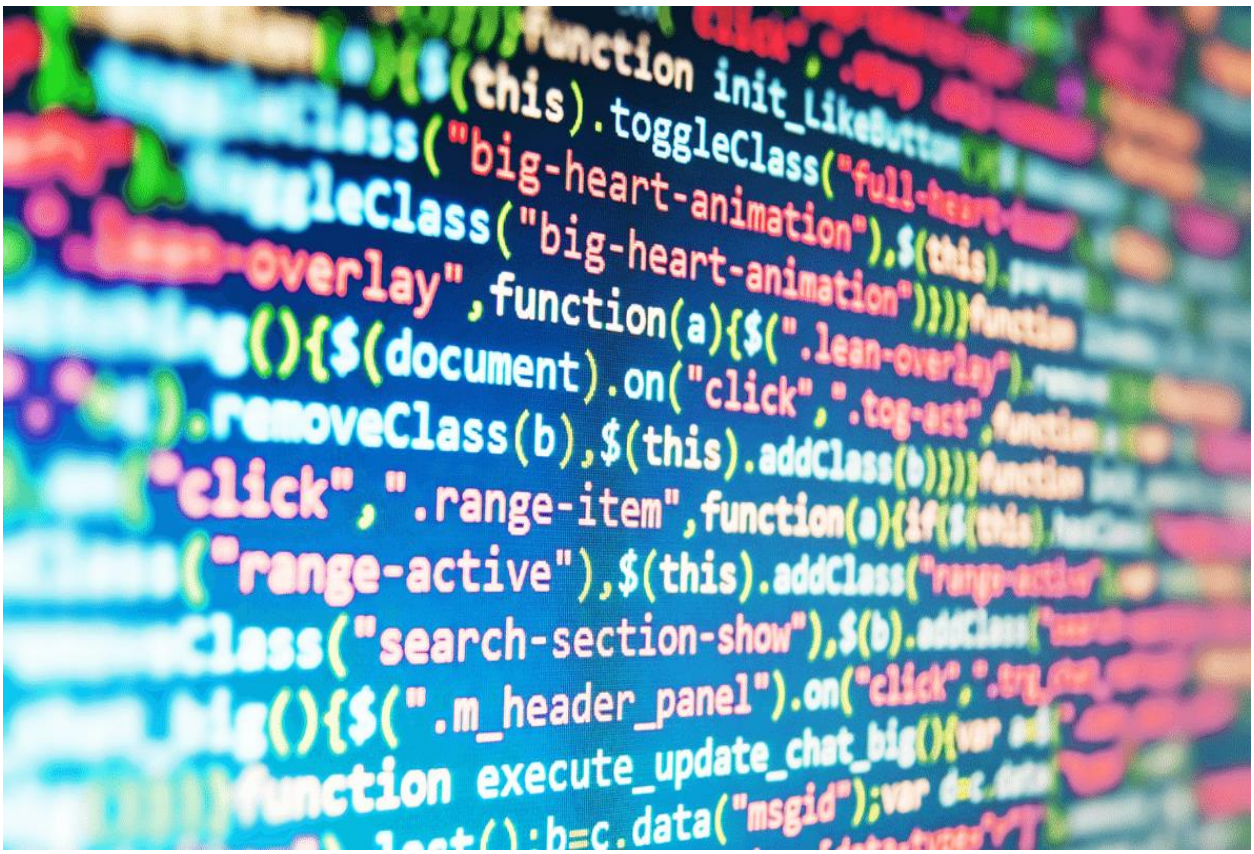


Detailed Syllabus for Computer Science & Engineering with Specialization in Big Data Analytics

UEM
श्रद्धावान लभते ज्ञानम्
UNIVERSITY OF ENGINEERING & MANAGEMENT
Good Education, Good Jobs

DEPT. OF COMPUTER SCIENCE & ENGINEERING

UNIVERSITY OF ENGINEERING & MANAGEMENT, JAIPUR

PREAMBLE

Education plays an enormously significant role in the building of a nation. There are quite a large number of educational institutions, engaged in imparting education in our country. Majority of them have entered recently into semester system to match with international educational pattern. However, our present education system is churning out youth who have to compete locally, regionally, nationally as well as globally. The present alarming situation necessitates transformation and/or redesigning of system, not only by introducing innovations but developing "learner-centric approach.

Majority of Indian higher education institutions have been following marks or percentage-based evaluation system, which obstructs the flexibility for the students to study the subjects/courses of their choice and their mobility to different institutions. There is need to allow the flexibility in education system, so that students depending upon their interests can choose inter-disciplinary, intra-disciplinary and skill-based courses. This can only be possible when choice based credit system (CBCS), an internationally acknowledged system, is adopted. The choice based credit system not only offers opportunities and avenues to learn core subjects but also explore additional avenues of learning beyond the core subjects for holistic development of an individual. The CBCS will undoubtedly facilitate benchmarking of our courses with best international academic practices.

Advantages of the choice based credit system:

•Shift in focus from the teacher-centric to student-centric education.

•Student may undertake as many credits as they can cope with (without repeating all courses in a given semester if they fail in one/more courses).

•CBCS allows students to choose inter-disciplinary, intra-disciplinary courses, skill oriented papers (even from other disciplines according to their learning needs, interests and aptitude) and more flexibility for students.

•CBCS makes education broad-based and at par with global standards. One can take credits by combining unique combinations.

•CBCS offers flexibility for students to study at different times and at different institutions to complete one course (ease mobility of students). Credits earned at one institution can be transferred to another institution.

CHOICE BASED CREDIT SYSTEM

The Indian Higher Education Institutions have been moving from the conventional annual system to semester system. Currently many of the institutions have already introduced the Choice Based Credit System. The semester system accelerates the teaching-learning process and enables vertical and horizontal mobility in learning. The credit based semester system provides flexibility in designing curriculum and assigning credits based on the course content and hours of teaching. The Choice Based Credit System provides a 'cafeteria' type approach in which the students can take courses of their choice, learn at their own pace, undergo additional courses to acquire more than the required credits and adopt an interdisciplinary approach to learning.

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

It has been a necessity to align higher education with the emerging needs of the economy so as to ensure that the graduates of higher education system have adequate knowledge and skills for employment and entrepreneurship since last few years. The higher education system has to incorporate the requirements of various industries in its curriculum, in an innovative and flexible manner while developing a well-groomed graduate. CSE department aims to encourage research and innovation in the field of Machine Learning, Cyber security, Artificial Intelligence and other allied areas such as Computational Theory, Cloud Computing, Blockchain Technology, Data Science, Big Data Analytics and many more. The objective of the Computer Science & Engineering Programme with Specialization in Big Data Analytics is to prepare students to undertake careers involving innovative technologies, develop a problem solving capability, or to opt for advanced studies for research oriented careers.

In order to give due importance to practical applications as well as theoretical aspects of CSE, the curriculum for the Computer Science & Engineering Programme with Specialization in Big Data Analytics covers most of the foundational aspects as well as develops engineering skills for problem solving.

JOB OPORTUNITIES

Booming IT sector in India has plenty of jobs for fresh computer science graduates. Candidates with a high percentage of mark and good communication skills as well as sound computer knowledge do not face problem in getting a job. Computer engineers can get jobs in non-IT companies like universities, research, private and public industries, government departments, business organizations, commercial organizations and the manufacturing sector, etc. Besides the Computer Engineers have plenty of options to work in IT companies in departments such as design, development, assembly, manufacture, and maintenance, etc. Software Developers: Software developers are professionals who are concerned with facets of the software development process which involves activities such as design and coding, computer programming, project management, etc. Big data allows organisations to understand the people they serve, choose better strategies, allocate their resources more effectively and operate smarter. For instance, right now big data analytics is playing an important part in understanding cyber security trends.

PROGRAMME EDUCATIONAL OBJECTIVES (PEO)

PEO 01: High Quality Engineering Design and Development Work: Graduates of the program will engage in the effective practice of computer science and engineering to identify and solve important problems in a diverse range of application areas.

PEO 02: Real Life Problem Solving: To educate students with proficiency in core areas of Computer science & Engineering and related engineering so as to comprehend engineering trade-offs, analyse, design, and synthesize data and technical concepts to create novel products and solutions for the real life problems.

PEO 03: Leadership: Graduates of the program will engage in successful careers in industry, academia and attain positions of importance where they have impact on their business, profession and community.

PEO 04: Lifelong Learning: Graduates of the program will adapt to contemporary technologies, tools and methodologies to remain at the frontier of computer science and engineering practice with the ability to respond to the need of a challenging environment.

PROGRAM OUTCOME (PO)

PO	Summary	Description
PO1	**Engineering knowledge**	Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
PO2	**Problem analysis**	Identify, formulate, research literature, and analyse complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
PO3	**Design /development of solutions**	Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
PO4	**Conduct investigations of complex problems**	Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
PO5	**Modern tool usage**	Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modelling to complex engineering activities with an understanding of the limitations.
PO6	**The engineer and society**	Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities

		relevant to the professional engineering practice.
PO7	**Environment and sustainability**	Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.
PO8	**Ethics**	Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
PO9	**Individual and team work**	Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
PO10	**Communication**	Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
PO11	**Project management and finance**	Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.
PO12	**Life-Long Learning**	Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change

TYPES OF COURSES

1. Courses are the subjects that comprise the Computer Science & Engineering Programme with Specialization in Big Data Analytics.
2. A course may be designed to comprise lectures, tutorials, laboratory work, fieldwork, outreach activities, project work, vocational training, viva, seminars, term papers, assignments, presentations, self-study etc. or a combination of some of these components.
3. The learning outcomes of each course will be defined before the start of a semester.
4. Following are the course types:

i. **Core Course (PCC):** This is a course, which is to be compulsorily studied by a student as a core requirement to complete the requirement of B.Tech in Computer Science & Engineering with Specialization in Big Data Analytics.

ii. **Elective Course:** An elective course is a course, which can be chosen from a pool of courses. It is intended to support the discipline of study by providing an expanded scope, enabling exposure to another discipline/domain and nurturing a student's proficiency and skill. An elective may be of following types:

a) **Discipline Elective Courses (DE):** It is an elective course that adds proficiency to the students in the discipline.

b) **Open Elective Courses (OE):** It is an open elective course taken from other engineering disciplines and enhances the generic proficiency and interdisciplinary perspective of students.

c) **Specialization Elective Courses (SEC):** This is a course, which is to be compulsorily studied by a student as a core requirement to complete the requirement of B.Tech in Computer Science & Engineering with Specialization in Big Data Analytics.

iii. **Obligatory Courses:**

a) **Mandatory Courses (MC)**: It can be taken from among a pool of foundation courses, which aim at value-based education. They may provide hands-on training to improve competencies and skills or provide education on human, societal, environmental and national values.

b) **Internship/Training/Project/Dissertation (INT):** Course designed to acquire special/advanced knowledge, such as supplement study/support study to a project work, and a candidate studies such a course on his own with an advisory support by a teacher/faculty member is called dissertation/project

c) **Humanities, Social Sciences & Management (HSM):** It is an elective course taken from non-engineering disciplines (humanities, social sciences and management) that broadens the perspective of an engineering student.

d) **Basic Science Courses (BSC):** It is based upon content that leads to fundamental knowledge enhancement in sciences, and basic engineering principles.

e) **Engineering Science Courses (ESC):** It is based upon content that leads to fundamental knowledge enhancement in basic Engineering Principles.

f) **NPTEL (NPT):** "Essential Studies for Professionals Skill & Skill Development for Professionals" courses designed to encourage and enrich the students for the technical and professional exams.

g) **General Studies Courses (GSC):** "Essential Studies for Professionals Skill & Skill Development for Professionals" courses designed to encourage and enrich the students for the technical and professional exams.

h) Mandatory Additional Requirements (MAR): A student has to do the following things to achieve the MAR points: The student should engage herself / himself in activities outside the curriculum. Join different types of Clubs of NSCBIP, write something for the wall magazine, remain active in outer society, participate in Tech Fests activities, etc.

5. Each credit course contributes certain credits to the programme. A course can be offered either as a full course (4 credits) or as a half course (2 credits). A full course is conducted with 3 hours of lectures and either 1 hour of tutorial or 2 hours of practical work per week. A half course is conducted with 2 hours of lectures. There are also some exceptional electives with 3 credits and 1 credit.

Definition of Credit:

1 Hr. Lecture (L) per week	1 Credit
1 Hr. Tutorial (T) per week	1 Credit
1 Hr. Practical (P) per week Or 2 Hr. Practical (Lab)/week	0.5 Credits Or 1 Credit

6. A project work/dissertation is considered as a special course involving application of the knowledge gained during the course of study in exploring, analyzing and solving complex problems in real life applications. A candidate completes such a course with an advisory support by a faculty member.
7. **Mandatory Courses** may be offered. They do not carry credits but aim at expanding knowledge or bridging deficiency in knowledge or skill.
8. A course may have pre-requisite course(s) that are given in the Semester-wise Course Allocation scheme.
9. A student can opt for a course only if he/she has successfully passed its pre- requisite(s).
10. A student has to register for all courses before the start of a semester.
11. **Program codes:** The codes for various undergraduate programmes are as follows:
    i. Civil Engineering: CE
    ii. Computer Science & Engineering: CS
    (Specialization in Big Data Analytics (CSBDA))
    iii. Electronics and Communication Engineering: EC
    iv. Electrical Engineering: EE
    v. Mechanical Engineering: ME
12. **Departmental Course Codes:** The codes for departmental core courses and discipline-specific electives are specific to each discipline. The first two characters are derived from departmental codes listed above. The third character is 'C' for core courses and 'D' for discipline-specific courses and 'INT' for Dissertation/Project/Training/Internship. This is followed by a digit sequence number:
    i. CSCyyy: Core Course
    ii. CSDyyy: Discipline-Specific Elective Courses

iii. BDAyyy: Specialization Elective Courses

iv. XXXyyy: Open Elective Courses (Depends on the respective Dept.)

v. INTyyy: Project/Training/Internship/ Dissertation

13. **Common Elective Course Codes:** All disciplines will follow a common code as shown below. The 3-digit sequence number 'yyy' is taken from the respective tables of different types of courses.

i. HSMyyy: Humanities, Social Sciences & Management Course

ii. BSCyyy: Basic Science Course

iii. MCyyy: Mandatory Course

iv. GSCyyy: General Studies Courses

v. MARyyy: Mandatory Additional Requirements

Here, yyy will be follow by a sequence of digit.

14. **General Electives:** A student may take a course under the category of General Elective (GE) offered by any other Department of the Institute under the categories of Core Course (CC) and Discipline Specific Electives (DE). However, such options shall be offered to a student as per prescribed guidelines of the Institute.

15. The opting of a course by the student will depend upon the requisites for that course and with the consent of the course advisor.

PROCESS FOR DESIGNING THE PROGRAM CURRICULUM

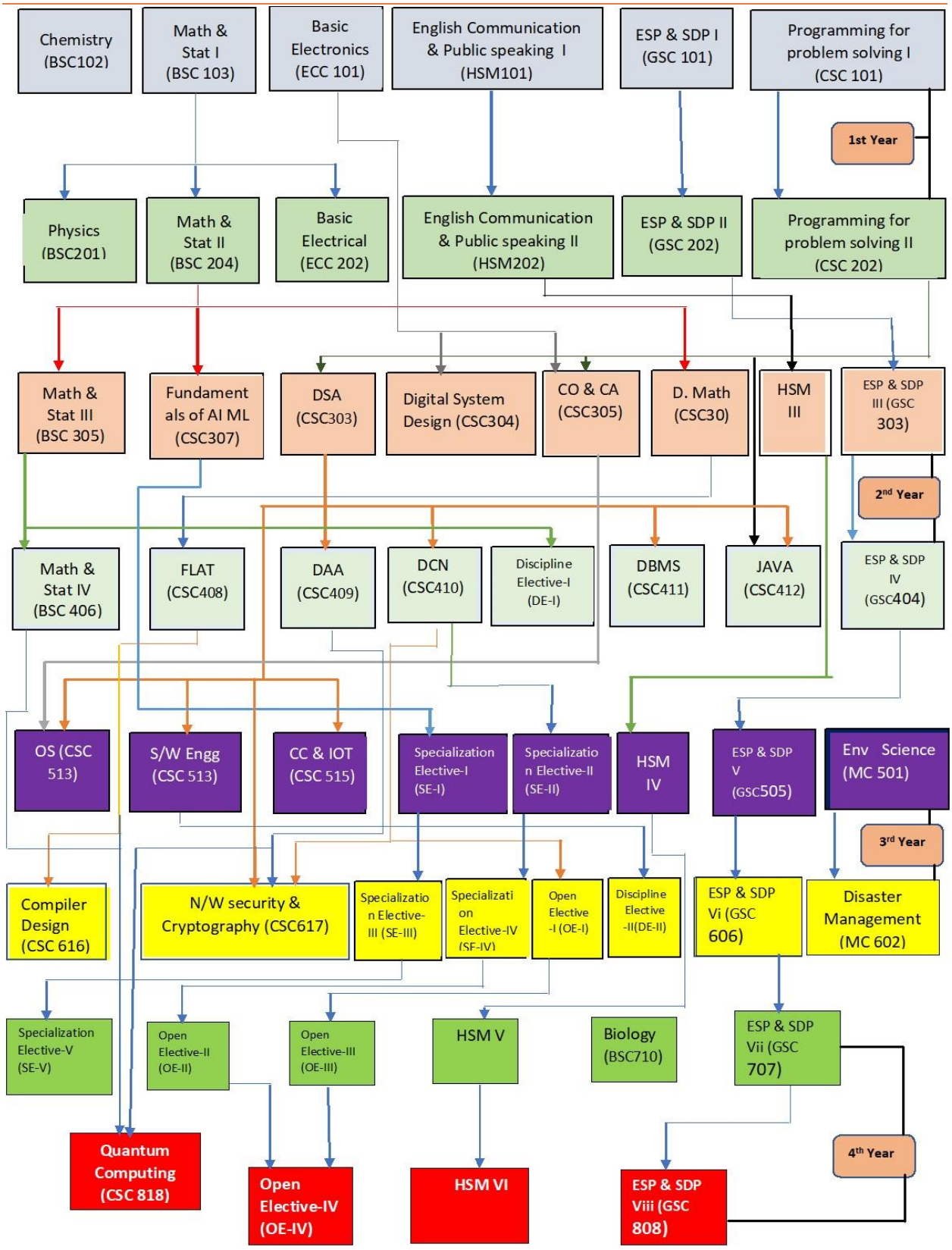
The process for designing the program curriculum involves consideration of the following aspects:

i) Feedback from stakeholders

ii) Institutional and Departmental Vision, Mission and PEOs

iii) Recent trends in industry and research

PREREQUISITE TREE



SCHEME – SEMESTER WISE COURSE ALLOCATION

First Semester Syllabus

Sl No.	Type	Subject Code	Topic	L	T	P	S	Credit Points
1.	BSC	BSC101	Chemistry	3	1	3	0	5.5
2.	BSC	BSC102	Mathematics-I (Calculus & Linear Algebra)	3	1	0	0	4
3.	ESC	ECC101	Basic Electronics Engineering	2	0	2	0	3
4.	ESC	MEC101	Engineering Graphics & Design	1	0	2	0	2
5.	ESC	CSC101	Programming for Problem Solving-I (C)	2	0	0	2	2
6.	HSM	HSM101	English Communication & Public Speaking Skills-I	0	0	2	0	1
7.	GSC	GSC101	ESP & SDP-I	2	0	0	2	2
			Total	13	2	9	4	19.5/28

#Students will undergo a mandatory Induction Program

Second Semester Syllabus

Sl No.	Type	Subject Code	Topic	L	T	P	S	Credit Points
1.	BSC	BSC203	Physics (Semi-Conductor Physics)	3	1	3	0	5.5
2.	BSC	BSC204	Mathematics-II(Probability & Statistics)	3	1	0	0	4
3.	ESC	EEC202	Basic Electrical Engineering	2	0	2	0	3
4.	ESC	MEC202	Workshop & Manufacturing Practices	1	0	2	0	2
5.	ESC	CSC202	Programming for Problem Solving-II (Python)	2	0	0	2	2
6.	HSM	HSM202	English Communication & Public Speaking Skills-II	1	0	3	0	2
7.	GSC	GSC202	ESP & SDP-II	2	0	0	2	2
8.	NPT	NPT201	(NPTEL/MOOCs)	-	-	-	-	2
			Total	15	0	9	4	22.5/28

#(NPT201)NPTEL/MOOCs are based on the respective year's offered courses.

Third Semester Syllabus

Sl No.	Type	Subject Code	Topic	L	T	P	S	Credit Points
1.	BSC	BSC305	Mathematics-III (Differential Calculus)	2	0	0	0	2
2.	PCC	CSC303	Data Structure & Algorithms	3	0	3	0	4.5
3.	ESC	ECS303	Analog Electronic Circuits	2	0	3	0	3.5
4.	PCC	CSC304	Digital Electronics	3	0	3	0	4.5
5.	PCC	CSC305	IT Workshop (Sci Lab/MATLAB)	1	0	4	0	3
6.	HSM	HSM---	Humanities-I	3	0	0	0	3
7.	GSC	GSC303	ESP & SDP-III	2	0	0	2	2
8.	MAR	MAR381	Mandatory Additional Requirements (MAR)	0	0	0	1	0.5
9.	NPT	NPT302	(NPTEL/MOOCs)	-	-	-	-	2
			Total	17	0	13	3	25/33

#(NPT302): NPTEL/MOOCs are based on the respective year's offered courses.

Suggestive Choice Based Subjects

Sl No	Type	Subject Code	Topic	L	T	P	Credit Points
1.	HSM	**HSM303**	Organizational Behavior	3	0	0	3
2.	**HSM**	**HSM304**	**Values and Ethics in Profession**	3	0	0	3
3	HSM	HSM305	Industrial Psychology	3	0	0	3

Fourth Semester Syllabus

Sl No.	Type	Subject Code	Topic	L	T	P	S	Credit Points
1.	PCC	CSC406	Discrete Mathematics	3	0	0	0	3
2.	PCC	CSC407	Operating System	3	0	3	0	4.5
3.	PCC	CSC408	Design & Analysis of Algorithms	3	0	3	0	4.5
4.	PCC	CSC409	Computer Organization & Architecture	3	0	3	0	4.5
5.	ESC	ECS404	Signals & System	3	0	0	0	3
6.	HSM	HSM---	Management-I	3	0	0	0	3
7.	GSC	GSC404	ESP & SDP-IV	2	0	0	2	2
8.	MC	MC401/402	Environmental Sciences/Disaster Management	0	0	0	2	0
9.	MAR	MAR484	Mandatory Additional Requirements (MAR)	0	0	0	1	0.5
10.	NPT	NPT403	(NPTEL/MOOCs)	-	-	-	-	2
			Total	20	0	9	5	27/34

#(NPT403): NPTEL/MOOCs are based on the respective year's offered courses.

Suggestive Choice Based Subjects

Sl No.	Type	Subject Code	Topic	L	T	P	Credit Points
1.	HSM	HSM406	Human Resource Development and Organizational Behavior	3	0	0	3
2.	HSM	**HSM407**	**Economics & Financial Accounting**	3	0	0	3
3.	HSM	HSM408	Economics for Engineers	3	0	0	3

Fifth Semester Syllabus

Sl No.	Type	Subject Code	Topic	L	T	P	S	Credit Points
1.	PCC	CSC510	Formal Language & Automata Theory	3	0	0	0	3
2.	PCC	CSC511	Data Base Management System	3	0	3	0	4.5
3.	PCC	CSC512	Object Oriented Programming Using Java	2	0	3	0	3.5
4.	PCC	CSC513	Software Engineering	2	0	2	0	3
5.	PEC/SEC	---	Professional/Specialization Elective-I	3	0	0	0	3
6.	HSM	HSM---	Humanities-II	3	0	0	0	3
7.	MC	MC503	Constitution of India/Essence of Indian Knowledge Tradition	0	0	0	0	0
8.	GSC	GSC505	ESP & SDP-V	2	0	0	2	2
9.	PTI	INT501	Internship/Project-I	0	0	0	1	1
10.	NPT	NPT504	(NPTEL/MOOCs)	-	-	-	-	2
			Total	19	0	6	3	25/28

#(NPT504): NPTEL/MOOCs are based on the respective year's offered courses.

Suggestive Choice Based Subjects

Sl No.	Type	Subject Code	Topic	L	T	P	Credit Points
1.	PEC*	CSP501	Embedded Systems	2	0	2	3
2.	PEC**	CSP502	AI & Machine Learning	2	0	2	3
3.	HSM	HSM509	Industrial Psychology	2	0	0	2
4.	HSM	**HSM510**	**Principle of Management**	2	0	0	2
5.	HSM	HSM511	Total Quality Management	2	0	0	2

Note: Students can opt any Professional Track from 5th Sem Onwards/Big data Analytics Specialization

***Track:** IOT, Cybersecurity & Blockchain Track

****Track:** AI & Machine Learning Track

Sixth Semester Syllabus

Sl No.	Type	Subject Code	Topic	L	T	P	S	Credit Points
1	PCC	CSC614	Compiler Design	3	0	3	0	4.5
2	PCC	CSC615	Computer Networks	3	0	3	0	4.5
3	PEC/SEC	---	Professional/Specialization Elective -II	2	0	2	0	3
4	PEC/SEC	---	Professional/Specialization Elective -III	3	0	0	0	3
5	OE	---	Open Elective-I	2	0	0	0	3
6	DE	CSD---	Discipline Elective-I	2	0	2	0	3
7.	GSC	GSC606	ESP & SDP-VI	2	0	0	2	2
8.	PTI	INT604	Internship/Industrial Training/Project-II	0	0	0	1	1
9.	NPT	NPT605	(NPTEL/MOOCs)	-	-	-	-	2
			Total	17	0	10	3	26/30

#(NPT605): NPTEL/MOOCs are based on the respective year's offered courses.

Suggestive Choice Based Subjects

Sl No.	Type	Subject Code	Topic	L	T	P	Credit Points
1.	PEC*	CSP5603	Blockchain Technology	3	0	0	3
2.	PEC**	CSP604	Soft Computing	3	0	0	3
3.	PEC*	CSP605	AI & Machine Learning	3	0	2	4
4.	PEC**	CSP606	Big Data Analytics	3	0	2	4
5.	OE	BSC607	Numerical Methods & Operation Research	2	0	2	3
6.	OE	BSC608	Operations Research	2	0	2	3
7.	OE	BSC609	Statistics for Data Analysis	2	0	2	3
8.	OE*	CSD601	Blockchain Technology	3	0	0	3
9.	OE**	CSD602	Big Data Analytics	3	0	2	4
10.	DE	CSD603	Web Technology	2	0	2	3
11.	DE	CSD604	Computer Graphics	2	0	2	3
13.	DE	**CSD605**	**Software Project Management**	2	0	2	3
14.	DE	CSD606	E-Commerce	2	0	2	3

Seventh Semester Syllabus

Sl No.	Type	Subject Code	Topic	L	T	P	Credit Points
1	PEC/SEC	---	Professional/Specialization Elective -IV	3	0	0	3
2	PEC/SEC	---	Professional/Specialization Elective -V	3	0	0	3
3.	OE	---	Open Elective-II	3	0	0	3
4.	BSC	BSC710	Biology	2	1	0	3
5.	GSC	GSC707	ESP & SDP-VII	2	0	2	2
6.	PTI	INT705	Internship/Industrial Training/Project-III	0	0	8	4
			Total	13	1	8	18/21

#Students will undergo project/training/internship in the industry / research organization / reputed Institute during the vacation

Suggestive Choice Based Subjects

Sl No.	Type	Subject Code	Topic	L	T	P	Credit Points
1.	PE*	CSP707	Blockchain Technology	3	0	0	3
2.	PE**	CSP708	Natural Language Processing	3	0	0	3
3.	PEC*	CSP709	Digital Forensics	3	0	0	3
4.	PEC**	CSP710	Computer Vision	3	0	0	3
5.	CSD*	CSD707	Cyber Law, IPR & Ethics	3	0	0	3
6.	CSD**	CSD708	Natural Language Processing	3	0	0	3
7.	OE	CSD709	Wireless Sensor Network & Network Security	3	0	0	3
8.	OE	CSD710	Neural Network and Application	3	0	0	3
9.	OE	CSD711	Real Time Operating System	3	0	0	3
10.	OE	CSD712	Distributed System	3	0	0	3
11.	BSC	BSC711	Statistics For Data Analytics	3	0	0	3
12.	BSC	BSC712	Statistical Methods For Decision Making	3	0	0	3
13.	BSC	BSC713	Exploratory Data Analysis	3	0	0	3
14.	BSC	BSC714	Graph Theory	3	0	0	3

Eighth Semester Syllabus

Sl No.	Type	Subject Code	Topic	L	T	P	Credit Points
1.	PEC/SEC	---	Professional/Specialization Elective -VI	3	0	0	3
2.	OE	---	Open Elective-III	3	0	0	3
3.	OE	---	Open Elective-IV	3	0	0	3
4.	GSC	GSC808	ESP & SDP - VIII	2	0	2	2
5.	PTI	INT806	Internship Industrial Training/Project-IV	0	0	12	6
6.	CC	CSC881	Grand Viva	0	0	0	2
			Total	11	0	14	19/25

#Students will undergo project/training/internship in the industry / research organization / reputed Institute during the vacation.

Suggestive Choice Based Subjects

Sl No.	Type	Subject Code	Topic	L	T	P	Credit Points
1.	PEC*	CSP811	Digital Forensics	3	0	0	3
2.	PEC**	CSP812	Deep Learning	3	0	0	3
3.	OE	CSD813	Data Mining & Data Ware Housing	3	0	0	3
4.	OE	CSD814	Information Theory & Coding	3	0	0	3
5.	OE	CSD815	Advanced Algorithms	3	0	0	3
6.	OE	CSD816	Digital Image Processing	3	0	0	3

Professional Elective Courses:

Subject Code	IOT, Cybersecurity & Blockchain Track	Subject Code	AI & Machine Learning Track
CSP501	Embedded Systems	CSP502	AI & Machine Learning
CSP603	Blockchain Technology	CSP504	Soft Computing
CSP605	AI & Machine Learning	CSP606	Big Data Analytics
CSP707	Cyber Security	CSP608	Natural Language Processing
CSP709	Cyber Law, IPR & Ethics	CSP710	Computer Vision
CSP811	Digital Forensics	CSP712	Deep Learning

Specialization Course:

Subject Code	Big Data Analytics
BDA501	Big Data Analytics
BDA602	Big Data Modeling & Management
BDA603	Statistics for Data Analytics
BDA704	Big Data Integration & Modeling
BDA705	Machine Learning With Big Data
BDA806	Graph Analytics for Big Data
BDA807	Managing Big Data with SQL

Note: Refer to Computer Science & Engineering Syllabus for "syllabus of other subject".

TITLE OF COURSE: BIG DATA ANALYTICS
COURSE CODE: BDA501
L-T-P: 3-0-0
CREDITS: 3

Pre-requisite: Should have knowledge of one Programming Language (Java preferably), Practice of SQL (queries and sub queries), exposure to Linux Environment.

Introduction:

The course enables students to Understand the Big Data Platform and its Use cases, Provide an overview of Apache Hadoop, Provide HDFS Concepts and Interfacing with HDFS , Understand Map Reduce Jobs ,Provide hands on Hadoop Eco System ,Apply analytics on Structured, Unstructured Data, Exposure to Data Analytics with R.

Course Outcomes (CO):

The students will be able to:
CO1: Identify Big Data and its Business Implications
CO2: List the components of Hadoop and Hadoop Eco-System
CO3: Access and Process Data on Distributed File System
CO4: Manage Job Execution in Hadoop Environment
CO5: Develop Big Data Solutions using Hadoop Eco System

Mapping of Course Outcomes (CO) and Program Outcomes (PO):

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓	✓	✓								✓
CO2	✓		✓		✓							✓
CO3	✓		✓						✓			✓
CO4	✓		✓		✓				✓	✓		✓
CO5	✓		✓								✓	✓

Course Contents:

Module-1: Introduction To Big Data And Hadoop: Types of Digital Data, Introduction to Big Data, Big Data Analytics, History of Hadoop, Apache Hadoop, Analysing Data with Unix tools, Analysing Data with Hadoop, Hadoop Streaming, Hadoop Echo System, IBM Big Data Strategy, Introduction to Infosphere BigInsights and Big Sheets.

Module-2: HDFS(Hadoop Distributed File System) The Design of HDFS, HDFS Concepts, Command Line Interface, Hadoop file system interfaces, Data flow, Data Ingest with Flume and Scoop and Hadoop archives, Hadoop I/O: Compression, Serialization, Avro and File-Based Data structures.

Module-3: Map Reduce Anatomy of a Map Reduce Job Run, Failures, Job Scheduling, Shuffle and Sort, Task Execution, Map Reduce Types and Formats, Map Reduce Features.

Module-4: Hadoop Eco System Pig: Introduction to PIG, Execution Modes of Pig, Comparison of Pig with Databases, Grunt, Pig Latin, User Defined Functions, Data Processing operators. Hive: Hive Shell, Hive Services, Hive Metastore, Comparison with Traditional Databases, HiveQL, Tables, Querying Data and User Defined Functions. HBase: HBasics, Concepts, Clients, Example, Hbase Versus RDBMS. Big SQL: Introduction

Module-5: Machine Learning: Introduction, Supervised Learning, Unsupervised Learning, Collaborative Filtering. Big Data Analytics with BigR.

Textbooks:

1. Tom White "Hadoop: The Definitive Guide" Third Edit on, O'reily Media, 2012.
2. Seema Acharya, Subhasini Chellappan, "Big Data Analytics" Wiley 2015.

Reference Books:

1. Michael Berthold, David J. Hand, "Intelligent Data Analysis", Springer, 2007.
2. Jay Liebowitz, "Big Data and Business Analytics" Auerbach Publications, CRC press (2013)
3. Tom Plunkett, Mark Hornick, "Using R to Unlock the Value of Big Data: Big Data Analytics with Oracle Enterprise and Oracle R Connector for Hadoop", McGraw-Hill/Osborne Media (2013), Oracle press

TITLE OF COURSE: BIG DATA MODELING & MANAGEMENT
COURSE CODE: BDA602
L-T-P: 3-0-2
CREDITS: 4

Pre-requisite: Introduction about Big data and Hadoop.

Introduction:

In this course, you will experience various data genres and management tools appropriate for each. You will be able to describe the reasons behind the evolving plethora of new big data platforms from the perspective of big data management systems and analytical tools.

Course Outcomes (CO):

The students will be able to:
CO1: Recognize different data elements in your own work and in everyday life problems.
CO2: Explain why your team needs to design a Big Data Infrastructure Plan and Information System Design.
CO3: Identify the frequent data operations required for various types of data.
CO4: Select a data model to suit the characteristics of your data.
CO5: Apply techniques to handle streaming data.

Mapping of Course Outcomes (CO) and Program Outcomes (PO):

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓				✓						✓
CO2	✓		✓	✓								✓

CO3	✓			✓								✓
CO4	✓			✓	✓							✓
CO5	✓	✓							✓			✓

Course Contents:

Module-1: Introduction to Big Data Modeling and Management, Data Ingestion, Data Storage, Data Quality, Data Operations, Data Scalability and Security, Energy Data Management Challenges at ConEd

Module-2: Big Data Modeling, Introduction to Data Models, Data Model Structures, Data Model Operations, Data Model Constraints, Introduction to CSV Data, What is a Relational Data Model?, What is a Semistructured Data Model?, Exploring the Relational Data Model of CSV Files, Exploring the Semistructured Data Model of JSON data, Exploring the Array Data Model of an Image, Exploring Sensor Data.

Module-3: Vector Space Model, Graph Data Model, Other Data Models, Exploring the Lucene Search Engine's Vector Data Model,Exploring Graph Data Models with Gephi.

Module-4: Data Model vs. Data Format, What is a Data Stream?, Why is Streaming Data different?, Understanding Data Lakes, Exploring Streaming Sensor Data.

Module-5: DBMS-based and non-DBMS-based Approaches to Big Data,From DBMS to BDMS, Redis: An Enhanced Key-Value Store, Aerospike: a New Generation KV Store, Semi structured Data – AsterixDB, Solr: Managing Text, Relational Data – Vertica.

Textbooks:

1. Hands-On Big Data Modeling, By James Lee , Tao Wei & Suresh Kumar Mukhiya
2. Data Management: Databases And Organizations 6th Editionby Richard T. Watson

Reference Books:

1. Big Data Principles and best practices of scalable realtime data systems, Nathan Marz and James Warren

TITLE OF COURSE: STATISTICS FOR DATA ANALYTICS

COURSE CODE: BDA603
L-T-P: 3-0-0
CREDITS: 3

Pre-requisite: Basic concepts Data base management system and SQL Query Language.

Introduction:

This statistics and data analysis course will pave the statistical foundation for our discussion on data science. You will learn how data scientists exercise statistical thinking in designing data collection, derive insights from visualizing data, obtain supporting evidence for data-based decisions and construct models for predicting future trends from data.

Course Outcomes (CO):

The students will be able to:

CO1: Data collection, analysis, and inference.
CO2: Data classification to identify key traits and customers.
CO3: Conditional Probability-How to judge the probability of an event, based on certain conditions.
CO4: How to use Bayesian modeling and inference for forecasting and studying public opinion.
CO5: Basics of Linear Regression.
CO6: Data Visualization: How to create use data to create compelling graphics.

Mapping of Course Outcomes (CO) and Program Outcomes (PO):

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓	✓		✓			✓				✓
CO2	✓			✓								✓
CO3	✓		✓									✓
CO4	✓	✓							✓			✓
CO5	✓	✓	✓									✓
CO6	✓				✓						✓	✓

Course Contents:

Module-1: Statistics and Big Data: What are Statistics and what is Big Data. How are Big Data problems being tackled? Why is it important for statistics to be one of the key disciplines for Big Data? What does statistics bring to Big Data and where are the opportunities?

Module-2: Introduction Statistical Thinking, Examples of Statistical Thinking, Numerical Data, Summary Statistics, From Population to Sampled Data, Different Types of Biases, Introduction to Probability, Introduction to Statistical Inference.

Module-3: Association and Dependence, Association and Causation, Conditional Probability and Bayes Rule, Simpsons Paradox, Confounding, Introduction to Linear Regression, Special Regression Models.

Module-4: Exploratory Data Analysis and Visualization, Goals of statistical graphics and data visualization, Graphs of Data, Graphs of Fitted Models, Graphs to Check Fitted Models, what makes a good graph, Principles of graphics.

Module-5: Introduction to Bayesian Modeling, Bayesian inference: combining models and data in a forecasting problem, Bayesian hierarchical modeling for studying public opinion, Bayesian modeling for Big Data.

Textbooks:

1. Statistics for Data Science, James D. Miller
2. Statistical Techniques for Data Analysis, By John K. Taylor, Cheryl Cihon

Reference Books:

1. An Introduction to Statistical Learning: with Applications in R By Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani

2. Statistical Models for Data Analysis, edited by Paolo Giudici, Salvatore Ingrassia, Maurizio Vichi

TITLE OF COURSE: BIG DATA INTEGRATION & MODELING
COURSE CODE: BDA704
L-T-P: 3-0-2
CREDITS: 4

Pre-requisite: Completion of Intro to Big Data is recommended

Introduction:

Nowadays, huge volume of data is collected from many heterogeneous data sources which are generating data in real-time with different qualities — which is called Big Data. The big data integration is very challenging especially after the traditional data integration techniques failed to handle it.

Course Outcomes (CO):

The students will be able to:
CO1: Retrieve data from example database and big data management systems
CO2: Describe the connections between data management operations and the big data processing patterns needed to utilize them in large-scale analytical applications
CO3: Identify when a big data problem needs data integration
CO4: Execute simple big data integration and processing on Hadoop and Spark platforms

Mapping of Course Outcomes (CO) and Program Outcomes (PO):

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓	✓		✓							✓
CO2	✓			✓								✓
CO3	✓	✓								✓		✓
CO4	✓		✓		✓				✓			✓

Course Contents:

Module-1: introduction to big data integration and processing, big data modeling and management, why is big data processing different? What is data retrieval?, querying two relations, subqueries, querying relational data with postgres.
Module-2: Retrieving Big Data: Querying JSON Data with MongoDB, Aggregation Functions, Querying Aerospike, Querying Documents is MongoDB, Exploring Pandas Data Frames.

Module-3: Big Data Integration: Overview of information integration, A Data integration Scenario, Integration for Multichannel Customer Analytics, Big Data Management and Processing Using Splunk and Datameer, why splunk?, Connected Cars with ford's OpenXC and Splunk, Big Data Management and Processing using Datameer, Installing splunk Enterprise on Windows, Installing splunk enterprise on Linux, Exploring Splunk Queries.

Module-4: Processing Big Data: Big Data Processing Pipelines, Some High-Level Processing Operations in Big Data Pipelines, Aggregation Operations in Big Data Pipelines, Typical Analytical Operations in Big Data Pipelines, Overview of Big Data Processing Systems, The Integration and Processing Layer, Introduction to Apache Spark, Getting Started with Spark, WordCount in Spark.

Module-5: Big Data Analytics using Spark: Spark Core: Programming In Spark using RDDs in Pipelines, Spark Core: Transformations, Spark Core: Actions, Spark SQL, Spark Streaming, Spark MLLib, Spark GraphX, Exploring SparkSQL and Spark DataFrames, Analyzing Sensor Data with Spark Streaming.

Textbooks:

1. Data Integration Blueprint And Modeling: Techniques For A Scalable And Sustainable Architecture (Paperback) (Ibm Press) 1st Edition By Anthony David Giordano
2. Managing Data In Motion: Data Integration Best Practice Techniques And Technologies (The Morgan Kaufmann Series On Business Intelligence) 1st Edition By April Reeve

Reference Books:

1. Principles of Data Integration 1st Edition by AnHai Doan, Alon Halevy , Zachary Ives

TITLE OF COURSE: MACHINE LEARNING WITH BIG DATA
COURSE CODE: BDA705
L-T-P: 3-0-2
CREDITS: 4

Pre-requisite: Completion of Intro to Big Data is recommended

Introduction:

This course provides an overview of machine learning techniques to explore, analyze, and leverage data. You will be introduced to tools and algorithms you can use to create machine learning models that learn from data, and to scale those models up to big data problems.

Course Outcomes (CO):

The students will be able to:
CO1: Design an approach to leverage data using the steps in the machine learning process.
CO2: Apply machine learning techniques to explore and prepare data for modeling.
CO3: Identify the type of machine learning problem to apply the appropriate set of techniques.
CO4: Construct models that learn from data using widely available open source tools.
CO5: Analyze big data problems using scalable machine learning algorithms on Spark.

Mapping of Course Outcomes (CO) and Program Outcomes (PO):

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓		✓		✓							✓
CO2	✓			✓	✓						✓	✓

CO3	✓	✓	✓								✓	✓
CO4	✓	✓	✓		✓							✓
CO5	✓			✓								✓

Course Contents:

Module-1: Introduction to Machine Learning With Big Data, Summary of Big Data Integration and Processing, Machine Learning Overview, Categories Of Machine Learning Techniques, Machine Learning Process, Goals and Activities in the Machine Learning Process, CRISP-DM, Scaling Up Machine Learning Algorithms

Module-2: Data Exploration: Data Terminology, Data Exploration, Data Exploration through Summary Statistics, Data Exploration through Plots, Exploring Data with KNIME Plots, Data Exploration in Spark.

Module-3: Data Preparation: Data Preparation, Data Quality, Addressing Data Quality Issues, Feature Selection, Feature Transformation, Dimensionality Reduction, Handling Missing Values in KNIME, Handling Missing Values in Spark.

Module-4: Classification: introduction to Classification, Building and Applying a Classification Model, Classification Algorithms, k-Nearest Neighbors, Decision Trees, Naïve Bayes, Classification using Decision Tree in KNIME, Classification in Spark.

Module-5: Evaluation of Machine Learning Models: Generalization and Overfitting, Overfitting in Decision Trees, Using a Validation Set, Metrics to Evaluate Model Performance, Confusion Matrix, Evaluation of Decision Tree in KNIME, Evaluation of Decision Tree in Spark.

Module 6: Regression, Cluster Analysis, and Association Analysis: Regression Overview, Linear Regression, Cluster Analysis, k-Means Clustering, Association Analysis, Association Analysis in Detail, Machine Learning With Big Data - Final Remarks, Cluster Analysis in Spark.

Textbooks:

1. Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners Book by Jared Dean
2. Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning Book by Shan Suthaharan.

Reference Books:

1. Big Data and Machine Learning in Quantitative Investment Book by Tony Guida
2. Machine Learning For Big Data Analysis, Edited By Siddhartha Bhattacharyya, Hrishikesh Bhaumik, Anirban Mukherjee, Sourav De

TITLE OF COURSE: GRAPH ANALYTICS FOR BIG DATA
COURSE CODE: BDA806
L-T-P: 3-0-0
CREDITS: 3

Pre-requisite: Basic concepts of Data structure and Data base management system.

Introduction:

Graph analytics is an emerging form of data analysis, one that works particularly well with complex relationships. It involves moving data points and relationships between data points into a graph format

(also known as nodes and links, or vertices and edges). When querying complex relationships or distant connections between data, graph analytics offers a solution that codes queries more efficiently and can output results in an easy-to-digest visual format.

Course Outcomes (CO):

The students will be able to:
CO1: Model A Problem into A Graph Database
CO2: Perform Analytical Tasks Over the Graph in A Scalable Manner.
CO3: Apply These Techniques to Understand the Significance Of Your Data Sets For Your Own Projects.

Mapping of Course Outcomes (CO) and Program Outcomes (PO):

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓	✓		✓				✓		✓	✓
CO2	✓	✓	✓		✓						✓	✓
CO3	✓	✓	✓		✓				✓		✓	✓

Course Contents:
Module-1: Graphs: What is a Graph?, Why Graphs?, Why Graphs? Example 1: Social Networking, Why Graphs? Example 2: Biological Networks, Why Graphs? Example 3: Human Information Network Analytics, Why Graphs? Example 4: Smart Cities, The Purpose of Analytics, What are the impact of Big Data's V's on Graphs?
Module-2: Graph Analytics: Focusing On Graph Analytics Techniques, Path Analytics, The Basic Path Analytics Question: What is the Best Path?, Applying Dijkstra's Algorithm, Inclusion and Exclusion Constraints, Connectivity Analytics, Disconnecting a Graph, Connectedness: Indegree and Outdegree, Community Analytics and Local Properties, Global Property: Modularity, Centrality Analytics.

Module-3: Bi-directional Dijkstra Algorithm, Goal-directed Dijkstra Algorithm, Power Law Graphs, Measuring Graph Evolution, Eigenvector Centrality, Key Player Problems.
Module-4: Graph Analytics Techniques: Hands-On: Downloading, Installing, and Running Neo4j, Hands-On: Getting Started With Neo4j, Hands-On: Modifying a Graph With Neo4j, Hands-On: Importing Data Into Neo4j, Hands-On: Basic Queries in Neo4j With Cypher, Hands-On: Path Analytics in Neo4j Using Cypher ,Hands-On: Connectivity Analytics in Neo4j With Cypher.
Module-5: Introduction: Large Scale Graph Processing, A Parallel Programming Model for Graphs, Pregel: The System That Changed Graph Processing, Giraph and GraphX, Beyond Single Vertex Computation, Introduction to GraphX: Hands-On Demonstrations, Hands On: Building a Graph, Hands On: Building a Degree Histogram, Hands On: Plot the Degree Histogram, Hands On: Network Connectedness and Clustering Components, Hands On: Joining Graph Datasets.

Textbooks:

1. Graph Analytics Using Big Data,By Rajat Mehta

Reference Books:

1. Graph Algorithms: Practical Examples in Apache Spark and Neo4j, By Mark Needham and Amy E. Hodler

TITLE OF COURSE: MANAGING BIG DATA WITH SQL
COURSE CODE: BDA807
L-T-P: 3-0-0
CREDITS: 3

Pre-requisite: Basic concepts Data base management system and SQL Query Language.

Introduction:

This course is an introduction to how to use relational databases in business analysis. You will learn how relational databases work, and how to use entity-relationship diagrams to display the structure of the data held within them. This knowledge will help you understand how data needs to be collected in business contexts, and help you identify features you want to consider if you are involved in implementing new data collection efforts.

Course Outcomes (CO):

The students will be able to:
CO1: Understand How Data Needs to Be Collected in Business Contexts.
CO2: Identify Features You Want to Consider If You Are Involved in Implementing New Data Collection Efforts
CO3: Understand How to Execute the Most Useful Query and Table Aggregation Statements For Business Analysts
CO4: Understand Query Practice Using Them with Real Databases.

Mapping of Course Outcomes (CO) and Program Outcomes (PO):

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓			✓		✓						✓
CO2	✓			✓		✓						✓
CO3	✓			✓		✓						✓
CO4	✓	✓	✓			✓					✓	✓

Course Contents:

Module-1: Problems with Having a Lot of Data Used by a Lot of People, How Relational Databases Help Solve Those Problems, Database Design Tools That Will Help You Learn SQL Faster.

Module-2: How Entity-Relationship Diagrams Work, Database Structures Illustrated by Entity-Relationship Diagrams, Relational Schemas, How to Make Entity-Relationship Diagrams using ERDPlus, How to Make Relational Schemas using ERDPlus.

Module-3: Queries to Extract Data from Single Tables: Introduction to Query Syntax, How to Use Jupyter Notebooks, How to Use Your Jupyter Account, How to Use Teradata Viewpoint and SQL Scratchpad.

Module-4: Queries to Summarize Groups of Data from Multiple Tables: What are Joins? Joins with Many to Many Relationships and Duplicates, A Note about Our Join Examples, Retrieve Your Data.

Module-5: Queries to Address More Detailed Business Questions: Design and execute subqueries, Introduce logical conditions into your queries using IF and CASE statements, Implement analyses that accommodate missing data or data mistakes, and Write complex queries that incorporate many tables and clauses.

Textbooks:

1. Sql on big data, technology, architecture, and innovation, authors: pal, sumit.

Reference Books:

1. Oracle database 11g pl/sql programming by mclaughlin, mcgraw hill**,** by mclaughlin.