SANKALCHAND PATEL UNIVERSITY
।। अथातो ज्ञानजिज्ञासा ।।

Ph. D. Program
(July-2020 batch)

Subject Code: 1RI3000102	Subject Title: Essentials of Disciplinary Research Subject-1 (BIG DATA TOOLS)
Pre-requisite Subject	Working knowledge of programming language and database concepts

Objectives of course:
The objective of the course is to:
- Understand basics of Big Data and Big Data Tools (Hadoop, MapReduce)

Learning outcomes:
At the end of this course, the student would be able
- To understand the importance of Big Data and Big Data Tools for solving real world problems.

Teaching Scheme (Credit Distribution)				Evaluation Scheme (% Marks)	Total Marks
Lecture	Tutorial	Practical	Credit	Continuous (Internal) Assessment	
3	1	--	4	100%	100

Unit	Content	Wt. (%)
1	**Overview – Big Data Tools-I:** **Big Data:** Introduction to Big Data, Big Data characteristics, Types of Big Data, Traditional vs. Big Data approach **Hadoop:**What is Hadoop? Core Hadoop Components, Hadoop Ecosystem (Hbase, Hive, Hcatalog, Pig, Sqoop, Oozie, Mahout, ZooKeeper), Physical Architecture, Hadoop limitations.	25
2	**Overview – Big Data Tools-II:** **NoSQL:** What is NoSQL? NoSQL business drivers, NoSQL case studies (Amazon Dynamo DB, Google's Big Table, MongoDb, Neo4j), NoSQL data architecture patterns (Key-value stores, Graph stores, Column family (Bigtable) stores, Document stores), Variations of NoSQL architectural patterns; Using NoSQL to manage big data **Map Reduce:** MapReduce and New Software stack (Distributed File Systems, Physical Organization of Compute Nodes), The Map Tasks, Grouping by Key, The Reduce Tasks, Combiners, Details of MapReduce Execution, Coping With Node Failures	25
3	**Recommendation Systems:** Recommendation Systems: A Model for Recommendation Systems, Content-Based Recommendations, Collaborative Filtering.	25
4	**Social Network Graphs:** Social Network Graphs: Application of social network mining, Social network as a graph, Types of social networks,	25

	Clustering of social Graphs, Direct discovery of communities in a social Graph, SimRank	

List of References:

1. Radha Shankarmani, M Vijayalakshmi, Big Data Analytics, 2nd Edition, Wiley
2. Jure Leskovec, AnandRajaraman, Jeffrey David Ullman, "Mining of Massive Datasets", Cambridge University Press, Second Edition, 2014.
3. Seema Acharya, Subhashini Chhellappan, BIG Data and Analytics, Willey

SANKALCHAND PATEL UNIVERSITY
।। अथातो ज्ञानजिज्ञासा ।।

Ph. D. Program
(July-2020 batch)

Subject Code: 1RI3000103	**Subject Title: Essentials of Disciplinary Research Subject-2 (BIG DATA ANALYTICS)**
Pre-requisite Subject	**Working knowledge of programming language and database concepts**

Objectives of course:
The objective of the course is to:
- Understand fundamental techniques used for Big data analytics

Learning outcomes:
At the end of this course, the student would be able
- To understand the mechanism of fundamental techniques used for the Big Data Analytics.

Teaching Scheme (Credit Distribution)				**Evaluation Scheme (% Marks)**	**Total Marks**
Lecture	**Tutorial**	**Practical**	**Credit**	**Continuous (Internal) Assessment**	
---	**2**	**---**	**2**	**100%**	**50**

Unit	Content	Wt. (%)
1	**Finding Similar items:** Nearest neighbor Search, Applications of Near-Neighbor Search, Similarity of documents (Plagiarism detection, Document clustering, News aggregation), Collaborative Filtering as a Similar-Sets Problem, Recommendations based on user ratings, Distance Measures (Definition of a Distance Measure, Euclidean Distances, Jaccard Distance, Cosine Distance, Edit Distance, Hamming Distance)	25
2	**Data Stream Mining:** The Stream Data Model: A Data-Stream-Management System, Examples of Stream Sources, Stream Queries, Issues in Stream Processing.	25
3	**Sampling Data in a Stream:** Sampling Data in a Stream, Filtering Streams (The Bloom Filter), Counting Distinct Elements in a Stream, Counting Ones in a Window, Decaying Windows.	25
4	**Link Analysis:** PageRank, Efficient Computation of Page Rank (PageRank implementation using MapReduce, Use of Combiners to Consolidate the Result Vector), Topic sensitive Page Rank, link Spam, Hubs and Authorities.	25

List of References:

1. Jiawei Han, Micheline Kamber, Jian Pei, "Data Mining Concepts and Techniques", Morgan Kaufman Publications, Third Edition, 2011.
2. VigneshPrajapati, "Big Data Analytics with R and Haoop", Packet Publishing 2013
3. Big Data Black Book, DreamTech