL

- Data manipulation with Hive,
- UDFs,
- Appending data into existing Hive table,
- custom map/reduce in Hive
- Writing HQL scripts

Session: 20, 21 & 22

- Introduction to Data Warehousing and Data Lakes
- Designing Data warehousing for an ETL Data Pipeline
- Designing Data Lakes for an ETL Data Pipeline
- ETL vs ELT
- Fundamentals of Airflow/Informatica
- Work management with Airflow/ Informatica
- Automating an entire Data Pipeline with Airflow/Informatica

Lab-Assessment:

- Create an airflow DAG/ Informatica for Extract -> Transform -> Load

Session: 23, 24 & 25

Apache Spark APIs for large-scale data processing

- Overview, Linking with Spark, Initializing Spark,
- Resilient Distributed Datasets (RDDs), External Datasets
- RDD v/s Data frames v/s Datasets
- Data frame operations
- Structured Spark Streaming
- Passing Functions to Spark, Working with Key-Value Pairs, Shuffle operations,
- RDD Persistence, Removing Data, Shared Variables, Deploying to a Cluster

*Suggested Teaching Guidelines for***Big Data Technologies PG-DBDA August
2024****Lab Assignment:**

- Run the provided Hadoop Streaming program using python

Session: 26

- Map Reduce with Spark
- Working with Spark with Hadoop
- Working with Spark without Hadoop and their Differences

Lab Assignment

- Execute all the provided code using step-runs for each and every codeline
- Setup the JDBC configuration and run the Spark JDBC Connectivity program
- Run the spark integrations using the provided code

Session: 27

- Data preprocessing
- EDA

Session: 28 & 29

- Introduction to Kafka
- Working with Kafka using Spark
- Spark streaming Architecture
- Spark Streaming APIs
- Building Stream Processing Application with Spark

Lab Assignment

- Execute the spark streaming with Kafka

Session: 30

- Setting up Kafka Producer and Consumer
- Kafka Connect API

Session: 31

- Spark SQL

Lab Assignment

- Run the sparkSQL programs using step-runs for each and every codeline
- Run all the SparkSQL programs
- Analyse the election data using spark and provide analysis

Session: 32 & 33

- Spark MLlib
- Predictive Analysis

*Suggested Teaching Guidelines for***Big Data Technologies PG-DBDA August
2024****Lab Assignment:**

- Deep Learning with Spark
- Connecting DB's with Spark
- Accessing and manipulating the DB's
- Demo: Capstone Project
- Create a complex workflow using bash operator, a simple workflow using python
- Create Using python airflow operator to read data from your local drive, ingest the data into your HDFS, and perform a spark WC

Suggested Teaching Guidelines for
Data Visualization - Analysis and Reporting
PG-DBDA August 2024

Duration: 26 Classroom hours + 24 Lab hours

Objective: To introduce students in Data Analytics, Visualization and Reporting

Prerequisites: Knowledge of Database Fundamentals and Big Data Technologies.

Evaluation method: Theory exam— 40% weightage
 Lab exam – 40% weightage
 Internal exam— 20% weightage

List of Books / Other training material

Text Book:

1. Communicating Data with Tableau, Ben Jones, O'Reilly, Shroff Publishers & Distributors, Tableau 8.1.

Reference Book:

1. Mastering Microsoft Power BI: Expert Techniques for Effective Data Analytics and Business Intelligence Book by Brett Powell
2. Designing Data Visualizations, by Steele,O'Reilly
3. Tableau your data, by Daniel G/ Wiley
4. Graphs Cookbook, Hrishi V. Mittal, Packt Publishing
5. Python Data Visualization Cookbook,Igor Milovanović, Packt Publishing
6. Learning Python Data Visualization, Chad Adams, Packt Publishing
7. Data Visualization with D3.js Cookbook,Nick Qui Zhu,Packt Publishing
8. Getting Started with D3, Mike Dewar,O'Reilly
9. Data Visualization with JavaScript
10. Data Visualization for Dummies
11. High Impact Data Visualization with Power View, Power Map, and Power BI
12. The Visual Organization: Data Visualization, Big Data, and the Quest for Better Decisions
13. Mastering Tableau 2021:- by Marleen Meier

Note:

- **Tool to be use: Tableau**

Session 1 & 2:

- Business Intelligence basic,
- Information gathering,
- Decision making,
- Managing BI,

Suggested Teaching Guidelines for
Data Visualization - Analysis and Reporting
PG-DBDA August 2024

- BI User Segmentation,
- Gathering BI Requirements,
- Content and Knowledge Management,
- Strategic Approach to BI
- Significance of visual analytics Information Visualization
- Data Representation
- Data collection and binding
 - Structured Data
 - Unstructured data

Session 3, 4 & 5:**MS EXCEL**

- Functions
- Formula
- Charts
- Pivots and Lookups
- Data Analysis Tool pack
 - 1. Descriptive Summaries
 - 2. Correlation
 - 3. Regression

Session 6**Data analytics Life Cycle:**

- Discovery,
- Data preparation
- Model planning
- Model building implementation
- Quality assurance
- Documentation
- Management approval
- Installation
- Acceptance and operation

Session 7 & 8

- Introduction to Tableau
- Intelligent data analysis
- Nature of Data
- Analytics Processes and tools
- Analysis vs. Reporting
- Modern Data Analytic Tools
- Data sources in Tableau

Suggested Teaching Guidelines for
Data Visualization - Analysis and Reporting
PG-DBDA August 2024

Session 9, 10 & 11

- Visualization Algorithms
- Visual Encodings
 - color, size, shape, lines, axes, scaling, annotation
- Taxonomy of data visualization (Some Types of charts, but not limited to)
 - Comparison charts – types of Bar chart, Box plots, Histograms, Gantt charts, Bullet graphs, side-by-side bar chart etc.
 - Tables – Text Tables, Highlight tables
 - Hierarchies and relationships – Pie chart, stacked bar, Tree map etc.
 - Changes over time – Line chart, dual lines, Area charts etc.
- Connections and relationships – scatter plots, Symbol maps, map, heat maps, packed bubble chart etc.

Session 12 & 13:

- Choosing appropriate visuals
- Applying calculations using functions, statistics
 - Numeric Calculations
 - String Calculations
 - Date calculations
 - LOD (Level of Detail) Expressions
- Data sorting, filters
- Interactive visualization
 - Event listeners/callbacks
 - Data updation
 - Visual updation
- Dashboard Design

Assignment-Lab:

- Load coffee chain dataset in Tableau and create required visuals. Also create the report for the same dataset using VBA tools in Excel.

Suggested Teaching Guidelines for
Practical Machine Learning PG-DBDA
August 2024

Duration: 60 hours Theory and 80 hours Lab

Objective: Practicing Machine Learning Algorithms

Prerequisites: Good knowledge of Python Programming and Statistics

Evaluation method: Theory exam– 40% weightage
 Lab exam – 40% weightage
 Internal exam– 20% weightage

List of Books / Other training material

Textbook:

1. Machine Learning using Python , Manaranjan Pradhan , U Dinesh Kumar Wiley India, 1st Edition, 9788126579907

Reference Book:

Note:

- PyTorch Framework should be taught in Lab Hours

Note: Each session having 2 Hours

Session 1, 2 & 3:

- Fundamentals of information theory
- What is machine learning?
- Algorithm types of Machine learning
- Probably Approximately Correct (PAC) Learning
- Uses of Machine learning
- Evaluating ML techniques
- Bias complexity trade off
- Vapnik-Chervonenkis (VC) Dimension
- Non-uniform learnability (Structural risk minimization, Occam's Razor and No Free Lunch Theorem)
- Regularization and Stability
- Model Selection and Validation
- Introduction to Scikit Learn
- Performing ML using Scikit Learn

Assignments:

- Explore scikit learn Library.
- Explore Datasets Online (can refer Kaggle, UCI ML, etc.)
 - Load dataset in google colab.
 - Print first five values and last five values in dataset.
 - check correlation between fields present in dataset

Suggested Teaching Guidelines for
Practical Machine Learning PG-DBDA
August 2024

Session 4 & 5:

- Clustering
- Hierarchical Clustering & K means
- Distance Measure and Data Preparation – Scaling & Weighting
- Evaluation and Profiling of Clusters
- Hierarchical Clustering
- Clustering Case Study
- Principal Component analysis: PCA, Kernel PCA
- Random Projections

Assignments:

- Download “mall_customers.csv” dataset from Kaggle.
- (a) Form n no. of clusters according to your observation.
 - (b) Get wss value for each cluster.
 - (c) find best K value

Session 6, 7 & 8:

- Evaluation metrics: Accuracy, F1-score, ROC AUC, Log Loss
- Evaluation metrics: MAE, RMSE, R2 score
- Decision Trees
- Classification and Regression Trees
- Random forest, Gradient boosting Machines, Model Stacking
- CAT Boost & Light GBM
- XG Boost

Assignments:

- Implement Random Forest, SVM, Logistic regression classification algorithm and check for classification report, f1 score for all three algorithms.

Session 8 & 9:

- Bayesian analysis and Naïve bayes classifier
- Assigning probabilities and calculating results
- Linear Discriminant Analysis
- K-Nearest Neighbors Algorithm

Assignments:

- Implement K-Nearest Neighbors Algorithm

Session 10 & 11:

- Linear Regression
- Logistic Regression
- Polynomial Regression
- Ridge Regression
- Lasso Regression
- Elastic Net Regression

Suggested Teaching Guidelines for
Practical Machine Learning PG-DBDA
August 2024

Assignments:

- Download Dataset, perform linear, Ridge, Lasso, Polynomial regression and check for MAE, MSE, RMSE and also check F1 score and explain with conclusion.

Session 12:

- Support Vector Machines
- Basic classification principle of SVM
- Linear and Nonlinear classification (Polynomial and Radial)

Assignments:

- Download Air Quality Dataset from Kaggle Predict Air Quality Index using Linear regression and classify it into five categories using SVM (i.e. Very good, good, moderate, poor, worst)

Session 13 & 14:

- Moving average, Exponential Smoothing, Holt's Trend Methods, Holt-Winters'
- Methods for seasonality
- Autocorrelation (ACF & PACF), Auto-regression, Auto-regressive Models, Moving Average Models
- ARMA & ARIMA

Assignments:

- What is Auto correlation, explain its purpose Also download one data set and calculate Auto correlation.
- Explain ARMA and ARIMA model, what is purpose of this models in time series and Explain difference between them.

Session 15 & 16:

- Recommendation Systems
 - Data Collection & Storage, Data Filtering
 - Collaborative Filtering
 - Factorization Methods
 - Evaluation Metrics: Recall, Precision, RMSE, Mean Reciprocal Rank, MAP at K, NDCG

Session 17, 18 & 19:

- Introduction to Deep Learning
- Introduction to Tensor flow and Keras
- Introduction to Auto-encoders
- Neural Network and its applications
- Single layer neural Network
- Activation Functions: Sigmoid, Hyperbolic Tangent, ReLu
- Overview of Back propagation of errors

Suggested Teaching Guidelines for
Practical Machine Learning PG-DBDA
August 2024

Assignments:

- Explore Tensor Flow and Keras Libraries.
- Implement Different Activation functions on datasets in Jupyter Notebook.

Session 20, 21 & 22:**Deep Learning Essentials**

- Early Stopping for Preventing Overfitting
 - Dropout
 - L1 and L2 Regularization
- Update of weights with single training set element, Batch Training, Mini-batch Training, Stochastic Gradient Descent
- Training Methods for Neural Network (High-Level Overviews only) / Optimizers
 - Classic Backpropagation
 - Momentum Backpropagation
 - ADAM

Assignments:

- Implement Different optimizers
- Implement Gradient Problems

Session 23 & 24**Convolutional Neural Network using PyTorch**

- Introduction to PyTorch Framework
- Pytorch vs Tensor flow
- Convolutional Concept
- Transfer Learning and Inception Network as one of its examples
- Data Augmentation
- Object Detection
- YOLO Algorithm (High-Level Overview)

Assignments:

- Install PyTorch. Explore the documentation of PyTorch Library.
- Implement YOLO Algorithm.

Session 25 & 26**Recurrent Neural Network (RNN) using Pytorch**

- RNN Concept
- Types of RNNs
- Vanishing gradients with RNNs
- Gated Recurrent Unit (GRU) - (High-Level Overview only)
- Long Short-Term Memory (LSTM) - (High-Level Overview only)

Suggested Teaching Guidelines for
Practical Machine Learning PG-DBDA
August 2024

Assignments:

- Implement RNN using PyTorch
- Implement LSTM and GRU

Case Studies:

- Time series example with LSTM

Session 27**Generative AI****Introduction to Transformers**

- What are transformers and why are they important?
 - A brief history of NLP and the rise of transformers (including the Attention Timeline)
 - Transformers
 - Introduction to BERT (Bidirectional Encoder Representations from Transformers):
 - Pre-training objectives of BERT and its impact on NLP tasks
 - Fine-tuning BERT for various NLP applications (e.g., sentiment analysis, question answering)
 - Understanding and Handling Text Data:
 - Text pre-processing techniques (e.g., tokenization, stemming, lemmatization)
 - Representing text data for machine learning models
 - Hands-on activity: Pre-processing text data and fine-tuning a pre-trained model for a specific task
 - Applications of transformers across various domains (e.g., computer vision, speech recognition)

Assignment:

- Hands-on activity: Exploring a pre-trained transformer model (e.g., using Google AI's Colab)

Session 28

- Core Concepts of AI architectures
 - Understanding the Encoder-Decoder Architecture:
 - The role of encoders in processing input sequences
 - The role of decoders in generating output sequences
 - Encoder-decoder for tasks like machine translation and text summarization
 - Attention Mechanisms:
 - Demystifying attention and its variants (e.g., self-attention, masked attention)
 - How attention helps models "focus" on relevant parts of the input
 - Hands-on activity: Implementing a simple attention mechanism

Suggested Teaching Guidelines for
Practical Machine Learning PG-DBDA
August 2024

Session 29 & 30

- Advanced Concepts & Applications
 - Introduction to Large Language Models (LLMs)
 - Capabilities and limitations of LLMs
 - Responsible development and use of LLMs
 - Reward Models and Alignment Strategies
 - Ensuring alignment between LLM goals and human values
 - Techniques for mitigating bias and promoting safety
- Practical Case Studies:
 - Exploring real-world applications of transformers and LLMs (e.g., chatbots, text generation, code completion)
- Deployment Considerations for LLMs:
 - Infrastructure requirements and challenges
 - Strategies for efficient deployment of LLMs in production environments

Assignments:

- Exploring an LLM API or building a simple application using a pre-trained model

Duration: 90 hours

Objective: To reinforce knowledge of general Aptitude & English

Prerequisites: Knowledge of Mathematics & English.

Evaluation method: Theory exam– 80% weightage
Internal Assessment– 20% weightage

List of Books / Other training

materialReference:

1. Quicker math by M. Tyra (BSC publication co. Pvt. Ltd)
2. Quantitative Aptitude by RS Aggarwal
3. Verbal & Non- Verbal Reasoning: RS Aggarwal
4. Quantitative Aptitude - Quantum CAT: Sarvesh K Verma
5. High School English Grammar & Composition Revised Edition Wren, Martin / S. Chand Publisher
6. How to prepare GRE by Barron's / galgotia publications pvt. Ltd
7. Oxford Guide to English Grammar 01 Edition John Eastwood / Oxford University Press Website to refer: www.indiabix.com
8. Business Communication by H S Mukerjee / Oxford University Press
9. Business Communication by R K Madhukar / Vikas Publishing House Pvt. Ltd.
10. Business Communication Essentials A skills-Based Approach to Vital Business English by Courtland Bovee, John Thill / Pearson
11. Effective Business Communication by Asha Kaul / Prentice Hall of India
12. Fundamental of Technical Communications by Meenakshi Raman, Sangeeta Sharma /Oxford University Press
13. English is easy by Chetan Anand Singh/ BSC publication Co. Pvt. Ltd
14. Communication Skills Publication Year 2011 Sanjay Kumar, Pushp Lata / Oxford University Press
15. Professional Communication Skills Praveen S R Bhatia / S. Chand Publishing

Note:

- *Each session having 2 Hours*
- *General Aptitude is of 40 Hours*
- *Effective Communication is of 50 Hours*

Part I – Aptitude

Session 1:

- Analogy
- Series Completion (Number, Alphabet, Letter Series)
- Coding- Decoding for Number, Alphabet and Letter

Session 2:

- Blood Relations

Session 3:

- Puzzle Test
 - Classification Type questions
 - Compression Type questions
 - Sequential order questions
 - Section based on given conditions
 - Questions involving family members

Session 4:

- Alphabet test
- Order of words
- Letter words problems
 - Rule detection
 - Alphabetical quibble
 - Word formation
- Logical sequence of words

Session 5:

- Number, Ranking and time
- Sequence Test
- Mathematical operations
- Arithmetic reasoning

Session 6:

- Logical reasoning
- Statement-Arguments
- Statement-Assumptions

Session 7:

- Statement- courses of Action
- Statement- Conclusions
- Deriving conclusion from passages

Session 8:

- HCF and LCM
- Fraction

Session 9:

- Number system

Session 10:

- Permutation & combination

Session 11:

- Ratio & Preparation
- Partnership

Session 12:

- Average
- Percentage

Session 13:

- Clock

Session 14:

- Probability

Session 15:

- Pipes and cisterns
- Problem on streams

Session 16:

- Time and work
- Work and Wages

Session 17:

- Problem on Trains
- Problem on Speed and Velocity

Session 18:

- Problem on Ages

Session 19:

- Profit and loss

Session 20:

- Simple Interest,
- Compound Interest

Part II -Effective Communication

Session 1:

Fundamentals of
Communication The Art of
Communication

- Vocabulary, spelling and grammar
- Fluency, pronunciation, intonation and accent
- Idioms
- Synonyms & Antonyms

Practice Sessions:

Practice words, spelling, intonation and correct pronunciation Practice idioms, synonyms & antonyms

Session 2:

Personality Development

- Greeting
- Etiquettes
- Body language
- Developing positive attitude
- Confidence building
- Questioning techniques

Practice Sessions:

Practice greeting, etiquettes and questioning

Session 3

English Grammar

- Nouns
- Pronouns
- Adjectives
- Articles
- Verbs
- Adverbs
- Prepositions
- Conjunctions

Practice Sessions:

Practice sentence making

Session 4:

English Grammar

- Active and passive voices
- Direct and indirect speeches

Practice Sessions:

*Practice speaking in active &
passive voices Practice direct &
indirect speaking*

Session 5:

Correct usage of English
Common mistakes in English communication

Practice Sessions:

Practice general English communication

Session 6:

Listening Skills

- Importance of listening
- Techniques for effective listening
- Audio synthesis
 - Listening to audio clips
 - Question-answers based on the listened audio clips

Practice Sessions:

Practice audio synthesis

Session 7:

Reading Skills

- Comprehension
 - Techniques

Practice Sessions:

Comprehension exercises

Session 8:

Written Communication

- Essay writing
 - Characteristics of a good essay
 - Types of essays
 - Structure of an essay (introduction, main body, conclusion)
- Letter writing
 - Types of letters
 - Parts of a letter
- Official emailing
 - Structure and etiquettes of email writing
 - Tips to write an impressive email

Practice Sessions:

Essay

writing

Letter

writing

email

writing

Session 9:

Public Speaking

- Speech design
- Informative speeches
- Speeches for special occasions (Introduction, Welcome, Felicitation, Thanks, etc.)
- Extempore & impromptu speeches

Practice Sessions:

Conduct various types of speeches

Session 10:

Presentation Skills

- How to conduct effective and engaging presentations?
- Organization & structure of presentation
- Design of slides in PPT
- Body language & voice

Practice Sessions:

Conduct presentations

using PPTFeedback of

presentations

Session 11:

Group Discussions

- What is a GD?
- Skills assessed in GD
- Common mistakes
- Common GD topics

Practice Sessions:

Conduct practice GDs with video

recording Playing and analysis of

GDs conducted

Session 12:

Personal Interviews

- Preparation for Interview
 - Qualities interviewers looking for

- Getting ready for Interviews
- Company research
- Overall approach
- Just before interview

Session 13:

Personal Interviews

- Introducing yourself
 - Importance of introduction
 - Structure of introduction

Practice Sessions:

Practice introduction

Analysis and feedback on introduction

Session 14:

Personal Interviews

- Facing job interviews
 - Confidence
 - Body language
 - Right mindset
- Tips for facing Interviews
 - What to do (and not do) during interviews?
 - Best practices and common mistakes of answering questions

Practice Sessions:

Practice common technical

questions Practice common

HR/behavioral questions Conduct

mock interviews

| S No | Module Name | No. of Hours | Theory | Lab | No of Questions for CCEE | CCEE Marks | Lab Marks | Internal Marks | Total Marks |
|------|----------------------------------------------------------|--------------|------------|------------|----------------------------------------------------------------------------------------|------------|------------|----------------|-------------|
| 1. | Linux Programming and Cloud Computing | 50 | 28 | 22 | Linux - 20 & CC - 20 | 40 | 40 | 20 | 100 |
| 2. | Practical Machine Learning | 140 | 60 | 80 | Machine Learning - 40 | 40 | 40 | 20 | 100 |
| 3. | Advance Analytics using Statistics | 90 | 46 | 44 | Advanced Analytics - 40 | 40 | 40 | 20 | 100 |
| 4. | Python & R Programming | 80 | 44 | 36 | Python - 30, R - 10 | 40 | 40 | 20 | 100 |
| 5. | Data Collection and DBMS (Principles, Tools & Platforms) | 90 | 44 | 46 | SQL - 12, NoSQL -04, MongoDB - 12, Cassandra - 12 | 40 | 40 | 20 | 100 |
| 6. | Big Data Technologies | 150 | 66 | 84 | Big Data - 5, Hadoop - 5, HIVE – 5, HBase - 5, Map Reduce - 5, Spark - 10, Airflow - 5 | 40 | 40 | 20 | 100 |
| 7. | Data Visualization - Analysis and Reporting | 50 | 26 | 24 | Visualization - 40 | 40 | 40 | 20 | 100 |
| 8. | Java Programming | 70 | 34 | 36 | Java - 40 | 40 | 40 | 20 | 100 |
| 9. | Aptitude & Effective Communication | 90 | 90 | - | - | - | - | (100) | Grade |
| 10. | Project | 90 | - | 90 | - | - | - | (100) | Grade |
| | Total Marks | 900 | 438 | 462 | | 320 | 320 | 160 | 800 |