Name of Institute: Indus Institute of Technology and Engineering
Name of Faculty: Prof. Naiswita Parmar

Course code: CE0703
Course name: Big Data Analytics

Pre-requisites:
1. Knowledge of one Programming Language (Java preferably)
2. Practice of SQL (queries and sub queries)
3. Exposure to Linux Environment

Credit points: 04
Offered Semester: VII

Course coordinator (weeks 12)
Full name: Prof. Naiswita Parmar
Department with seating location: Computer Engineering Department, 4th Floor, Bhanwar Building
Telephone: 3427
Email: naiswitaparmar.ce@indusuni.ac.in
Consultation times: 04:15 PM to 05:00 PM

Course lecturer (weeks 12)
Full name: Prof. Naiswita Parmar
Department with seating location: Computer Engineering Department, 4th Floor, Bhanwar Building
Telephone: 3427
Email: naiswitaparmar.ce@indusuni.ac.in
Consultation times: 04:15 PM to 05:00 PM

Students will be contacted throughout the session via mail with important information relating to this course.

Course Objectives

By participating in and understanding all facets of this course a student will:

1) Be Understand the Big Data Platform and its Use cases
2) Provide an overview of Apache Hadoop
3) Provide HDFS Concepts and Interfacing with HDFS
4) Understand Map Reduce Jobs
5) Provide hands on Hadoop Eco System
6) Apply analytics on Structured, Unstructured Data.

Course Outcomes (CO)

1. ***Understand*** the concepts of Big data and challenges in processing Big Data
2. ***Analyze*** Hadoop architecture and eco-system.
3. ***Learn*** conceptual understanding of Hadoop Distributed File System.
4. ***Apply*** the concepts of map and reduce and functional programming
5. ***Apply*** appropriate techniques and tools to solve actual Big Data problems.
6. ***Analyze*** on Big Data Applications Using Pig and Hive.
7. ***Create*** a complete business data analytics solution

Course Outline

This course aims have been offered with the aim of providing the students with in depth knowledge about big data platform and explore the big data analytics techniques. The curriculum includes any real time recommendation system case study.

Method of delivery

Face to face lectures, Online Lecture, self-study material, Active Learning Techniques, PowerPoint Presentation and Assignments

Study time

Lecture hours: 3 hours
Lab hours: 2 hours

CO-PO Mapping (PO: Program Outcomes)

Program Outcomes (POs)

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, review research literature, and analyze 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 modeling 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.

Program Specific Outcomes (PSOs)

At the end of the program, the Computer Engineering student:

PSO1. Basics of Computer System: Should able to understand the principles and working of computer systems. Students can assess the hardware and software aspects of computer systems.

PSO2. Program Design: Design and develop computer programs in the areas related to algorithms, networking, web design, cloud computing, IoT and data analytics.

PSO3.Software Development: Should able to understand the structure and development methodologies of software systems with the use of a various programming languages and open source platforms.

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO 1	3	3	1	1	-	-	-	-	-	-	-	-
CO 2	1	3	1	3	2	-	-	-	-	-	-	-
CO 3	2	1	-	-	-	-	-	-	-	-	-	-
CO 4	1	2	2	2	3	1	-	-	-	-	-	-
CO 5	-	-	1	2	3	2	-	-	-	-	-	-
CO 6	1	2	3	-	2	-	-	-	-	2	-	-
CO 7	-	1	2	2	3	-	-	-	-	-	1	-

Blooms Taxonomy and Knowledge retention (For reference)

(Blooms taxonomy has been given for reference)

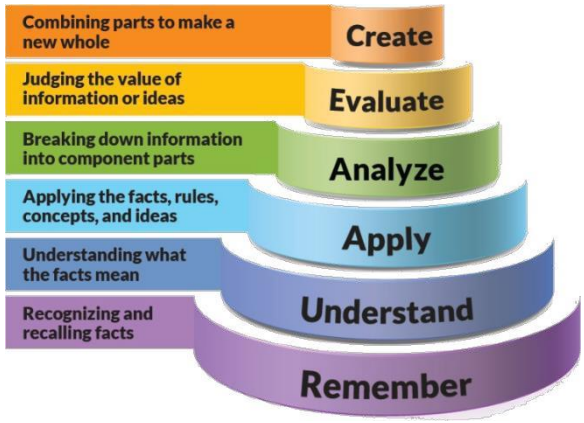


Figure 1: Blooms Taxonomy

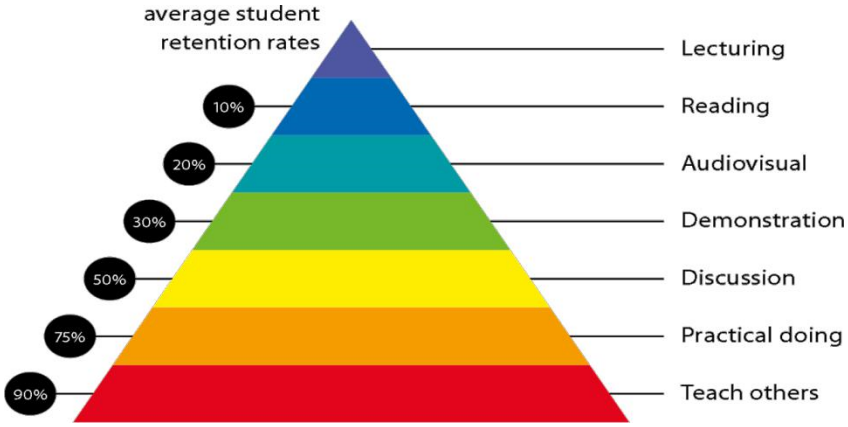


Figure 2: Knowledge retention

Graduate Qualities and Capabilities covered

(Qualities graduates harness crediting this Course)

General Graduate Qualities	Specific Department of _____Graduate Capabilities
Informed Have a sound knowledge of an area of study or profession and understand its current issues, locally and internationally. Know how to apply this knowledge. Understand how an area of study has developed and how it relates to other areas.	**1 Professional knowledge, grounding & awareness**
Independent learners Engage with new ideas and ways of thinking and critically analyze issues. Seek to extend knowledge through ongoing research, enquiry and reflection. Find and evaluate information, using a variety of sources and technologies. Acknowledge the work and ideas of others.	**2 Information literacy, gathering & processing**
Problem solvers Take on challenges and opportunities. Apply creative, logical and critical thinking skills to respond effectively. Make and implement decisions. Be flexible, thorough, innovative and aim for high standards.	**4 Problem solving skills**
Effective communicators Articulate ideas and convey them effectively using a range of media. Work collaboratively and engage with people in different settings. Recognize how culture can shape communication.	**5 Written communication** **6 Oral communication** **7 Teamwork**
Responsible Understand how decisions can affect others and make ethically informed choices. Appreciate and respect diversity. Act with integrity as part of local, national, global and professional communities.	**10 Sustainability, societal & environmental impact**

Practical work:

	Lab Number	Practical	CO Addressed
	1	**Study of Hadoop ecosystem**	2,3
	2	**Programming exercises on Hadoop**	2,3
	3	**Programming exercises in No SQL**	1
	4	**Implementing simple algorithms in Map- Reduce (3) -Matrix multiplication, Aggregates, joins, sorting, searching etc.**	4
	5	**Implementing any one Frequent Itemset algorithm using Map-Reduce**	4
	6	**Implementing any one Clustering algorithm using Map-Reduce.**	4
	7	**Implementing any one data streaming algorithm using Map-Reduce.**	4
	8	**Mini Project: One real life large data application to be implemented**	6,7
	9	**Twitter data analysis**	1,2,5
	10	**Fraud Detection and Text Mining**	1,2
	11	**Survey Paper**	5,6
	12	**Survey Paper**	5,6

Lecture/tutorial times

(Give lecture times in the format below)

Lecture /Tutorial/Practicals	Day	Time	
Lecture	Monday	11:10 am – 12:10 pm	7-CSE
Lecture	Monday	2:00 pm – 3:00 pm	7-CE
Lecture	Tuesday	11:10 am – 12:10 pm	7- CE
Lecture	Tuesday	12:10 pm- 1:20 pm	7-CS
Lecture	Wednesday	10:00 am – 11:00 am	7-IT
Lecture	Wednesday	12:10 pm – 1:20 pm	7-CE
Lecture	Thursday	10:00 am – 11:00 am	7-CS
Lecture	Thursday	12:10 pm – 1:10 pm	7-IT
Lecture	Friday	10:00 am – 11:00 am	7-IT

Attendance Requirements

The University norms states that it is the responsibility of students to attend all lectures, tutorials, seminars and practical work as stipulated in the course outline. Minimum attendance requirement as per university norms is compulsory for being eligible for semester examinations.

Details of referencing system to be used in written work

Text books

1. Boris lublinsky, Kevin t. Smith, AlexeyYakubovich, “Professional Hadoop Solutions”,
2. Wiley, Second Edition,2015, ISBN: 9788126551071

Reference books

1. Chris Eaton,Dirkderooset al. , “Understanding Big data ”, McGraw Hill, 2012. ISBN: ISBN 978-0-07-179053-6
2. Tom White, “HADOOP: The definitive Guide”, O Reilly 2012, ISBN-10: 1449311520
3. VigneshPrajapati, “Big Data Analyticswith R and Haoop”, Packet Publishing 2013, ISBN-10: 178216328X
4. Learning Spark: Lightning-Fast Big Data Analysis Paperback by Holden Karau, ISBN-10:9351109941 Additional Materials

Additional Material

1. NPTEL- Lecture https://onlinecourses.nptel.ac.in/noc16_cs12/preview
2. Analytics http://nptel.ac.in/courses/106106142/https://nptel.ac.in/courses/106105166/
3. Hadoop https://www.edureka.co/big-data-and-hadoop
4. Spark https://www.coursera.org/courses?query=apache spark

ASSESSMENT GUIDELINES

Your final course mark will be calculated from the following:

CIE – Theory – 60 Marks	
Mid Semester Exam	40 Marks
Case Study/Assignment	10 Marks
Presentation	05 Marks
Attendance	05 Marks
CIE – Practical – 60 Marks	
Lab Experiment Conduction and Regularity	20 Marks
Case Study Implementation & Presentation	40 Marks

SUPPLEMENTARY ASSESSMENT

Students who receive an overall mark less than 40% in internal component or less than 40% in the end semester will be considered for supplementary assessment in the respective components (i.e internal component or end semester) of semester concerned. Students must make themselves available during the supplementary examination period to take up the respective components (internal component or end semester) and need to obtain the required minimum 40% marks to clear the concerned components.

Practical Work Report/Laboratory Report:

A report on the practical work is due the subsequent week after completion of the class by each group.

Late Work

Late assignments will not be accepted without supporting documentation. Late submission of the reports will result in a deduction of -% of the maximum mark per calendar day

Format

All assignments must be presented in a neat, legible format with all information sources correctly referenced. **Assignment material handed in throughout the session that is not neat and legible will not be marked and will be returned to the student.**

Retention of Written Work

Written assessment work will be retained by the Course coordinator/lecturer for two weeks after marking to be collected by the students.

University and Faculty Policies

Students should make themselves aware of the University and/or Faculty Policies regarding plagiarism, special consideration, supplementary examinations and other educational issues and student matters.

Plagiarism - Plagiarism is not acceptable and may result in the imposition of severe penalties. Plagiarism is the use of another person's work, or idea, as if it is his or her own - if you have any doubts at all on what constitutes plagiarism, please consult your Course coordinator or lecturer. Plagiarism will be penalized severely.

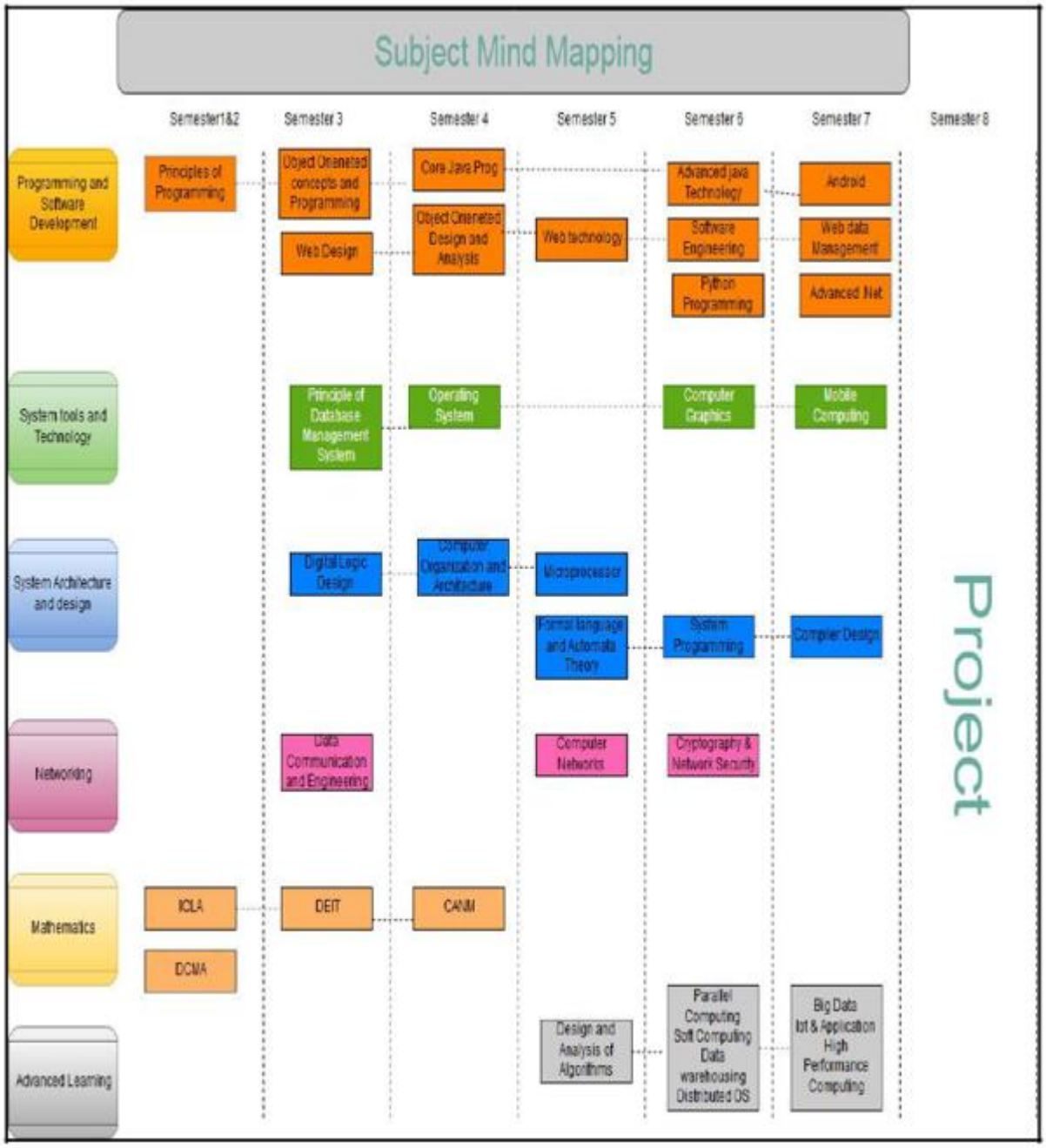
Do not copy the work of other students.
Do not share your work with other students (except where required for a group activity or assessment)

Course schedule (subject to change)

(Mention quiz, assignment submission, breaks etc as well in the table under the Teaching Learning Activity Column)

	Week #	Topic & contents	CO Addressed	Teaching Learning Activity (TLA)
	Weeks 1	Introduction– distributed file system– Big Data and its importance, Four Vs, Drivers for Big data	1	Quiz, Assignment
	Weeks 2	Big data analytics, and Big data applications. Algorithms using map-reduce.	1, 3	Quiz, Assignment
	Week 3	Apache Hadoop & HadoopEcoSystem	2,3	Quiz, Assignment
	Week 4	Moving Data in and out of Hadoop	2,3	Quiz, Assignment
	Week 5	Understanding inputs and outputs of MapReduce , Data Serialization	2,3,4	Quiz, Assignment
	Week 6	Data Serialization	4	Quiz, Assignment
	Week 7	HDFS-Overview, Installation and Shell, Java API; Hive Architecture and Installation, Comparison with Traditional Database, HiveQL Querying Data- Sorting And Aggregating	3	Quiz, Assignment
	Week 8	Map Reduce Scripts, Joins &Subqueries, HBase concepts- Advanced Usage, Schema Design, Advance Indexing	3,4	Project
	Week 9	PIG, Zookeeper- how it helps in monitoring a cluster, HBase uses ZookeeperandhowtoBuild Applications with Zookeeper.	5	Quiz, Assignment
	Week 10	Introduction to Data Analysis with Spark, Downloading Spark and Getting Started, Programming with RDDs	5,6	Quiz, Assignment
	Week 11	Machine Learning with MLlib. Oozie	6	Quiz, Assignment
	Week 12	Pig-Introduction, Joining datasets and	5,6	Case Study

		other advanced topics; Hive		


Subject Mind Mapping
Semester1&2
Semester 3
Semester 4
Semester 5
Semester 6
Semester 7
Semester 8
Programming and Software Development
Principles of Programming
Object Oriented concepts and Programming
Web Design
Core Java Prog
Object Oriented Design and Analysis
Web technology
Advanced Java Technology
Software Engineering
Python Programming
Android
Web data Management
Advanced Net
System tools and Technology
Principle of Database Management System
Operating System
Computer Graphics
Mobile Computing
System Architecture and design
Digital Logic Design
Computer Organization and Architecture
Microprocessor
Formal language and Automata Theory
System Programming
Compiler Design
Networking
Data Communication and Engineering
Computer Networks
Cryptography & Network Security
Mathematics
ICLA
DCMA
DEIT
CANM
Advanced Learning
Design and Analysis of Algorithms
Parallel Computing Soft Computing Data warehousing Distributed OS
Big Data Iot & Application High Performance Computing
Project