Suggested Teaching Guidelines for

Python & R Programming

e-DBDA September 2020

Duration: 36 Classroom hours + 24 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

Reference Book:

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

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:

Suggested Teaching Guidelines for
Python & R Programming
e-DBDA September 2020

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:

- Pass, 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 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

Suggested Teaching Guidelines for
Python & R Programming
e-DBDA September 2020

- Modifying Dictionaries
- Deleting Items from Dictionaries

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
- Declare, assign and retrieve values from Lists
- Accessing list
- Operations in 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

Session 5:

- Function and Methods
- Defining a function
- Calling a function
- Types of functions
- Function Arguments
- Anonymous functions
- Global and local variables
- Using Optional and Named Arguments
- Using type, str, dir, and Other Built-In Functions

Suggested Teaching Guidelines for

Python & R Programming

e-DBDA September 2020

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 fofunon".

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:

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

Q. 2.We can define sum from 1 to x (i.e. 1 + 2 + ... + x) recursively as follows for integer x ≥ 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. 3. Define a function overlapping () that takes two lists and returns True if they have at least one member in common, False otherwise.

Q. 4.Write a function find_longest_word() that takes a list of words and returns the length of the longest one.

Q. 5.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

Suggested Teaching Guidelines for

Python & R Programming

e-DBDA September 2020

Q. 6.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. 7.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
- 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 6:

- Working with Tuples
- Introducing Tuples
- Accessing tuples
- Operations

Session 7&8:
Advanced Python:

- Object Oriented Python
- OOPs concept
- What's an Object?
- Indenting Code
- Native Data types
- Declaring variables
- Referencing Variables
- Object References
- Class and object
- Attributes, Inheritance
- Overloading & Overriding
- Data hiding
- Regular Expressions Using python
- Object Oriented Linux Environment

Session 9:

- Operations Exception
- Exception Handling
- Except clause
- Try finally clause
- User Defined Exceptions

Suggested Teaching Guidelines for

Python & R Programming

e-DBDA September 2020

Session 10 & 11:

- Working with Pandas
- Data wrangling with Pandas
- Working with NumPy
- Data cleaning with Python

Session 12 & 13:

- Working with beautiful soup
- Working with matplotlib, seaborn
- Working with ggplot, plotly

R-Programming:

Session 14:

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

Session 15:

- 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

Session 16:

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

Session 17:

- Functions
- Built in functions in R (numeric, character, statistical)
- Interactive reporting with R markdown
- Introduction to R Shiny

Session 18:

- 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

Suggested Teaching Guidelines for

Practical Machine Learning e- DBDA September 2020

Duration: Duration: 42 Classroom hours + 38 Lab hours

Objective: Practicing Machine Learning Algorithms

Prerequisites: Good knowledge of Python Programming and Statistics

Evaluation method: Theory exam– 40%
Lab Exam - 40%
Internal exam- 20%

List of Books / Other training material
Reference Book:

1. Introduction to Machine Learning with Python - A Guide for Data Scientists, Muller Andreas / Shroff Publishers
2. Machine Learning with R by Brett Lantz
3. Machine Learning for Big Data: Hands- On for Developwer by Jasaon Bell, Wiley
4. Machine Learning: Hands-on for Developers and Technical Professionals
5. Machine Learning: A Bayesian and Optimization Perspective
6. Introduction to Machine Learning, Third Edition
7. R in Action, Robert Kabakoff

Note:

- Pytorch Framework should be taught in Lab Hours

Note: Each session having 2 Hours
Session 1

- What is machine learning?
- Algorithm types of Machine learning
- Supervised and Unsupervised Learning
- Uses of Machine learning
- Evaluating ML techniques
- Introduction to Scikit Learn

Session 2

- Clustering
- Hierarchical Clustering &K means
- Distance Measure and Data Preparation – Scaling & Weighting
- Evaluation and Profiling of Clusters
- Hierarchical Clustering
- Principal Component analysis

Session 3& 4

- Decision Trees
- Classification and Regression Trees
- Concept of Model Ensembling
- Random forest, Gradient boosting Machines, Model Stacking
- CAT Boost

- XG Boost

Session 5 & 6

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

Session 7 & 8:

- Linear Regression
- Logistic Regression
- Polynomial Regression
- Stepwise Regression
- Ridge Regression
- Lasso Regression
- ElasticNet Regression

Session 9:

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

Session 10 &11:

- 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

Session 12:

- ML in Real Time
- Algorithm Performance Metrics
- ROC and AOC
- Confusion Metrix
- F1 Score
- MSE and MAE

Session 13:

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

Session 14:

- Anomaly detection

Suggested Teaching Guidelines for

Practical Machine Learning e- DBDA September 2020

- Point, Contextual and Collective Anomaly
- Supervised and Unsupervised anomaly detection

Session 15:

- Introduction to NLP
- Working with NLTK
- Word2Vec
- GloVe word vectors
- Sentiment Classification

Session 16 & 17 :

- Introduction to Deep Learning
- Introduction to Tensorflow 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

Session 18

Deep Learning Essentials

- Early Stopping for Preventing Overfitting
- Dropout
- Training Methods for Neural Network (High-Level Overviews only)
  - 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)
- Classic Backpropagation
- Momentum Backpropagation
- ADAM
  - L1 and L2 Regularization

Session 19 & 20

Convolutional Neural Network using PyTorch

- Introduction to PyTorch Framework
- Pytorch vs Tensorflow
- Convolutional Concept
- Inception Network
- Transfer Learning
- Data Augmentation
- Object Detection
- YOLO Algorithm (High-Level Overview)

Session 21

Recurrent Neural Network (RNN) using Pytorch

- RNN Concept
- Types of RNNs

Suggested Teaching Guidelines for

Practical Machine Learning e- DBDA September 2020

- 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

Object Oriented Programming with Java 8
e-DBDA September 2020

Duration: 26 Classroom hours + 24 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
Reference:

1. Java - The Complete Reference by Herbert Schildt / Tata Mcgraw Hill Education
2. Java Server Programming (J2EE 1.7 Edition) Black Book by Dreamtech Software Team
3. Java 8 Programming Black Book by Dreamtech Press
4. Core Java : Fundamentals - Volume 1 Gary Cornell, Cay S. Horstmann/ Pearson
5. Programming in Java by Sachin Malhotra, Saurabh Choudhary / Oxford University Press
6. Core Java : Advanced Features - Volume 2 Gary Cornell, Cay S. Horstmann/ Pearson
7. Beginning Java 2 by Ivor Horton; Wrox Publication
8. The Complete Reference Java Eight Edition, Herbert Schidt/ TMH
9. Object-Oriented Analysis and Design with applications by Booch
10. Core Java 8 for Beginners by Sharanam Shah, Vaishali Shah / Shroff Publishers & Distributors
11. Murach's Java Programming 4th edition by Joel Murach / Shroff Publishers & Distributors
12. Advanced Java programming by Uttam K Roy / Oxford University press
13. Sun Certified Enterprise Architect For Java EE Study Guide by Cade, 2nd Edition (Paperback)
14. Programming in Java by Sachin Malhotra, Saurabh Choudhary / Oxford University Press
15. Professional Java EE Design Patterns by Murat Yener, Alex Theedom, Reza Rahman

Note: Each session having 2 Hours

Session 1&2

- Java 8 Basics :Overview of Java, Features of Java, Scope of variables
- Object Oriented Concepts
- 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:

- 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)

Session 4:

- 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

Assignment – Reading:
Study the book Java FAQ

Assignment – Tutorial:
Compare syntactical similarities and dissimilarities between Java and C++

Session 5:

- 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:

- 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:

- Enumerations, Auto boxing, and Annotations
- Lambda Expressions
- Java 8 New Features

Session 8 & 9 :

- Java API: java.util, java.lang, java.math, java.io

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

Session 10:

- 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:

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

Session 12:

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

Session 13:

- Synchronization, Inter-Thread communication
- Executor Framework overview
- Files
- Byte Streams and Unicode Character Streams
- Persistence of objects
- Object Serialization Methods

Suggested Teaching Guidelines for

Linux Programming and Cloud Computing
e-DBDA September 2020

Duration: 22 Classroom hours + 18 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

Linux Programming
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)

Cloud Computing
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 AnthonyT 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

Linux Programming
Session 1 & 2:

Linux History and Operation

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

Installing and Configuring Linux (Ubuntu and CentOS)

- Introduction to Installation and Media Types
- Performing a Custom Linux Server Installation
- Run Levels and the Startup/Shutdown Sequence
- 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 an its variants, touch, echo, who, whoami, ps, kill,makefile,etc.)

Assignment –Lab:
Getting Acquainted with the Linux Environment
Use various commands in Linux system.

Session 3

Gaining confidence with Linux

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

VI/vim/gedit editior

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

Session 4, 5, & 6

Linux shell programming

- Introduction to Shells
  a. What is 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
  - Get start with simple scripts (User variable, expr, multiple command)
  - Wild cards (* and ?)
  - Command line arguments
  - Arithmetic in shell scripts
  - *Read* and *echo* commands in shell scripts
  - The *tput* command
  - Taking decisions:
    - if-then-fi
    - if-then-else-fi
    - The test command (file tests, string tests)
    - Nested if-elses
    - The case control structure
- The loop control structure
  a. The while, until and for loop structures
  b. The break and continue statements
- Shell metacharacters
- Command line expansion
- Directory stacks manipulation

- Job control, history and processes
- Built-ins and functions
- Shell Files

Assignment –Lab:
Review Exercises

Sessions 7:

- Introduction to cloud
- What computing paradigms are there?
- Characteristics and benefits
- Understanding Cloud Vendors (AWS/Azure/GCP)
- Definition
- Characteristics
- Components

Lab Assignments:

- Study about cloud and other similar configuration
- Explore available solutions
- Cloud Architecture

Session 8 & 9:

- Introduction to SaaS
- Pros and Cons of SaaS Model
- Traditional Packaged software Vs SaaS
- SaaS examples
- Introduction to IaaS
- Examples
- Introduction to virtualization
- Types and Uses of Virtualization
- Virtual Machine Provisioning
- Virtual Machine Migration Services
- Private Cloud Computing Deployment
- Introduction to PaaS
- Storage as Service(RAID)
- Challenges of cloud environment
- Hypervisor
- Comparisons of web services
- Organizational Scenarios of Clouds

Lab Assignments:

- Provide a solution on cloud as SAAS using available systems.

Sessions 10:

- Administering & Monitoring cloud services,

- benefits and limitations,
- Deploy application over cloud.
- Comparison among SAAS, PAAS, IAAS,
- Cloud Computing Basics,
- Cloud Products and Solutions,
- Cloud Pricing,
- Compute Products and Services,

Session 11:

- Elastic Cloud Compute
- Dashboard
- Launching Linux VM
- Accessing Linux VM
- Launching & Accessing Windows server VM

Lab Assignments:

- Study about cloud and other similar configuration
- Exposure to big data technologies on cloud

Suggested Teaching Guidelines for

Data Visualization - Analysis and Reporting
e-DBDA September 2020

Duration: **16 Classroom hours + 14 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
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

Note:

- **Each session having 2 Hours**
- **Tool to be use: PowerBI**

Session 1:
Lecture

- BI basic,
- Information gathering,
- Decision making,
- Managing BI,
- 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 2 & 3:
Lecture

Suggested Teaching Guidelines for

Data Visualization - Analysis and Reporting
e-DBDA September 2020

MS EXCEL (Theory – 06 Hrs. and Lab – 06 Hrs.)

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

Session 4
Lecture
Data analytics Life Cycle:

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

Session 5
Lecture

- Introduction to Power BI
- Intelligent data analysis,
- Nature of Data,
- Analytic Processes and Tools,
- Analysis vs. Reporting
- Modern Data Analytic Tools

Session 6, 7
Lecture

- 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 – Bar chart, Box plots, Histograms, Gannt charts, Glyph chart, Sanky diagam, Word Cloud etc.
  - Hierarchies and relationships – Pie chart, stacked bar, Tree map etc.
  - Changes over time – Line chart, sparklines, candlestick/ohlc etc.
  - Connections and relationships – scatter lots, bubble plots, radial network, heat maps, etc.

Session 8:
Lecture

- Choosing appropriate visuals
- Applying calculations, statistics
- Data sorting, filters
- Interactive visualization

- Event listeners/callbacks
- Data updation
- Visual updation

- Dashboard Design

Suggested Teaching Guidelines for

Data Collection and DBMS (Principles, Tools & Platforms)
e- DBDA September 2020

Duration: 22 Classroom hours + 18 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
Reference:

1. MongoDB in Action by DreamTechss
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 having 2 Hours

Session 1:

- 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:

- 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:

- 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

Session 4:

- 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

Session 5 & 6:

- 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:

- 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 8:

- 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 9:

- 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 10:

- 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 11:

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

Suggested Teaching Guidelines for

Big Data Technologies e-DBDA September 2020

Duration: 60 Classroom hours + 60 Lab hours

Objective: To reinforce knowledge of BigData Technologies such as Hadoop, Map reduce, Hive, Spark (PySpark), Airflow, Kafka.

Prerequisites: Knowledge of Linux command, SQL and Core Java, Python

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

List of Books / Other training material

Reference:

1. Hadoop The Definitive Guide 4thEdition by O'Rellay (Author: - Tom White)
2. Hadoop In Practice by Manning (Author: - ALEX HOLMES)
3. Pro Hadoop by Aprss(Author:-Jason Venner)
4. Hadoop with python
5. Hadoop Real-World Solutions Cookbook by Packet publication (Author: Jonathan R. Owens, Jon Lentz,Brian Femiano)
6. Hadoop In Action by Manning Publications (Author: - CHUCK LAM)
7. Data Architecture: A Primer for the Data Scientist: Big Data, Data Warehouse and Data Vault
8. Big Data Made Easy: A Working Guide to the Complete Hadoop Toolset
9. 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
10. Hadoop: The Definitive Guide, SPD
11. Big Data, Black Book by DreamTech
12. Programming Hive by O'Rellay (Author:- Edward Capriolo, Dean Wampler, and Jason RutherglenEdward Capriolo, Dean Wampler, and Jason Rutherglen)

Note: Each session having 2 Hours

Introduction to Bigdata and Hadoop (Theory- 12 Hrs and Lab- 04 Hrs)
Session: 1 & 2
Lecture

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 Data Curation,
- 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

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: 3 & 4

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

Hadoop Installation and Cluster Configuration(Lab – 02 Hrs)

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,

Session: 5

Hadoop Architecture

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

Session: 6

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.

Map Reduce(Theory – 08 Hrs & Lab – 08 Hrs)
Session: 7

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, Multiple Inputs),
- Output Formats (Text Output, Binary Output, Multiple Output).
- Distributed Cache

Session: 8

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

Session: 9
Map Reduce Streaming

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

Hadoop ETL
Session: 10

- Hadoop ETL Development,
- ETL Process in Hadoop,
- Discussion of ETL functions,
- Data Extractions,
- Need of ETL tools,
- Advantages of ETL tools.

HBASE (6 hours Theory + 4 hours lab)
Session: 11

Introduction to HBase

- Overview of HBase

- HBase architecture
- Installation

Session: 12 and 13

The HBase Admin 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

Hive (Theory – 08 Hrs & 8 Hrs Lab)

Session: 14

The Hive Data-ware House

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

Session: 15

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.

Session: 16

Querying with Hive QL

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

Session: 17

More on Hive QL

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

Suggested Teaching Guidelines for

Big Data Technologies e-DBDA September 2020

Apache Airflow(Theory – 06 Hrs & Lab – 06 Hrs)
Session: 18, 19 and 20

- 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
- Work management with Airflow
- Automating an entire Data Pipeline with Airflow

Introduction to Apache Spark& Kafka (Theory – 20 Hrs & Lab – 28 Hrs)
Session: 21, 22 and 23

Apache Spark APIs for large-scale data processing

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

Session: 24

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

Session: 25 and 26 & 27

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

Session: 28

- Setting up Kafka Producer and Consumer
- Kafka Connect API

Session: 29 and 30

- Spark SQL
- Spark Scripts
- Spark MLlib

Lab Assignment

- Deep Learning with Spark

Lab Assignment

- Connecting DB's with Spark

Suggested Teaching Guidelines for
Big Data Technologies e-DBDA September 2020

- Accessing and Manipulating the DB's

Lab Assignment

- Demo: Capstone Project

Suggested Teaching Guideline for

Aptitude e- DBDA September 2020

Duration: 30 Classroom hours

Objective: To reinforce knowledge of general Aptitude & English

Prerequisites: Knowledge of Mathematics

List of Books / Other training material

Reference:

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

Note: Each session having 2 Hours

Session 1:
Lecture

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

Session 2:
Lecture

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

Session 3:
Lecture

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

Session 4:
Lecture

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

Session 5:
Lecture

- Logical reasoning
- Statement-Arguments
- Statement-Assumptions

Session 6:
Lecture

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

Session 7:
Lecture

- HCF and LCM
- Fraction
- Number system
- Permutation & combination

Session 8:
Lecture

- Ratio & Preparation
- Partnership

Session 9:
Lecture

- Average
- Percentage

Session 10:
Lecture

- Clock
- Probability

Session 11:
Lecture

- Pipes and cisterns
- Problem on streams

Session 12:
Lecture

- Time and work
- Work and Wages

Session 13:
Lecture

- Problem on Trains
- Problem on Speed and Velocity

Session 14:
Lecture

- Problem on Ages
- Profit and loss

Session 15:
Lecture

- Simple Interest,
- Compound Interest

Suggested Teaching Guidelines for

Advanced Analytics e-DBDA September 2020

Duration: 36 Classroom hours + 34 Lab hours

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

Prerequisites: Good Knowledge of Basic Mathematics

Evaluation method: Theory exam– 40%
Lab Exam - 40%
Internal exam- 20%

List of Books / Other training materials

Reference:

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

Note:

- **Each session mentioned is for theory and 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**
- **Trainer must teach 'Scipy' package in detail.**

Session 1:

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

Session 2:

- Intelligent data analysis,
- Nature of Data,
- Analytic Processes and Tools,
- Analysis vs. Reporting
- Modern Data Analytic Tools

- Visualization and Exploring Data

Session 3:

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

Session 4:

- Sample& population, Uni-variate and bi-variate sampling, re-sampling
- Sample Spaces and Events
- Joint, Conditional and Marginal Probability
- Bayes' Theorem

Session 5 & 6:

- Random Variable
- Probability Distribution and Data
  - Continuous and discrete distribution – (Normal, Binomial, and Poisson distribution)
- Central Limit Theorem

Session 7:

- Sampling and Estimation
- Statistical Interfaces
- Concepts of Correlation
- Covariance
- Outliers

Session 8 & 9:

- Predictive modelling and analysis
  - Application
  - Types
  - Benefits and challenges
  - The Future of predictive modelling
  - The Limitations of Predictive modelling
  - Predictive modelling Tools

Session 10 & 11:

- 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;

Session 12:

- Prescriptive Modelling
  - Difference between predictive and prescriptive modelling

- How prescriptive analytics works?
- Examples and use cases

Session 13:

- Regression Analysis
- Forecasting Techniques

Session 14:

- Simulation and Risk Analysis
- Optimization, Linear, Nonlinear

Session 15:

- Overfitting and Its Avoidance:
  - Generalization,
  - Holdout Evaluation Vs Cross Validation;

Session 16:

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

Session 17:

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

Session 18:

- Factor Analysis,
- Directional Data Analytics,