Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore

Shri Vaishnav Institute of Computer Applications

Name of the Program: BSC (Data Science)

COURSE CODE	CATEGORY	COURSE NAME	L	T	P	CREDITS	TEACHING & EVALUATION SCHEME				
							THEORY			PRACTICAL	
							END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*
BCABDA402	Major	Big Data Technologies (Hadoop)	2	0	2	3	60	20	20	30	20

Legends: L – Lecture; T – Tutorial/Teacher Guided Student Activity; P – Practical; Q/A – Quiz/Assignment/Attendance; MST – Mid Semester Test.

***Teacher Assessment** shall be based on following components: Quiz/Assignment/Project/Participation in class activities, given that no component shall exceed more than 10 marks

Course Education Objectives (CEOs): The objective of this course is to teach students about:

1. To familiarize the students with the need and scope of the subject.
2. Provide an exposure giving a strong foundation to the data analytics practices.
3. Big data computing technologies, Watson studio
4. Understand the challenges posed by distributed applications and how ZooKeeper is designed to Handle.
5. create a basis for the use of advanced investigative and computational methods to convert information to useful knowledge.
6. Develop an understanding of how business analytics is actually performed
7. Develop an understanding of the complete open-source Hadoop ecosystem and its near term future direction
8. Covers foundational techniques and tools required for data science and big data analytics like Hadoop, NoSQL MapReduce, BIGSQL, Watson studio.

Course Outcomes (COs): After the completion of the course the student will be able to:

1. understanding importance and scope of the subject
2. understand the technological foundations of the Big Data
3. differentiate the structured and unstructured data and different databases to store structured and unstructured data like Mongo DB, HBase etc.
4. store and manipulate the different types of data
5. have a knowledge of HADOOP and Hadoop ecosystem and its uses in Big Data
6. understand and apply the Big Data Analytics
7. Understand the functions and features of HDP.

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore

Shri Vaishnav Institute of Computer Applications

Name of the Program: BSC (Data Science)

COURSE CODE	CATEGORY	COURSE NAME	L	T	P	CREDITS	TEACHING & EVALUATION SCHEME				
							THEORY			PRACTICAL	
							END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*
BCABDA402	Major	Big Data Technologies (Hadoop)	2	0	2	3	60	20	20	30	20

8. To model and Design Big Data for analytics
9. Understand the MapReduce model v1 and review java code.
10. understand the professional and ethical responsibility
11. to produce the good decision makers who can use empirical approaches, wide range of data analytic techniques to problem solving.

PRE- REQUISITES:

This course requires the familiarity with linear algebra, calculus, matrix operations, probability theory, statistics, programming, Database Management System, Data Mining and Warehousing

Syllabus

Unit I

Big Data: Introduction and basics, Evolution of Data Management, Definition, Importance, Big Data Types, Structured and unstructured Data, Sources of bigstructured data and unstructured data, Architecture of Big Data Management System, Stages of Big Data Management,

Big Data Technology Foundations: Technology Components, virtualization, distributed computing, Cloud and Big Data, Integration of data types into a big data environment.

Unit II

Big Data and operational Databases: relational, non relational, key-value pair, document, column oriented, graph, spatial databases, MapReduce, Hadoop, Hadoop Foundation and Ecosystem, Appliances and Big Data Warehouse, Big data Implementation, Big Data Applications. roles in building a complete big data, solution to common business problems.

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore

Shri Vaishnav Institute of Computer Applications

Name of the Program: BSC (Data Science)

COURSE CODE	CATEGORY	COURSE NAME	L	T	P	CREDITS	TEACHING & EVALUATION SCHEME				
							THEORY			PRACTICAL	
							END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*
BCABDA402	Major	Big Data Technologies (Hadoop)	2	0	2	3	60	20	20	30	20

Hadoop and HDFS: need for a big data strategy in terms of parallel reading of large data files and internode network speed in a cluster, Hadoop Distributed File System (HDFS), function of the NameNode and DataNodes in a Hadoop cluster, files are stored and blocks ("splits") are replicated. Hive, Sqoop.

Unit III

Big Data Analytics: Introduction, Basic and Advanced Analytics, Drivers, Pillars of Analytics: descriptive, predictive and prescriptive. Core Components of analytical data architecture, Performance issues, Parallel vs. distributed processing, Shared nothing data architecture and Massive parallel processing, Elastic scalability, Data loading patterns.

Data Analytics lifecycle: Discovery, Data Preparation, Model Planning, Model Building, Communicating results and findings.

Unit IV

Apache Ambari: The purpose of Apache Ambari in the HDP stack, architecture of Ambari and Ambari' relation to other services and components of a Hadoop cluster, components of Ambari and their functions, initiating start and stop services from Ambari Web Console.

Introduction of Hortonworks Data Platform – HDP: functions and features of HDP, the IBM value-add components, basics of IBM Watson Studio, description of the purpose of the value-add components

Unit V

MapReduce and YARN: MapReduce model v1, the limitations of Hadoop 1 and MapReduce, handling the Mapper class with Java, Reducer class and the program driver needed to access MapReduce, the YARN model, Hadoop 2/YARN versus Hadoop 1

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore

Shri Vaishnav Institute of Computer Applications

Name of the Program: BSC (Data Science)

COURSE CODE	CATEGORY	COURSE NAME	L	T	P	CREDITS	TEACHING & EVALUATION SCHEME				
							THEORY			PRACTICAL	
							END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*
BCABDA402	Major	Big Data Technologies (Hadoop)	2	0	2	3	60	20	20	30	20

ZooKeeper, Slider, and Knox: The challenges with distributed applications and how ZooKeeper is designed to handle them, ZooKeeper: role within the Apache Hadoop infrastructure, the realm of Big Data management, real-world applications, services to manage distributed systems, ZooKeeper CLI and ZooKeeper services.

TEXT BOOKS:

1. "Introduction to InfosphereBigInsights", IBM Career Education
2. "Changing Business with Data Insight", IBM Career Education
3. "Big Insights Analytics for Business Analysts", IBM Career Education
4. "Hadoop: The Definitive Guide Paperback – 2015" by Tom White, Shroff Publishers & Distributers Private Limited - Mumbai; Fourth edition (2015).
5. "Big Data and Hadoop" by V. K. Jain, Khanna Publishers; 1 edition (1 June 2015)
6. "Big Data for Dummies" by Judith Hurwitz, Alan Nugent, Dr. Fern Halper and Marcia Kaufman Wiley, ISBN: 978-1- 118-50422-2, 2013
7. "Data Analytics, Models, Algorithms for Intelligent Data Analysis by Thomas A., Springer Vieweg , ISBN 978-3-8348-2589-6, 2013
8. "HadoopFor Dummies", Dirk deRoos, For Dummies, 2014

REFERENCE BOOKS:

1. "Big Data: A Revolution That Will Transform How We Live, Work, and Think" by Kenneth Cukier, Viktor Mayer-Schönberger; Mariner Books; Edition (2014)
2. **"**Big Data: Using Smart Big Data, Analytics and Metrics to Make Better" by Bernard Marr; Wiley; Edition 1st (2015)
3. "Mining of Massive Datasets, Chapter 2", by Ullman, Rajaraman,
4. "MapReduce and Parallel DBMS" by Stonebrakeret al.,: Friends or Foes?, Communications of the ACM, January 2010.
5. "MapReduce: A Flexible Data Processing Tool" by Dean and Ghemawat, Communications of the ACM, January 2010.

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore

Shri Vaishnav Institute of Computer Applications

Name of the Program: BSC (Data Science)

COURSE CODE	CATEGORY	COURSE NAME	L	T	P	CREDITS	TEACHING & EVALUATION SCHEME				
							THEORY			PRACTICAL	
							END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*
BCABDA402	Major	Big Data Technologies (Hadoop)	2	0	2	3	60	20	20	30	20

List of Practical:

1. Installation and configuring Hadoop, configuring HDFS,
2. Installation of Zookeeper, Pig, Sqoop and Hbase
3. Running jobs on Hadoop
4. Working on HDFS
6. Hadoop streaming
7. Creating Mapper function using python.
8. Creating Reducer function using python
9. Python iterator and generators
10. Twitter data sentimental analysis using Flume and Hive
11. Business insights of User usage records of data cards
12. Wiki page ranking with Hadoop
13. Health care Data Management using Apache Hadoop ecosystem

Chairperson Board of Studies Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore	Chairperson Faculty of Studies Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore	Controller of Examinations Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore	Joint Registrar Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore

Shri Vaishnav Institute of Computer Applications

Name of the Program: BSC (Data Science)

COURSE CODE	CATEGORY	COURSE NAME	L	T	P	CREDITS	TEACHING & EVALUATION SCHEME				
							THEORY			PRACTICAL	
							END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*
BCABDA404	Major	Basics of Computer Networks	3	0	0	3	60	20	20	0	0

Legends: L - Lecture; T - Tutorial/Teacher Guided Student Activity; P – Practical; C - Credit;
Q/A – Quiz/Assignment/Attendance, MST - Mid Sem Test.
***Teacher Assessment** shall be based on following components: Quiz/Assignment/Project/Participation in class (Given that no component shall be exceed 10 Marks)

Course Educational Objectives (CEOs):

- To provide an introduction to the fundamental concepts on data communication and the design of computer networks.
- To get familiarized with the basic protocols of computer networks.

Course Outcomes (COs): After the successful completion of this course students will be able to

- Identify the different components in a Communication System and their respective roles.
- Describe the technical issues related to the local Area Networks
- Identify the common technologies available in establishing LAN infrastructure.

UNIT-I

Introduction: Computer Network, Data communication, Network Topologies, Layered Network Architecture-Review of ISO-OSI Model., Transmission Media: Guided and unguided.

UNIT-II

Data Security and Integrity: Parity Checking Code, Cyclic redundancy checks (CRC), Hemming Code, Flow and error control, Go-Back-N protocol, sliding window protocol. Contention Protocol- Stop-Go-Access Protocol.

UNIT-III

Data Link Layer: Simplex, Half duplex and Full duplex, Inter Networking, Layer 1 connections-Repeater, Hubs, Layer 2 connections-Bridges, Switches, Layer 3 connections-Routers, Gateways.

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore

Shri Vaishnav Institute of Computer Applications

Name of the Program: BSC (Data Science)

COURSE CODE	CATEGORY	COURSE NAME	L	T	P	CREDITS	TEACHING & EVALUATION SCHEME				
							THEORY			PRACTICAL	
							END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*
BCABDA404	Major	Basics of Computer Networks	3	0	0	3	60	20	20	0	0

UNIT-IV

Wide Area Network: Introduction, Network routing, least cost routing algorithms, Dijkstra's algorithm, Internetworking.

UNIT-V

Transport and upper layers in OSI Model: Transport layer functions, Network Security, email, Multimedia.

Text Books:

1. A.S.Tanenbaum, "Computer Network", 4thaddition,PHI
2. Forouzan "Data Communication and Networking 3ed", TMH
3. J.F.Hayes, "Moduling and Analysis of Computer Communication Networks", Plenum Press
4. D.E.Comer, "Internetworking with TCP/IP", Volume Ist&IInd, PHI
5. Willium Stalling, "Data & Computer communications", Maxwell Macmillan International Ed.
6. D.Bertsekas and R.Gallager,"Data Networks", 2ndEd. ,PHI.
7. G.E. Keiser,"Local Area Networks ", McGraw Hill, International Ed.

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore

Shri Vaishnav Institute of Computer Applications

Name of the Program: BSC (Data Science)

COURSE CODE	CATEGORY	COURSE NAME	L	T	P	CREDITS	TEACHING & EVALUATION SCHEME				
							THEORY			PRACTICAL	
							END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*
BSCDSMA401	DCC	Time Series Analysis	3	0	0	3	60	20	20	0	0

Course Objective

To introduce the students to the fundamentals of the Time Series Analysis.

Course Outcomes

After the successful completion of this course students will be able to:

1. *Forecast the trend pattern exhibited by the given data by using various methods.*
2. *Run and interpret time series models and regression models for time series.*
3. *Analyze and estimate the cyclic components using special processes.*

Course Content:

UNIT – I

Introduction, component of time series: - trend, periodic changes, irregular component. Analysis of time series: - mathematical models for time series, uses of time series.

UNIT– II

Measurement of trend: - graphic method, method of semi averages, method of curve fitting by principle of least square, growth curves and their fitting, moving average method, approximation to moving averages.

UNIT – III

Measurement of seasonal variation: - method of simple averages, ratio to trend method, ratio to moving average method, link relative method, de-seasonalisation of data. Method of cyclic variation.

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore

Shri Vaishnav Institute of Computer Applications

Name of the Program: BSC (Data Science)

COURSE CODE	CATEGORY	COURSE NAME	L	T	P	CREDITS	TEACHING & EVALUATION SCHEME				
							THEORY			PRACTICAL	
							END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*
BSCDSMA401	DCC	Time Series Analysis	3	0	0	3	60	20	20	0	0

UNIT – IV

Auto regression series: - first order auto regression series (markoff's series), second order auto regression series (yule's series), general auto regression.

UNIT – V

Study of the stationary processes: (1) moving average (ma), (2) auto regressive (ar), (3) Auto regressive moving averages (arma) and (4) Auto Regressive integrated moving Averages (arima) models.

SUGGESTED READING:

1. S.C. Gupta and V.K. Kapoor, Fundamentals of applied Statistics, S Chand & Co.
2. Jim Frost, Introduction to Statistics: An Intuitive Guide for Analyzing Data and Unlocking Discoveries, Jim Frost MS.
3. Douglas C. Montgomery, Cheryl L. Jennings and Murat Kulahci. Introduction to Time Series Analysis and Forecasting.
4. James D. Hamilton. Time Series Analysis.

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore

Shri Vaishnav Institute of Computer Applications

Name of the Program: BSC (Data Science)

COURSE CODE	CATEGORY	COURSE NAME	L	T	P	CREDITS	TEACHING & EVALUATION SCHEME				
							THEORY			PRACTICAL	
							END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*
BSMAT 402	DCC	Numerical Methods	3	0	0	3	60	20	20	0	0

Course Objective

To introduce the students with the fundamentals of the Numerical Methods.

Course Outcomes

After the successful completion of this course students will be able to:

1. *Apply various numerical methods to obtain numerical solution of definite integration and algebraic and transcendental equations.*
2. *Learn various techniques of getting solution of linear system of equation.*
3. *Use various techniques of interpolation methods.*
4. *Learn various techniques of numerical differentiation and numerical Integration.*
5. *Find numerical solution of ordinary differential equation.*

Course Content:

UNIT – I

Numerical errors, Numerical methods for solving nonlinear equations: Method of bisection, Secent method, False position method, Newton-Raphson method, Fixed point method.

UNIT – II

Linear Equations: Direct Methods for Solving Systems of Linear Equations (Gauss-Jordan Elimination, LU Decomposition), Iterative Methods (Jacobi, Gauss- Seidel Reduction Methods).

Shri Vaishnav Vidyapeeth Vishwavidyalaya, Indore

Shri Vaishnav Institute of Computer Applications

Name of the Program: BSC (Data Science)

COURSE CODE	CATEGORY	COURSE NAME	L	T	P	CREDITS	TEACHING & EVALUATION SCHEME				
							THEORY			PRACTICAL	
							END SEM University Exam	Two Term Exam	Teachers Assessment*	END SEM University Exam	Teachers Assessment*
BSMAT402	DCC	Numerical Methods	3	0	0	3	60	20	20	0	0

UNIT – III

Introduction to Interpolation, Calculus of finite differences, Difference Operators, Relation between Operators, Interpolation with equal interval: Newton's Forward and Backward Interpolation formula.

UNIT – IV

Interpolation with unequal interval: Newton divided difference interpolation formula and Lagrange's interpolation Formula, Lagrange's Inverse interpolation Formula.

UNIT – V

Numerical Differentiation and Integration: Discrete Approximation of Derivatives: Forward and Backward Difference Forms, Numerical Integration: Simson's 1/3, Simson's 3/8. Ordinary Differential Equations: Euler Method, Runge¬ Kutta's Method.

Reference Book:

1. S. S. Sastry, Introductory Methods of Numerical Analysis, P11.1 Learning Pvt. Ltd.
2. Balaguruswamy, Numerical Methods. Tata McGraw Hill Publication, New York.
3. Numerical Methods: R.K. Jain, S.R.K. Iyenger, New Age International Publisher.
4. Higher Engineering Mathematics: B.S. Grewal, Khanna Publisher