YMCA University of Science and Technology, Faridabad
BCA Scheme of Studies / Examination
Semester – VI

Course No.	Course Title	Schedule				Sessional Marks/ Internal	Marks for End Term Examination		Total Marks	Credits
		L	T	P	Total		Theory	Practical		
BCA-DS-311	Big Data – II	3	-		3	25	75	-	100	3
BCA-DS-312	Machine Learning – II	3	-		3	25	75	-	100	3
BCA-DS-313	Employability Skills	3	-		3	25	75	-	100	3
BCA-DS-314	Bio Informatics	3	-		3	25	75	-	100	3
As per list above	General Elective - IV	3	-		3	25	75	-	100	3
BCA-DS-315	Big Data – II Lab	-		4	4	25	-	50	75	2
BCA-DS-316	Machine Learning – II Lab	-		4	4	25	-	50	75	2
BCA-DS-317	Presentation	-		2	2	25	-		25	1
BCA-DS-318	Group Discussion	-		2	2	25	-		25	1
	Total				27	225	375	100	700	21

Note: Exam duration will be as under
    (a) Theory exams will be of 3 hours duration
    (b)Practical exams will be of 3 hours duration

BCA-DS-311: BIG DATA II
B.C.A VI SEMESTER

No. of Credits: 3

L T   P   Total

3 0   0   3

Sessional: 25 Marks

Theory :   75 Marks

Total :   100 Marks

Duration of Exam: 3 Hours

Note: Examiner will be required to set *Seven* questions in all having two parts. Part I will have Question Number 1 consisting of total 10 parts (short-answer type questions) covering the entire syllabus and will carry 15 marks. In Part II, there will be *Six* questions. Examiner will set one and a half questions from each Unit of the syllabus and each question will carry 15 marks. Question Number 1 will be compulsory. In addition to compulsory question, student will have to attempt four more questions from Part II.

COURSE OBJECTIVES:
This course is to help students learn, understand, and practice big data analytics and machine learning approaches, which include the study of modern computing big data technologies and scaling up machine learning techniques focusing on industry applications. Mainly the course objectives are: conceptualization and summarization of big data and machine learning, trivial data versus big data, big data computing technologies, machine learning techniques, and scaling up machine learning approaches.

SYLLABUS

UNIT – I
Hadoop I/O The Writable Interface, Writable Comparable and comparators, Writable Classes: Writable wrappers for Java primitives, Text, Bytes Writable, Null Writable, Object Writable and Generic Writable, Writable collections, implementing a Custom Writable: Implementing a Raw Comparator for speed, Custom comparators

UNIT – II
Pig: Introduction, Pig Architecture, Pig Latin Data model, Pig Latin operators, Pig Diagnost ic operators, Pig Data manipulation, built-in functions, user defined functions, Scripting with Pig Latin

UNIT – III
Hive: Getting Started with Apache Hive, Hive architecture, working with Hive Data Types, Creating and Managing Databases and Tables, views and indexes, Hive Data Manipula t ion Language, Querying and Analyzing Data.

UNIT – IV
Spark: Introduction, Spark Architecture, Advantages over traditional data approaches, Spark Ecosystem, Spark for Big Data processing and its applications
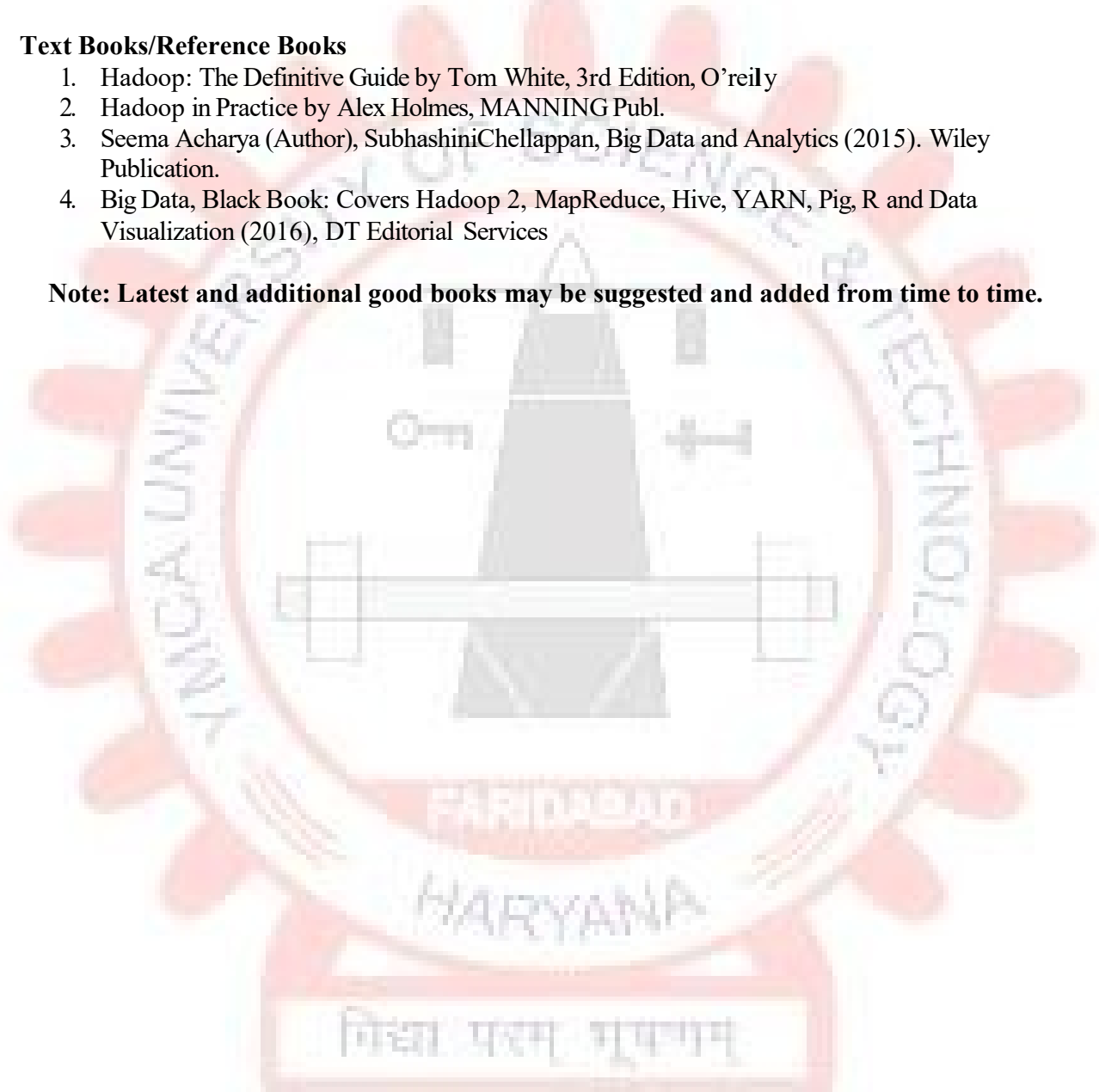
COURSE OUTCOMES:
1. To introduce programming tools PIG & HIVE in Hadoop ecosystem.
2. Preparing for data summarization, query, and analysis.
3. Applying data modeling techniques to large data sets.
4. Creating applications for Big Data analytics.
5. Building a complete business data analytic solution.

Text Books/Reference Books
1. Hadoop: The Definitive Guide by Tom White, 3rd Edition, O'reily
2. Hadoop in Practice by Alex Holmes, MANNING Publ.
3. Seema Acharya (Author), SubhashiniChellappan, Big Data and Analytics (2015). Wiley Publication.
4. Big Data, Black Book: Covers Hadoop 2, MapReduce, Hive, YARN, Pig, R and Data Visualization (2016), DT Editorial Services

Note: Latest and additional good books may be suggested and added from time to time.

BCA-DS-312: MACHINE LEARNING - II
B.C.A VI SEMESTER

No. of Credits: 3		Sessional:	25 Marks
L T P Total		Theory :	75 Marks
			100
3 0 0 3		Total :	Marks
		Duration of Exam:	3 Hours

Note: Examiner will be required to set *Seven* questions in all having two parts. Part I will have Question Number 1 consisting of total 10 parts (short-answer type questions) covering the entire syllabus and will carry 15 marks. In Part II, there will be *Six* questions. Examiner will set one and a half questions from each Unit of the syllabus and each question will carry 15 marks. Question Number 1 will be compulsory. In addition to compulsory question, student will have to attempt four more questions from Part II.

COURSE OBJECTIVES:
This course is designed to help students learn to design and implement machine learning solutio ns to classification, regression, and clustering problems; and be able to evaluate and interpret the results of the algorithms.

SYLLABUS

UNIT-1
COMBINING DIFFERENT MODELS: Evaluating Machine Learning algorithms and Model Selection, Introduction to Statistical Learning Theory, Ensemble Methods: Boosting, Bagging, Random Forests.

UNIT- II
DIMENSIONALITY REDUCTION: Dimensionality Reduction, Linear Discriminant Analysis – Principal Component Analysis, Kernel PCA– Factor Analysis – Independent Component Analysis

UNIT –III
LEARNING WITH NEURAL NETWORKS: Perceptron, multilayer neural networks, learning neural networks structures – Deep Learning and Feature Representation Learning

UNIT IV
REINFORCEMENT LEARNING: Reinforcement Learning overview, Elements of Reinforcement Learning, Generalization in reinforcement learning, policy search, adaptive dynamic programming.

COURSE OUTCOMES:
After completing this course, student will be able
1. To have a good understanding of numerical approaches to learning (optimization and integration).

2. Develop an idea of how to choose a probabilistic model to describe a particular type of data.

3. To know how to evaluate a learned model in practice.

4. To understand the role of machine learning in massive scale automation.

5. Understand the mathematics necessary for constructing novel machine learning solutions.

Text Books/Reference Books

1. K. P. Murphy, ―Machine Learning: A probabilistic perspective, MIT Press, 2012.
2. M. Mohri, A. Rostamizadeh, and A. Talwalkar, ―Foundations of Machine Learning, MIT Press, 2012.
3. D. Barber, ―Bayesian Reasoning and Machine Learning, Cambridge University Press, 2012.
4 . P. Flach, ―Machine Learning: The art and science of algorithms that make sense of data, Cambridge University Press, 2012.

Note: Latest and additional good books may be suggested and added from time to time.

BCA-DS-313: EMPLOYABILITY SKILLS
B.C.A VI SEMESTER

No. of Credits: 3

L T   P   Total

3 0   0   3

Sessional: 25 Marks

Theory :   75 Marks

     100

Total :   Marks

Duration of Exam: 3 Hours

Note: Examiner will be required to set *Seven* questions in all having two parts. Part I will have Question Number 1 consisting of total 10 parts (short-answer type questions) covering the entire syllabus and will carry 15 marks. In Part II, there will be *Six* questions. Examiner will set one and a half questions from each Unit of the syllabus and each question will carry 15 marks. Question Number 1 will be compulsory. In addition to compulsory question, student will have to attempt four more questions from Part II.

COURSE OBJECTIVES:
1. Review and evaluate your personal attributes and employability skills which are necessary for successful employment and career development.
2. Investigate employment opportunities, complete a job application including a CV for for a specific job opportunity.
3. Understanding of employability skills and personal attributes in relation to job selection methods
4. Understanding of the skills and attributes required to maintain and successfully develop in employment.

SYLLABUS

UNIT –I
VERBAL ABILITY and BASIC QUANTITATIVE APTITUDE: Synonyms, Antonyms and One word substitutes, Speed, Time and Distance, Time and Work, Linear Equations, Progressions (Sequences & Series), Permutation and Combination, Probability, Functions, Set Theory, Number Systems, LCM and HCF, Percentages, Collection and Scrutiny of data: Primary data, questionnaire and schedule; secondary data, their major sources including some government publications

UNIT II
LOGICAL REASONING – I: Number and Letter Series, Calendars, Clocks, Cubes, Venn Diagrams, Binary Logic, Seating Arrangement, Logical Sequence, Logical Matching, Logical Connectives, Syllogism. Blood Relations; concept of a statistical population and sample from a population; qualitative and quantitative data.

UNIT-III
MEASURES OF CENTRAL TENDENCY: Objective of averaging, characteristics of good average, types of average, arithmetic mean of grouped and ungrouped data, correcting incorrect values, weighted arithmetic mean, median - median of grouped and ungrouped data merit and

limitation of median, computation of quartile, decile and percentile Mode - calculation of mode of grouped and ungrouped data, merits and limitation of mode, relationship between mean, median and mode. Geometric mean and Harmonic mean.

UNIT-IV
PRESENTATION OF DATA: Construction of tables with one or more factors of classificat io n; Diagrammatic and Graphical representation of non-frequency data; Frequency distribut io n, cumulative frequency distribution and their graphical representation - histogram, Column Graphs, Bar Graphs, Line Charts, Pie Chart, Data Interpretation – Introduction and approach

COURSE OUTCOMES:
After completing this course student will be able:
1. To help students explore their values and career choices through individual skill assessments
2. To make realistic employment choices and to identify the steps necessary to achieve a goal
3. To develop and practice self-management skills for the work site
4. To explore and practice basic communication skills
5. To learn skills for discussing and resolving problems on the work site
6. To assess and improve personal grooming
7. To promote safety awareness including rules and procedures on the work site

Text Books/Reference Books:

1. Bajpai, N. Business Statistics, Pearson, 2010
2. Sharma J.K., Business Statistics, Pearson Education India, 2010.
3. Richard I Levin, David S. Rubin: Statistics for Management, Pearson Prentice Hall Education Inc. Ltd, NewDelhi, 5th Ed.
4. Anderson; David R, Dennis J. Sweeney and Thomas A. Williams, Quantitative Methods for Business, Prentice-Hall, WestPublishing Company, 1996.
5. CAT Complete course, UPKAR publications

Note: Latest and additional good books may be suggested and added from time to time.

BCA-DS-314: BIO INFORMATICS
B.C.A VI SEMESTER

No. of Credits: 3			Sessional:	25 Marks
L T	P	Total	Theory :	75 Marks
3 0	0	3	Total :	100Marks
			Duration of Exam:	3 Hours

Note: Examiner will be required to set *Seven* questions in all having two parts. Part I will have Question Number 1 consisting of total 10 parts (short-answer type questions) covering the entire syllabus and will carry 15 marks. In Part II, there will be *Six* questions. Examiner will set one and a half questions from each Unit of the syllabus and each question will carry 15 marks. Question Number 1 will be compulsory. In addition to compulsory question, student will have to attempt four more questions from Part II.

COURSE OBJECTIVES:
The basic objective is to give students an introduction to the basic practical techniques of bioinformatics. Emphasis will be given to the application of bioinformatics and biologica l databases to problem solving in real research problems. The students will become familiar with the use of a wide variety of internet applications, biological database and will be able to apply these methods to research problems.

SYLLABUS

UNIT-I
HISTORY, SCOPE AND IMPORTANCE: Important contributions - sequencing development - aims and tasks of Bioinformatics - applications of Bioinformatics – challenges and opportunities - Computers and programs – internet - world wide web – browsers – EMB net – NCBI.

UNIT-II
DATABASES - TOOLS AND THEIR USES: Importance of databases - nucleic acid sequence databases - protein sequence data bases - structure databases - bibliographic databases and virtual library - specialized analysis packages.

UNIT-III
INTRODUCTION TO BIOINFORMATICS ALGORITHMS: Algorithms and Complexit y Biological algorithms versus computer algorithms – The change problem –Correct versus Incorrect Algorithms – Recursive Algorithms – Iterative versus Recursive Algorithms – Big-O Notations – Algorithm Design Techniques.

UNIT-IV
UNIX COMMANDS: Advanced Unix Commands-Introduction-ls-cat-more-, Advanced Unix commands-mv-rm-rmdir-uniq-sort-, Advanced Unix commands-grep. PERL: Introduction to Perl-scalars, Arrays-Using standard Perl Modules-Perl regular expressions I.

BIOPERL: Installation and usage of bioperl modules

COURSE OUTCOMES:

1. The basic objective is to give students an introduction to the basic practical techniques of bioinformatics. Emphasis will be given to the application of bioinformatics and biologica l databases to problem solving in real research problems.
2. The students will become familiar with the use of a wide variety of internet applicatio ns, biological database and will be able to apply these methods to research problems
3. The student will be aware about the bioinformatics methods including accessing the major public sequence databases, use of the different computational tools to find sequences, analysis of protein and nucleic acid sequences by various software packages.
4. The students will be able to predict the secondary and tertiary structures of protein sequences.

Text Books / Reference books

1. T K Attwood, D J parry-Smith, Introduction to Bioinformatics, Pearson Education, 1st Edition,

11th Reprint 2005.

2. S. Ignacimuthu, S.J., Basic Bioinformatics, Narosa Publishing House, 1995.

3. Neil C. Jones and Pavel A. Pevzner, An Introduction to Bioinformatics Algorithms, MIT Press,

First Indian Reprint 2005.

4. Harshawardhan P BAL, Perl Programming for Bioinformatics, Tata McGraw Hill, 2003.

Note: Latest and additional good books may be suggested and added from time to time.

BCA-DS-315: BIG DATA –II LAB
BCA VI Semester

No. of Credits: 2	Sessional:	25 Marks
L T P Total	Practical:	50 Marks
0 0 4 4	Total :	75 Marks
	Duration of Exam:	3 Hours

List of Programs:

I) Working with Pig using various built-in and user-defined functions.

1. Installation of Cloud era to work with Pig.

2. Execute various commands and queries.

3. Loading data

4. Produce histogram.

5. Sum the word counts for each word length using the SUM function with the FOREACH GENERATE command.

6. Copy the data file into HDFS.

7. Importing CSV files, creating tables in pig.

8. Facebook analysis by installing plug-into see the details like: liked pages of friends etc. and many other activities.

II) Working with Hive

1. Installation of Hive

2. Introduction to Hive shell

3. Making a script file in Hive.

4. Loading dataset, loading tables.

5. Creating tables, creating databases, editing tables.

6. Apply various queries like:

7. Different ways of querying through Interactive shell window or using Hive script.

III) Introduction to Spark framework and Spark installation

1. Understand the purpose of SparkContext.
2. Initialize Spark with the various Programming Languages.
3. Passing functions to Spark and run some Spark Examples.

BCA-DS-316: MACHINE LEARNING –II LAB
BCA VI Semester

No. of Credits: 2
  L T   P   Total
0 0   4   4

Sessional: 25 Marks
Practical: 50 Marks
Total :   75 Marks
Duration of Exam: 3 Hours

List of Programs:
1. Compute the accuracy of different classifiers (KNN, Naïve Bayes and SVM), considering a test data set.
2. Develop a machine learning method to classify your incoming mail.
3. Develop a machine learning method to Predict stock prices based on past price variation.
4. Estimate the precision, recall, accuracy, and F-measure of the decision tree classifier on the text classification task.
5. Write a program to implement Random Forest Approach using sample dataset.
6. Write a program to implement dimensionality reduction using Principal component Analysis.
7. Build an Artificial Neural Network by implementing the Backpropagation algorithm and test the same using appropriate data sets.
8. Develop a machine learning method to implement recommendation system to rate movies, books, etc.