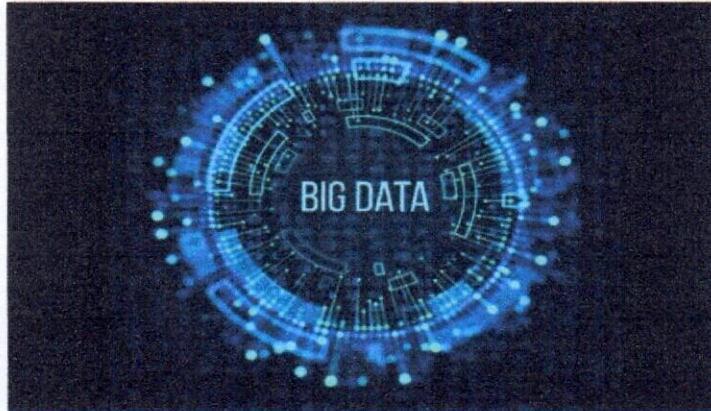


## ANNEXURE 'D'

### IFTM UNIVERSITY School of Computer Science & Applications

#### *Big Data*



This subject aims to cover the handling of large amount of data i.e Big Data. With this technology an organization or individual can obtain, store, transform and analyze large amounts of data to solve specific problems. Big data is a combination of structured, semi-structured and unstructured data collected by organizations that can be mined for information and used in machine learning projects, predictive modeling and other advanced analytics applications.

Systems that process and store big data have become a common component of data management architectures in organizations, combined with tools that support big data analytics uses.

Big Data allows organizations to detect trends, and spot patterns that can be used for future benefit. It can help to detect which customers are likely to buy products, or help to optimize marketing campaigns by identifying which advertisement strategies have the highest return on investment.

**Module Objectives:** This module intends to:

- ✓ Students able to know about the concept and challenge of big data (3 V's: volume, velocity, and variety).
- ✓ Students able to applying skills and tools to manage and analyze the big data.
- ✓ Provides information about new business opportunities in alternate fields of business.
- ✓ Provides information about best practices and improvements to enhance profitability and productivity.
- ✓ Improves the overall operation and leads to an increase in customer satisfaction.

Focusing on the above-mentioned objectives, a teacher's ability to improve technical skills can promote a positive development in students, by focusing on their competence, their tolerance to mistakes and their ability to set goals. So, the module has been designed by breaking down big tasks into easy steps and emphasizing the importance of mistakes in achieving success which can help students to avoid being overwhelmed by work or fearing failure.

*Dr. Arvind Kumar Shukla & Mr. Lalit Johri*

[Module Facilitator]

*Sanjeev Bhowmik*

Registrar  
IFTM University  
Moradabad

## Module Structure

### Program Schedule: Big Data

S.NO.	OBJECTIVES	COURSE CONTENT	HOURS
1	Introduction to Big Data	Big Data introduction – -Big data: definition and taxonomy - Big data value for the enterprise - Setting up the demo environment - First steps with the Hadoop “ecosystem” The 3 V's, their challenges and application domains. The Hadoop ecosystem: - Introduction to Hadoop - Hadoop components: MapReduce/Pig/Hive/HBase - Loading data into Hadoop - Handling files in Hadoop - Getting data from Hadoop	7
2	Collection of Big Data	Eventual Consistency and NoSQL systems MongoDB- Google BigTable Querying big data with Hive - Introduction to the SQL Language - From SQL to HiveQL	5
3	Large-Scale Data Analytics Systems	Auto-Parallel Data Programming; -Map Reduce -Hive and Parallel Databases Basic Statistical Analysis: - Fraud and Benford's Law, -Bayesian Introduction, - Heteroskedasticity	6
4	Machine Learning Systems for Big Data	Big data & Machine learning - Quick intro to Machine learning - Big Data & Machine Learning - Machine learning tools - Spark & SparkML- H2O-Azure ML Graph Analytics: Graph structures (diameter, connectivity, centrality), PageRank, Triangle counting.	6
5	Sentiment Analysis & Data Visualization	Sentiment Analysis, Data Visualization: Data types and dimensions; Visual encoding and perception	6
8	Review & Feedback		1

*Sanjew @ [Signature]*

Registrar  
IFTM University  
Moradabad