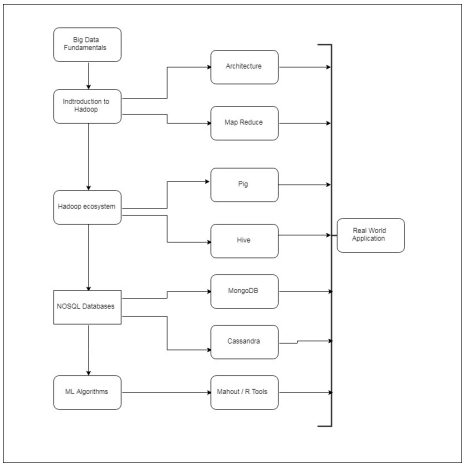
Course Outcomes

CO No	Course Outcome Description	Blooms Taxonomy Level (BTL)
15CSE334-CO1	Understand fundamental concepts of Big Data and its technologies	L2
15CSE334-CO2	Apply concepts of MapReduce framework for optimization.	L3
15CSE334-CO3	Analyze appropriate NoSQL database techniques for storing and processing large volumes of structured and unstructured data.	L4
15CSE334-CO4	Apply data analytics solutions using Hadoop ecosystems	L3
15CSE334-CO5	Explore modern reporting tools for Machine learning.	L4

CO-PO Mapping

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
CO1	3	2	1										2	1
CO2	1	3	2										2	1
CO3		3	2	1									2	1
CO4		3	2	1									2	1
CO5				2	3								2	1

Concept Map



Course Plan

Lec Hrs	TOPICS	DESCRIPTION	CO	BTL
1	Types of Digital Data	Data types and its Characteristics	CO1	L2
2	Evolution of Big Data	History of Big data	CO1	L2
3	Introduction to Big Data	Definition of Big Data-Challenges with Big Data-3Vs of Big Data-Non Definitional traits of Big Data	CO1	L2
4	Big data vs BI and Datawarehouse Vs Hadoop environment	BI Vs Big data, Data warehouse Vs Hadoop environment	CO1	L2
5 to 6	Big Data Analytics	Classification of analytics- Data Science-Terminologies in Big Data	CO1	L2
7	Big Data Theorems	CAP Theorem- BASE Concept.	CO1	L2
8	NoSQL datatypes	Types of Databases- Advantages- NewSQL- SQL Vs. NOSQLVs NewSQL.	CO1	L2
9 to 10	Introduction to Hadoop	Features- Advantages- Versions- Overview of Hadoop Eco systems- Hadoop distributions- Hadoop vs. SQL – RDBMS vs. Hadoop.	CO2	L3
11 to 13	Hadoop Architecture	Hadoop Components- Architecture	CO2	L3
14 to 15	Hadoop Distributed Systems	HDFS Daemons, Anatomy File read and Write , Replica replacement, HDFS commands, Special features.	CO2	L3
16 to 21	MapReduce Concepts	Mapper- Reducer- Combiner- Partitioner- Searching- Sorting- Compression	CO2	L3
22	Hadoop 2 (YARN)	Architecture-Interacting with Hadoop Eco systems	CO2	L2
		Periodical I		
23 to 30	Hive	Architecture- data type- File format- HQL- SerDe- User defined functions	CO4	L3
31 to 38	Pig	Features- Anatomy-Pig on Hadoop- Pig Philosophy- Pig Latin overview- Data types- Running pig- Execution modes of Pig-HDFS commands- Relational operators- Eval Functions- Complex data type- Piggy Bank- User defined Functions- Parameter substitution- Diagnostic operator.	CO4	L4
39 to 46	MongoDB	Introduction – Features-Data types- Mongo DB Query language- CRUD operations –Arrays-Functions: Count- Sort- Limit-Skip- Aggregate- Map Reduce. Cursors-Indexes- Mongo Import – Mongo Export.	CO3	L4
		Periodical II		
47 to 54	Cassandra	Introduction – Features- Data types -CQLSH- Key spaces- CRUD operations- Collections-Counter- TTL-Alter commands- Import and Export- Querying System tables	CO3	L4

Lec Hrs	TOPICS	DESCRIPTION	CO	BTL
55 to 56	R tool	Introduction- Connecting to Mongo DB- Connecting to Cassandra	CO5	L3
57 to 60	Introduction to Machine Learning	Linear Regression- Clustering- Collaborative filtering- Association rule mining-Decision tree.	CO5	L4

CO – PO Mapping - Justification

Engineering Graduates will be able to:

1. **Engineering knowledge**: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

CO1	3	Understanding basic concepts of relational databases, networking and Java
CO2	1	Basic architecture of hadoop

2. **Problem analysis**: Identify, formulate, review research literature, and analyse complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

CO1	2	Analyse differences between SQL and NOSQL databases
CO2	3	Designing solutions for Big data using Map Reduce concept
CO3	3	Designing NOSQL databases
CO4	3	Developing retrieval strategies for hadoop ecosystem

3. **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

CO1	1	Choice of tools for Big Data solutions
CO2	2	Analysing the efficiency of map reduce programming
CO3	2	Designing map reduce techniques in Mongo dB and Cassandra
CO4	2	Developing map reduce and ETL solutions for Big Data

4. **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.

CO3	1	Learning new strategies and complex data types in NOSQL databases
CO4	1	Handling new features in hadoop ecosystem
CO5	2	Exploring R / Mahout

5. **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.

CO5	3	Adapting machine learning algorithms to Big Data Environment.

6. **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.
   NA

7. **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.
   NA

8. **Ethics**: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
   NA

9. **Individual and team work**: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
   NA

10. **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.
    NA

11. **Project management and finance**: Demonstrate knowledge and understanding of 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.
NA

12. **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.
NA

15CSE335 BIOINFORMATICS 3-0-0 3

Syllabus

Unit 1
Introduction: The Central Dogma – Killer Application – Parallel Universes – Watson's Definition – Top-Down vs Bottom-Up Approach – Information Flow – Conversance – Communications. Database and Networks: Definition – Data Management – Data Life Cycle – Database Technology – Interfaces – Implementation – Networks: Communication Models – Transmission Technology – Protocols – Bandwidth –Topology – Contents – Security – Ownership – Implementation.

Unit 2
Search Engines and Data Visualization: Search Process – Technologies – Searching and Information Theory – Computational Methods – Knowledge Management – Sequence Visualizations – Structure Visualizations – User Interfaces – Animation vs Simulation. Statistics, Data Mining and Pattern Matching: Statistical Concepts – Micro Arrays – Imperfect Data – Basics – Quantifying – Randomness – Data Analysis – Tools Selection – Alignment – Clustering – Classification – Data Mining Methods – Technology – Infrastructure Pattern Recognition – Discovery.

Unit 3
Machine Learning – Text Mining – Pattern Matching Fundamentals – Dot Matrix Analysis – Substitution Matrix – Dynamic Programming – Word Method – Bayesian Method – Multiple Sequence Alignment Tools. Modelling Simulation and Collaboration: Drug Discovery Fundamentals – Protein Structure – System Biology Tools – Collaboration and Communication – Standards – Issues – Case Study.