BHARATI VIDYAPEETH
(DEEMED TO BE UNIVERSITY), PUNE (INDIA)

(Established u/s 3 of the UGC Act, 1956 vide Notification No.F.9-15/95-U-3 of the Govt. of India)

'A+ ' Grade Accreditation by NAAC

"Social Transformation Through Dynamic Education"

SCHOOL OF DISTANCE EDUCATION

PROGRAMME GUIDE

OF

DIPLOMA IN BIG DATA
(DBD)
(UNDER FACULTY OF MANAGEMET STUDIES)

BHARATI VIDYAPEETH (DEEMED TO BE UNIVERSITY), PUNE
SCHOOL OF DISTANCE EDUCATION
(Under Faculty of Management Studies)

DIPLOMA IN BIG DATA
(Choice Based Credit and Grade System)

1.   AIM AND OBJECTIVE

To introduce learner with Big Data Concept, decision making by doing analysis on the data and managing the data using Big Data Tools like Apache HADOOP, SPARK, MAPREDUCE, R. Students will get Knowledge of Decision making using analysis on the Big Data, also they will get Hands-on Big Data tools- HADOOP, SPARK, MAPREDUCE, R.

2. ADMISSION PROCEDURE

The Application Form is available on website of BVDU School of Distance Education i.e. *distance.bharatividyapeeth.edu*. The candidate will have to apply for admission to any academic programme of his / her choice thorough online. The candidate will be admitted provisionally to the programme on verification of the eligibility for admission. He / She will be asked to complete the eligibility requirement by submitting the required Marksheets, Bonafide Certificate (if required), Leaving/Transfer Certificate, Educational Gap Certificate (if required), Aadhaar Card etc. After verification of required documents candidate admission will be confirmed.

3. ELIGIBILITY:

The candidate must be a Graduate of any Faculty either of this University or of any other University whose degree is recognized, or equivalent to the corresponding degree of this University.

4. DURATION:

The duration of diploma programme conducted through open and distance learning mode is a One Year although student may complete the programme gradually within a maximum period of Three years.

5. MEDIUM OF COURSE:

The medium of instruction and examination is English only.

6. COURSE STRUCTURE :

Sr. No.	Name of the Course	Maximum Marks	No. of Credits	Study Input (in hrs.)
1.	Big Data	100	8	240 (120L + 120 P)
2.	HADOOP	100	8	240 (120L + 120 P)
3.	MapReduce	100	8	240 (120L + 120 P)
4.	R	100	8	240 (120L + 120 P)
Total		**400**	**32**	**960 hrs.**

7. CREDIT SYSTEM:

This Diploma Programme consists of four courses. Each course is allotted 8 credits. Each credit constitutes 30 hrs. Hence the study input hours for each course shall be 240 hrs. The study input has to be fulfilled as follows:

Study Input = 120 hrs (per course) Credit = 8 credits (per course)			
Counseling sessions (24 hrs)		Self-Learning Material (SLM) Size of e-content	Assignments
Lecture	Practical		
120 hrs	120 hrs	6 Units (Per Subject)	2

- **Lecture**
  - It shall be a face to face interaction of the student with the faculty which shall include the Teaching, Counseling sessions, difficulty solving etc.

- **Practical**
  - It shall provide a Lab Session where the students will perform Hands-on Practical on all the Programming Languages will be conducted by faculty in Computer Labs to give practical knowledge and Industrial Exposure to students.

8. SCHEME OF EXAMINATION:

The examination for the diploma programme will be held at the end of the academic session. The medium of instruction and the examination shall be in English. Each course shall be evaluated out of 100 marks out of which 80 marks are for a written examination which will be conducted by the university & 20 marks for submission of assignments.

9. STANDARD OF PASSING:

In order to pass in the course and to earn the assigned credits the student must obtain 40% marks at the written examination and 40% marks in evaluation of assignments.

For all courses, both UE and IA constitute separate heads of passing (HoP). In order to pass in such courses and to earn the assigned credits, the learner must obtain a minimum grade point of 5.0 (40% marks) at UE and also a minimum grade point of 5.0 (40% marks) at IA. . A student who fails at UE in a course has to reappear only at UE as backlog candidate and clear the Head of Passing. Similarly, a student who fails in a course at IA has to reappear only at IA as backlog candidate and clear the Head of Passing to secure the GPA required for passing.

The 10 point Grades and Grade Points according to the following table:

Range of Marks (%)	Grade	Grade Point
80≤Marks≤100	O	10
70≤Marks<80	A+	9
60≤Marks<70	A	8
55≤Marks<60	B+	7
50≤Marks<55	B	6
40≤Marks<50	C	5
Marks < 40	D	0

Table 1

The performance at UE and IA will be combined to obtain GPA (Grade Point Average) for the course. The weights for performance at UE and IA shall be 80% and 20% respectively. GPA is calculated by adding the UE marks out of 80 and IA marks out of 20.The total marks out of 100 are converted to grade point, which will be the GPA.

Formula to calculate Grade Points (GP):

Suppose that „Max" is the maximum marks assigned for an examination or evaluation, based on which GP will be computed. In order to determine the GP, Set x = Max/10 (since we have adopted 10 point system). Then GP is calculated by the following formulas

Range of Marks	Formula for the Grade Point
8x ≤ Marks≤10x	10
5.5x ≤ Marks<8x	Truncate (M/x) +2
4x ≤ Marks<5.5x	Truncate (M/x) +1

Table 2

Two kinds of performance indicators, namely the Semester Grade Point Average (SGPA) and the Cumulative Grade Point Average (CGPA) shall be computed at the end of each term. The SGPA measures the cumulative performance of a learner in all the courses in a particular semester, while the CGPA measures the cumulative performance in all the courses since his/her enrollment. The CGPA of learner when he /she completes the programme is the final result of the learner.

The SGPA is calculated by the formula

$$SGPA = \frac{\sum Ck * GPk}{\sum Ck}$$

where, Ck is the Credit value assigned to a course and GPk is the GPA obtained by the learner in the course. In the above, the sum is taken over all the courses that the learner has undertaken for the study during the Semester, including those in which he/she might have failed or those for which he/she remained absent. **The SGPA shall be calculated up to two decimal place accuracy.**

The CGPA is calculated by the following formula

$$CGPA = \frac{\sum C_k * GP_k}{\sum C_k}$$

where, Ck is the Credit value assigned to a course and GPk is the GPA obtained by the learner in the course. In the above, the sum is taken over all the courses that the learner has undertaken for the study from the time of his/her enrollment and also during the semester for which CGPA is calculated. **The CGPA shall be calculated up to two decimal place accuracy.**

The formula to compute equivalent percentage marks for specified CGPA:

	(10 * CGPA) - 10	If 5.00 ≤ CGPA < 6.00
	(5 * CGPA) **+** 20	If 6.00 ≤ CGPA < 8.00
% marks (CGPA)	(10 * CGPA) - 20	If 8.00 ≤ CGPA < 9.00
	(20 * CGPA) - 110	If 9.00 ≤ CGPA < 9.50
	(40 * CGPA) - 300	If 9.50 ≤ CGPA ≤ 10.00

Table 3

10. AWARD OF HONOURS:

A student who has completed the minimum credits specified for the programme shall be declared to have passed in the programme. The final result will be in terms of letter grade only and is based on the CGPA of all courses studied and passed. The criteria for the award of honours are given below.

Range of CGPA	Final Grade	Performance Descriptor	Equivalent Range of Marks (%)
9.5≤CGPA ≤10	O	Outstanding	80≤Marks≤100
9.0≤CGPA ≤9.49	A+	Excellent	70≤Marks<80
8.0≤CGPA ≤8.99	A	Very Good	60≤Marks<70
7.0≤CGPA ≤7.99	B+	Good	55≤Marks<60
6.0≤CGPA ≤6.99	B	Average	50≤Marks<55
5.0≤CGPA ≤5.99	C	Satisfactory	40≤Marks<50
CGPA below 5.0	F	Fail	Marks below 40

Table 4

Course 1: Big Data
Credits : 8

Course Objectives: To introduce learner with Big Data Concepts, Lifecycle and the role of the data scientists. It will also focus on the decision making by doing analysis on the data. In this course the learner will get the information using case study.

Learning Outcome:
- Good knowledge of Big Data Concepts
- Knowledge of Decision making using analysis on the Big Data

Unit No.	Contents
1.	**Introduction:** Big Data History, The Big Data Business Opportunity- Business Transformation Imperative, Big Data Business Model, Business Impact of Big Data
2.	**Big Data In Organization:** Data Analytics Lifecycle, Data Scientist Roles and Responsibilities – Discovery, Data Preparation, Model Planning, Model Building, Communicate Results, Operationalize, New Organizational Roles, Liberating Organizational Creativity.
3.	**Decision Theory and Strategy:** Business Intelligence Challenge, Big Data User Interface Ramifications, Human Challenge of Decision Making, Strategy for Decision Making- Big Data Strategy Document, Case Study.
4.	**Value Creation Process:** Understanding Big Data Value Creation, Value Creation Drivers, Michael Porter's Value Creation Models- Michael Porter's Five Forces Analysis, Michael Porter's Value Chain Analysis, Case Study.
5.	**Big Data User Experience:** The Unintelligent User Experience, Understanding the Key Decisions to Build a Relevant User Experience, Using Big Data Analytics to Improve Customer Engagement, Uncovering and Leveraging Customer Insights, Big Data can Power a New Customer Experience.
6.	**Big Data Use Cases:** The Big Data Envisioning Process –1. Research Business Initiatives, 2. Acquire and Analyze your Data, 3. Brainstorm New Ideas, 4. Prioritize Big Data Use Cases, 5. Document Next Steps, The Prioritization Process.

Suggested Readings (Books, Websites etc):

Big Data- Understanding How Big Data Power Big Business –By Bill Schmarzo

Edureka lectures   **Link: -** https://www.youtube.com/watch?v=A02SRdyoshM

Suggested MOOC:
- **Coursera**
- **Udemy**
- **edX**

Course 2: HADOOP

Credits : 8

Course Objectives: To introduce learners with HADOOP Architecture. The course also focuses on the HADOOP Distributed File System. Learners will also get practical knowledge of HADDOP Environment setup.

Learning Outcome:
- Good Knowledge about HADOOP.
- Working of HADOOP Distributed File System.
- Hands-on experience on HADOOP.

Unit No.	Contents
1.	**Big Data Overview:** What is Big Data?, What Comes Under Big Data?, Benefits of Big Data, Big Data Technologies, Operational vs. Analytical Systems, Big Data Challenges
2.	**Introduction To HADOOP:** Hadoop Architecture, Advantages of Hadoop
3.	**HDFS:** Hadoop Distributed File System, How Does Hadoop Work?
4.	**Environment Setup:** Pre-installation Setup, Installing Java, Downloading HADOOP , HADOOP Operation Modes, Installing HADOOP in Standalone Mode, Installing HADOOP in Pseudo Distributed Mode, Verifying HADOOP Installation
5.	**HDFS Overview:** Features of HDFS, HDFS Architecture, Goals of HDFS
6.	**HDFS Operations:** Starting HDFS, Listing Files in HDFS, Inserting Data into HDFS, Retrieving Data from HDFS, Shutting Down the HDFS, HDFS Command Reference.

Suggested Readings (Books, Websites etc):

Big Data- Understanding How Big Data Power Big Business –By Bill Schmarzo

Edureka lectures   **Link: -** https://www.youtube.com/watch?v=A02SRdyoshM

Tutorials Point- www.tutorialspoint.com

Suggested MOOC:
- **Coursera**
- **Udemy**
- **edX**

Course 3: MapReduce

Credits : 8

Course Objectives: to introduce learners with MapReduce. Learners will also get the knowledge about multi-node and clusters in MapReduce. Practical hands-on knowledge will be also provided to learners.

Learning Outcome:
- Good Knowledge about MapReduce.
- Practical knowledge and Programmimg in MapReduce.

Unit No.	Contents
1.	**Big Data And HADOOP Overview:** Introduction to HADOOP, HDFS Overview
2.	**MapReduce:** What is MapReduce?, The Algorithm, Inputs and Outputs (Java Perspective), Terminology
3.	**MapReduce Case Study:** Example Scenario, Compilation and Execution of Process Units Program
4.	**Multi-Node:** Installing Java, Creating User Account, Mapping the nodes, Configuring Key Based Login, Installing HADOOP, Configuring HADOOP, Installing HADOOP on Slave Servers, Configuring HADOOP on Master Server, Starting HADOOP Services
5.	**Multi-Node Cluster:** Adding a New DataNode in the HADOOP Cluster, Adding a User and SSH Access, Set Hostname of New Node, Start the DataNode on New Node
6.	**Removing Node From Cluster:** Removing a DataNode from the HADOOP Cluster

Suggested Readings (Books, Websites etc):

Big Data- Understanding How Big Data Power Big Business –By Bill Schmarzo
Edureka lectures   **Link: -** https://www.youtube.com/watch?v=A02SRdyoshM
Suggested MOOC:

- **Coursera**
- **Udemy**
- **edX**

Course 4: R Programming

Credits : 8

Course Objectives: To introduce learners with R-Programming Language. It will also focus on Software Environment for statistical analysis, representation and reporting. Learners will pursue practical knowledge on R-Programming.

Learning Outcome:
- Good knowledge about R-Programming.
- Knowledge about statistical analysis on data and preparing Reports.
- Hands-on programming knowledge of R-Programming.

Unit No.	Contents
1.	**Basics of R:** Introduction, R-Studio, Programming with R, Basic Data Types, Vectors, Arrays, Classes
2.	**Factors and Data Frames:** Introduction to Factors- *Factor Levels, Summarizing factor, Ordered Factors, Comparing ordered Factors*; Introduction to Data Frame- *Creating Data Frame*, Subsetting of Data Frames, Sorting Data Frames
3.	**Lists:** Introduction, Creating a List, Accessing List Elements, Manipulating List Elements, Merging Lists
4.	**Conditionals and Control Flow:** Relational Operators, Relational Operators and Vectors, Logical Operators- *AND, OR, NOT*; Logical Operators and Vectors, Conditional Statements.
5.	**Iterative Programming in R:** Introduction, While Loop, For Loop, Looping Over List- *Loops for Vectors, Loops for Matrices, Loops for Data Frames, Loops for Lists.*
6.	**Functions in R :** Introduction, Writing a Function in R, Nested Functions, Function Scooping, Recursion, Loading an R Package, Charts and Graphs, Data Interfaces

Suggested Readings (Books, Websites etc):

Tutorials Point- www.tutorialspoint.com

Suggested MOOC:

- **Coursera**
- **Udemy**
- **edX**