B. Tech with

MINOR

in

DATA SCIENCE

Academic Regulations, Course Structure and Syllabus

Effective from 2023-24 admitted batches

Offered by

Department of Computer Science and Engineering

KSRM College of Engineering (A) – Kadapa

(Approved by AICTE, Accredited by NAAC with A+ Grade and NBA and Affiliated to JNTUA, Anantapuramu)

ELIGIBILITY / REGISTRATION / AWARD OF MINOR

The primary objective of a minor degree is to provide students with a secondary area of study to broaden their knowledge, enhance their skill set, and potentially improve their career prospects. It allows students to explore interests beyond their major, potentially leading to a more well-rounded and competitive profile.

i) Minor degree is introduced by the respective departments offering B. Tech. programs and is applicable to all B.Tech (Regular and Lateral Entry) students admitted in Engineering & Technology.

ii) There is NO minimum CGPA requirement to register for Minor degree program. However, the student shouldn't have any course backlog **up to III Semester in the major degree.**

iii) A student shall earn **additional 18 credits for award of Minor degree** offered by the department other than parent department. This is in addition to 163 credits by a regular student and 123 Credits by a Lateral Entry student for the award of Major degree.

iv) A student is permitted to register for Minor in IV Semester after the results of III Semester are declared. Students shall register and pass in all the courses prescribed and being offered from V semester under the respective Minor degree.

v) Students have to attend classwork for courses under Minor degree beyond regular academic hours meant for major degree. Students can also undergo the courses under Minor through any proctored online platforms with the prior approval of the BoS Chairman and the HoD of the respective department offering Minor degree.

vi) The attendance for the registered courses under Minor and regular courses offered for Major degree in a Semester will be considered separately.

vii) A student shall have an aggregate of 75% attendance in all courses registered under Minor in that particular semester to become eligible for attending Semester-End examinations.

viii) The registration for the Minor will be cancelled, if the student is detained due to lack of attendance in Major.

ix) The registration for the Minor will be cancelled, if the student fails in any course of either Minor / Major in any semester from V to VIII Semester.

x) A student registered for Minor shall pass in all subjects that constitute the requirement for the Minor degree program. No class/division (i.e., second class, first class and distinction, etc.) will be awarded for Minor degree program.

xi) A separate grade sheet will be issued for the Minor degree courses semester-wise..

xii) If a student drops or is terminated from the Minor program, the additional credits so far earned cannot be converted into open or core electives; they will remain extra.

xiii) The Minor will be mentioned in the degree certificate as Bachelor of Technology in Civil Engineering with Minor in Mechanical Engineering.

xiv) There shall be a minimum enrolment of 20% OR 20 enrollments from the list of

eligible students to offer Minor program.

xv) There is no fee for registration of courses for Minor program offered.

xvi) A student can register for either Minor / Honor but not both.

xvii) Student shall submit an application for either Minor / Honor at least one week before the commencement of the V Semester.

MINOR PROGRAMS OFFERED

Offering Department	Title	Who can Register
Civil Engineering	Civil Engineering	B.Tech. EEE / ME / ECE / CSE / AIML / CSE(AIML) / CSE(DS)
Mechanical Engineering	3D Printing	B.Tech. CE / EEE / ECE / CSE / AIML / CSE(AIML) / CSE(DS)
	Industrial Engineering	B.Tech. CE / EEE / ECE / CSE / AIML / CSE(AIML) / CSE(DS)
Electrical and Electronics Engineering	Energy Systems	B.Tech. CE / ME / ECE / CSE / AIML / CSE(AIML) / CSE(DS)
Electronics and Communication Engineering	Embedded Systems	B.Tech. CE / EEE / ME / CSE / AIML / CSE(AIML) / CSE(DS)
Computer Science and Engineering	Computer Science and Engineering	B.Tech. CE / EEE / ME / ECE
	Data Science	B.Tech. CE / EEE / ME / ECE
	Computer Science and Engineering – Artificial Intelligence and Machine Learning	B.Tech. CE / EEE / ME / ECE

COURSE STRUCTURE

for

MINOR

in

DATA SCIENCE

S.No	COURSE CODE	COURSE TITLE	Semester Offered	L	T	P	IM	EM	CR
1	2332571M	Introduction to Data Science	V	3	0	0	30	70	3
2	2332572M	Big Data Technologies	V	3	0	0	30	70	3
3	2332573M	Data Science Lab	V	0	0	3	30	70	1.5
4	2332671M	Data Engineering	VI	3	0	0	30	70	3
5	2332672M	Introduction to Machine Learning and Deep Learning	VI	3	0	0	30	70	3
6	2332673M	Data Engineering Lab	VI	0	0	3	30	70	1.5
7	2332771M	NoSQL Databases	VII	3	0	0	30	70	3
			Total	**15**	**0**	**6**	**210**	**490**	**18**

2332571M	MINOR IN DATA SCIENCE INTRODUCTION TO DATA SCIENCE (CE,ME,ECE,EEE)	L	T	P	C
		3	0	0	3

Pre-Requisites: NIL

Course Outcomes:

On successful completion of the course, student will be able to

CO1. Develop foundational knowledge and hands-on skills necessary for a career in data science.
CO2. Apply machine learning techniques and Python libraries to real-world data science problems.
CO3. Analyze and process large-scale data using Big Data frameworks and NoSQL databases.
CO4. Analyze graph databases and text mining tools for real-world data problems.
CO5. Design and develop data visualization solutions and dashboards.

SYLLABUS:

UNIT - I: DATA SCIENCE IN A BIG DATA WORLD (07 Periods)

Introduction to Data science, benefits and uses, facets of data, data science process in brief, big data ecosystem and data science.
Data Science process: Overview, defining goals and creating project charter, retrieving data, cleansing, integrating and transforming data, exploratory analysis, model building, presenting findings and building applications on top of them.

UNIT- II: MACHINE LEARNING (10 Periods)

Applications of machine learning in Data science, role of ML in DS, Python tools like sklearn, modelling process for feature engineering, model selection, validation and prediction, types of ML, Semi-supervised learning.
Handling large data: problems and general techniques for handling large data, programming tips for dealing large data, case studies on DS projects for predicting malicious URLs, for building recommender systems.

UNIT- III: NOSQL MOVEMENT FOR HANDLING BIGDATA (09 Periods)

Distributing data storage and processing with Hadoop framework, case study on risk assessment for loan sanctioning, ACID principle of relational databases, CAP theorem, base principle of NoSQL databases, types of NoSQL databases.

UNIT- IV: TOOLS AND APPLICATIONS OF DATA SCIENCE (09 Periods)

Introducing **Neo4j** for dealing with graph databases, graph query language **Cypher**, Applications graph databases, Python libraries like nltk and SQLite for handling Text mining and analytics.

UNIT- V: DATA VISUALIZATION AND PROTOTYPE APPLICATION DEVELOPMENT **(10 Periods)**

Data Visualization options, Cross filter, the JavaScript Map Reduce library, Creating an interactive dashboard with dc.js, Dashboard development tools.

Total Periods: 45

TEXT BOOKS:

T1. Introduction to Data Science using Python Tools, Davy Cielen, Amo D. B. Meysman and Mohamed Ali, Manning Publications Co, Dreamtech Press, 2016.

T2. Data Science with Jupyter, Prateek Gupta, BPB Publications, 2019 for Basics.

REFERENCE BOOKS:

R1. Data Science from Scratch, Joel Grus, OReilly, 2019.

R2. Doing Data Science: Straight Talk From The Frontline, Cathy O'Neil and Rachel Schutt, O'Reilly, First Edition, 2013

Web Resources:

1. https://nptel.ac.in/courses/106106179
2. https://onlinecourses.nptel.ac.in/noc22_cs32/preview
3. https://nptel.ac.in/courses/106106212

2332572M	MINOR IN DATA SCIENCE BIG DATA TECHNOLOGIES (CE,ME,ECE,EEE)	L	T	P	C
		3	0	0	3

Pre-Requisites: NIL

Course Outcomes:

On successful completion of the course, student will be able to

CO1. Understand the characteristics, challenges, and tools of Big Data.
CO2. Implement storage and retrieval mechanisms in HDFS and NoSQL databases.
CO3. Develop MapReduce and Spark-based applications.
CO4. Apply big data analytics techniques to process real-world large-scale datasets.
CO5. Integrate big data platforms with machine learning and business intelligence solutions.

SYLLABUS:

UNIT-I: INTRODUCTION TO BIG DATA AND HADOOP ECOSYSTEM (10 Periods)

Definition, Characteristics of Big Data (Volume, Variety, Velocity, Veracity, Value), Types of Data: Structured, Semi-Structured, and Unstructured, Traditional vs Big Data Systems, Big Data Challenges and Benefits, Introduction to Hadoop: Architecture and Components, Hadoop Distributed File System (HDFS): Features, Design, Blocks, YARN and MapReduce Overview, Hadoop Ecosystem Components: Pig, Hive, HBase, Sqoop, Flume.

UNIT- II: MAPREDUCE PROGRAMMING AND HADOOP TOOLS (10 Periods)

MapReduce Programming Model: Mapper, Reducer, Partitioner, InputSplit and Record Reader, Combiner, Writing MapReduce Programs in Java, Advanced MapReduce Concepts: Counters, Joins, Secondary Sort, Hive: Data Warehousing Concepts, HiveQL, Partitions, Buckets, Pig: Data Flow, Pig Latin Scripts, Data Import & Export with Sqoop, Real-Time Data Collection using Flume.

UNIT - III: NOSQL DATABASES AND HBASE (9 Periods)

Introduction to NoSQL Databases, Types of NoSQL: Key-Value, Document, Column, Graph, Differences between RDBMS and NoSQL, HBase Data Model: Column Families, Regions, Tables, HBase Architecture and Internals, HBase CRUD Operations using Java, Integration of HBase with Hadoop, Case Study: Big Data Storage in Social Media.

UNIT-IV: APACHE SPARK AND BIG DATA ANALYTICS (08 Periods)

Apache Spark: RDDs and DAG Execution Model, Spark Core and Spark SQL, DataFrames and Datasets in Spark, Spark Streaming: Architecture and DStreams, Spark MLlib: Machine Learning on Big Data, GraphX: Graph Processing in Spark, Performance Tuning and Optimization in Spark.

UNIT-V: APPLICATIONS AND CASE STUDIES IN BIG DATA (08 Periods)

Big Data in Healthcare: Predictive Analysis, Genomics, Big Data in Finance: Fraud Detection, Risk Analytics, Big Data in E-Commerce: Customer Behavior, Personalization, Sentiment Analysis using Big Data, Big Data for Smart Cities and IoT, Big Data and Cloud Computing Integration (AWS, GCP, Azure), Data Privacy, Security.

Total Periods: 45

TEXT BOOKS:

T1. Hadoop: The Definitive Guide, Tom White, O'Reilly Media.
T2. Big Data Analytics, V. Srinivasa Subramanian, Wiley India.
T3. Mining of Massive Datasets, Anand Rajaraman and Jeffrey D. Ullman, Cambridge University Press.

REFERENCE BOOKS:

R1. Hadoop in Action, Chuck Lam, Manning Publications.
R2. Taming the Big Data Tidal Wave, Bill Franks, Wiley.
R3. Hadoop in Practice, Alex Holmes, Manning Publications.
R4. Big Data, Big Analytics: Emerging Business Intelligence, Michael Minelli, Wiley.

WEB RESOURCES:

1. Coursera – Big Data Specialization by UC San Diego
2. edX – Big Data Analysis with Apache Spark (UC Berkeley)
3. Udacity – Data Engineering Nanodegree

2332573M	MINOR IN DATA SCIENCE BIG DATA SCIENCE LAB (CE,ME,ECE,EEE)	L	T	P	C
		0	0	3	1.5

Pre-Requisites: Introduction to Programming

Course Outcomes:

On successful completion of the course, student will be able to

CO1. Construct and manipulate multi-dimensional arrays using NumPy for scientific computing.
CO2. Perform data wrangling tasks such as filtering, grouping, merging, and cleaning using pandas.
CO3. Import and handle various file formats including CSV, JSON, Excel, and database sources using Python.
CO4. Implement data preprocessing techniques like scaling, encoding, and imputation for ML readiness.
CO5. Generate/create different types of visualizations (bar, pie, box, scatter, etc.) using matplotlib.

LIST OF EXERCISES/ LIST OF EXPERIMENTS:

Minimum Ten experiments are to be conducted.

1. Creating a NumPy Array
    a. Basic ndarray
    b. Array of zeros
    c. Array of ones
    d. Random numbers in ndarray
    e. An array of your choice
    f. Imatrix in NumPy
    g. Evenly spaced ndarray
2. The Shape and Reshaping of NumPy Array
    a. Dimensions of NumPy array
    b. Shape of NumPy array
    c. Size of NumPy array
    d. Reshaping a NumPy array
    e. Flattening a NumPy array
    f. Transpose of a NumPy array

3. Expanding and Squeezing a NumPy Array
    a. Expanding a NumPy array
    b. Squeezing a NumPy array
    c. Sorting in NumPy Arrays
4. Indexing and Slicing of NumPy Array
    a. Slicing 1-D NumPy arrays
    b. Slicing 2-D NumPy arrays
    c. Slicing 3-D NumPy arrays
    d. Negative slicing of NumPy arrays
5. Stacking and Concatenating Numpy Arrays
    a. Stacking ndarrays
    b. Concatenating ndarrays
    c. Broadcasting in Numpy Arrays
6. Perform following operations using pandas
    a. Creating dataframe
    b. concat()
    c. Setting conditions
    d. Adding a new column
7. Perform following operations using pandas
    a. Filling NaN with string
    b. Sorting based on column values
    c. groupby()
8. Read the following file formats using pandas
    a. Text files
    b. CSV files
    c. Excel files
    d. JSON files
9. Read the following file formats
    a. Pickle files
    b. Image files using PIL
    c. Multiple files using Glob
    d. Importing data from database
10. Demonstrate web scraping using python
11. Perform following preprocessing techniques on loan prediction dataset

a. Feature Scaling
b. Feature Standardization
c. Label Encoding
d. One Hot Encoding

12. Perform following visualizations using matplotlib
    a. Bar Graph
    b. Pie Chart
    c. Box Plot
    d. Histogram
    e. Line Chart and Subplots
    f. Scatter Plot

REFERENCE BOOKS/LABORATORY MANUALS:

1. Introduction to Data Science using Python Tools, Davy Cielen, Amo D. B. Meysman and Mohamed Ali, Manning Publications Co, Dreamtech Press, 2016.
2. Data Science with Jupyter, Prateek Gupta, BPB Publications, 2019 for Basics.
3. Data Science from Scratch, Joel Grus, OReilly, 2019.
4. Doing Data Science: Straight Talk From The Frontline, Cathy O'Neil and Rachel Schutt, O'Reilly, First Edition, 2013

Web Resources:

1. https://www.analyticsvidhya.com/blog/2020/04/the-ultimate-numpy-tutorial-for-data- science-beginners/
2. https://www.analyticsvidhya.com/blog/2021/07/data-science-with-pandas-2-minutes- guide-to-key-concepts/
3. https://www.analyticsvidhya.com/blog/2020/04/how-to-read-common-file-formats- python/
4. https://www.analyticsvidhya.com/blog/2016/07/practical-guide-data-preprocessing- python-scikit-learn/
5. https://www.analyticsvidhya.com/blog/2020/02/beginner-guide-matplotlib-data- visualization-exploration-python/6.
6. https://www.nltk.org/book/ch01.html

2332671M	MINOR IN DATA SCIENCE BIG DATA ENGINEERING (CE,ME,ECE,EEE)	L	T	P	C
		3	0	0	3

Pre-Requisites: NIL

Course Outcomes:

On successful completion of the course, student will be able to

CO1. Describe and differentiate the key concepts, roles, responsibilities, and skills in the field of Data Engineering.
CO2. Analyze the Data Engineering Life Cycle and identify best practices, principles, and undercurrents such as security and orchestration.
CO3. Design and evaluate data architectures and ingestion strategies based on different source systems and enterprise requirements.
CO4. Apply appropriate data storage, transformation, and modeling techniques for both batch and streaming data.
CO5. Develop strategies for efficient querying, data transformation, and serving data for business analytics, machine learning, and reverse ETL pipelines.

SYLLABUS:

UNIT - I: INTRODUCTION TO DATA ENGINEERING (07 Periods)

Definition, Data Engineering Life Cycle, Evolution of Data Engineer, Data Engineering Versus Data Science, Data Engineering Skills and Activities, Data Maturity, Data Maturity Model, Skills of a Data Engineer, Business Responsibilities, Technical Responsibilities, Data Engineers and Other Technical Roles.

UNIT - II: DATA ENGINEERING LIFE CYCLE (08 Periods)

Data Life Cycle Versus Data Engineering Life Cycle, Generation: Source System, Storage, Ingestion, Transformation, Serving Data.
Major undercurrents across the Data Engineering Life Cycle: Security, Data Management, DataOps, Data Architecture, Orchestration, Software Engineering.

UNIT - III: DESIGNING GOOD DATA ARCHITECTURE (11 Periods)

Enterprise Architecture, Data Architecture, Principles of Good Data Architecture, Major Architecture Concepts.
Data Generation in Source Systems: Sources of Data, Files and Unstructured Data, APIs, Application Databases (OLTP), OLAP, Change Data Capture, Logs, Database Logs, CRUD, Source System Practical Details.

UNIT - IV: STORAGE (11 Periods)

Raw Ingredients of Data Storage, Data Storage Systems, Data Engineering Storage Abstractions, Data warehouse, Data Lake, Data Lakehouse.

Ingestion: Data Ingestion, Key Engineering considerations for the Ingestion Phase, Batch Ingestion Considerations, Message and Stream Ingestion Considerations, Ways to Ingest Data.

UNIT - V: QUERIES, MODELING AND TRANSFORMATION (08 Periods)

Queries, Life of a Query, Query Optimizer, Queries on Streaming Data, Data Modelling, Modeling Streaming Data, Transformations, Streaming Transformations and Processing.

Serving Data for Analytics, Machine Learning and Reverse ETL: General Considerations for serving Data, Business Analytics, Operational Analytics, Embedded Analytics, Ways to serve data for analytics and ML, Reverse ETL.

Total Periods: 45

TEXT BOOKS:

T1. Fundamentals of Data Engineering, 1. Joe Reis, Matt Housley, O'Reilly Media, Inc., June 2022, ISBN: 9781098108304

REFERENCE BOOKS:

R1. Data Engineering with Python, Paul Crickard, Packt Publishing, October 2020.
R2. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, Ralph Kimball, Margy Ross, Wiley, 3rd Edition, 2013.
R3. Data Pipelines Pocket Reference: Moving and Processing Data for Analytics, James Densmore, O'Reilly Media, 1st Edition, 2021

WEB RESOURCES:

1. https://elearn.nptel.ac.in/shop/completed-courses/short-term-programs-completed/introduction-to-data-engineering-using-azure/?v=c86ee0d9d7ed

2332672M	MINOR IN DATA SCIENCE BIG INTRODUCTION TO MACHINE LEARNING AND DEEP LEARNING (CE,ME,ECE,EEE)	L	T	P	C
		3	0	0	3

Pre-Requisites: Python Programming, Linear Algebra and Calculus, Probability and Statistics.

Course Outcomes:

On successful completion of the course, student will be able to

CO1. Understanding ML taxonomy (supervised, unsupervised, reinforcement), data pipeline, and overview of tools used in ML.
CO2. Implement and evaluate supervised learning algorithms like linear regression, decision trees, SVM, and KNN using real datasets.
CO3. Apply unsupervised learning techniques such as clustering and dimensionality reduction to explore patterns in unlabelled data.
CO4. Construct and train basic feedforward neural networks using deep learning frameworks such as Tensor Flow or Keras.
CO5. Design and evaluate convolutional and recurrent neural networks for applications in image and sequence data analysis.

SYLLABUS:

UNIT - I: INTRODUCTION TO MACHINE LERNING (07 Periods)

Definition, Scope, Types of machine learning, Applications in real-world domain (healthcare, finance, NLP, vision, etc.,) Data types, ML workflow and pipeline, Overview of tools: Python, scikit-learn, pandas, numpy.

UNIT - II: SUPERVISED LEARNING (10 Periods)

Linear regression, Logistic regression, Decision tress, Randome forests, K-Nearest Neighbors, Support Vector Machines, Evaluation metrics: Accuracy, Presicion, Recall, F1-score, ROC.

UNIT - III: UNSUPERVISED LEARNING (09 Periods)

Clustering: Types of clustering, K-Means, Hierarchical clustering.
Dimensional reduction: PCA, t-SNE
Association rule mining: Apriori, FP-Growth

UNIT - IV: INTRODUCTION TO DEEP LEARNING (09 Periods)

Introduction to neural networks, Perceptron, Activation functions, Feedforward neural networks, Backpropagation and Gradient descent, Regularization techniques: dropout, batch normalization, Overview of TensorFlow and Keras.

UNIT - V: CONVOLUTIONAL AND RECURRENT NEURAL NETWORKS

(10 Periods)

CNN: Architecture, Filters, Pooling, Image classification use cases.
RNN and LSTM: Sequence modeling and Time-series applications.
Basics of transfer learning and pre-trained models, Applications in computer vision and natural language processing.

Total Periods: 45

TEXT BOOKS:

T1. Machine Learning, Tom M. Mitchell, McGraw Hill Education.
T2. Deep Learning, Ian Goodfellow, Yoshua Bengio, and Aaron Courville, MIT Press.

REFERENCE BOOKS:

R1. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, Aurélien Géron, O'Reilly, 2nd Edition.
R2. Introduction to Machine Learning, Ethem Alpaydin, MIT Press, 4th Edition.
R3. Deep Learning with Python, Francois Chollet, Manning Publications.

WEB RESOURCES:

1. https://onlinecourses.nptel.ac.in/noc23_ee87/preview

2332673M	MINOR IN DATA SCIENCE BIG DATA ENGINEERING LAB (CE,ME,ECE,EEE)	L	T	P	C
		0	0	3	1.5

Pre-Requisites: NIL

Course Outcomes:

On successful completion of the course, student will be able to

CO1. Install, configure, and operate open-source tools like NiFi, Airflow, Elasticsearch, and PostgreSQL.
CO2. Design and implement data pipelines to ingest, clean, transform, and load data from various sources.
CO3. Work with structured and unstructured data using Python, NiFi, and Airflow.
CO4. Monitor, version control, and deploy data pipelines in a simulated production environment.
CO5. Build dashboards for data visualization and support analytical decision-making.
CO6. Design and deploy end-to-end production-ready data pipelines incorporating validation, atomicity, idempotency, and performance monitoring.

LIST OF EXERCISES/ LIST OF EXPERIMENTS:

Minimum Ten experiments are to be conducted.

1. Installing and configuring Apache NiFi, Apache Airflow
2. Installing and configuring Elasticsearch, Kibana, PostgreSQL, pgAdmin 4
3. Reading and Writing files
    a. Reading and writing files in Python
    b. Processing files in Airflow
    c. NiFi processors for handling files
    d. Reading and writing data to databases in Python
    e. Databases in Airflow
    f. Database processors in NiFi
4. Working with Databases
    a. Inserting and extracting relational data in Python
    b. Inserting and extracting NoSQL database data in Python
    c. Building database pipelines in Airflow
    d. Building database pipelines in NiFi
5. Cleaning, Transforming and Enriching Data
    a. Performing exploratory data analysis in Python
    b. Handling common data issues using pandas
    c. Cleaning data using Airflow
6. Building the Data Pipeline
7. Building a Kibana Dash Board
8. Perform the following operations
    a. Staging and validating data

   b. Building idempotent data pipelines
   c. Building atomic data pipelines
9. Version Control with the NiFi Registry
   a. Installing and configuring the NiFi Registry
   b. Using the Registry in NiFi
   c. Versioning your data pipelines
   d. Using git-persistence with the NiFi Registry
10. Monitoring Data Pipelines
   a. Monitoring NiFi in the GUI
   b. Monitoring NiFi using processors
   c. Monitoring NiFi with Python and the REST API
11. Deploying Data Pipelines
   a. Finalizing your data pipelines for production
   b. Using the NiFi variable registry
   c. Deploying your data pipelines
12. Building a Production Data Pipeline
   a. Creating a test and production environment
   b. Building a production data pipeline
   c. Deploying a data pipeline in production

REFERENCE BOOKS/LABORATORY MANUALS:

R1. Data Engineering with Python, Paul Crickard , Packt Publishing, October 2020.
R2. Data Engineering with Python, Paul Crickard (Packt Publishing)
R3. Streaming Data, Andrew G. Psaltis (O'Reilly Media)
R4. Fundamentals of Data Engineering, Joe Reis and Matt Housley (O'Reilly Media)

WEB RESOURCES:

NiFi, Airflow, and Elasticsearch official documentation

1. https://nifi.apache.org/
2. https://airflow.apache.org/
3. https://www.elastic.co/

2332771M	MINOR IN DATA SCIENCE BIG NOSQL DATABASES (CE,ME,ECE,EEE)	L	T	P	C
		3	0	0	3

Pre-Requisites: Database Management Systems

Course Outcomes:

On successful completion of the course, student will be able to

CO1. Discuss what type of NoSQL database to implement based on business requirements (key-value, document, full text, graph, etc.) Interpret and summarize on NoSQL, Cassandra.
CO2. Apply NoSQL data modeling from application specific queries Make use of Apache Spark, RDDs etc. to work with datasets.
CO3. Use Atomic Aggregates and denormalization as data modeling techniques to optimize query processing.
CO4. Distinguish various NOSQL database stores.
CO5. Apply Indexing Techniques of NoSQL.

SYLLABUS:

UNIT - I: Introduction to NoSQL (10 Periods)

Introduction to NoSQL: Definition and Introduction, Sorted Ordered Column-Oriented Stores, Key/Value Stores, Document Databases, Graph Databases, Examining Two Simple Examples, Location Preferences Store, Car Make And Model Database, Working With Language Bindings.

UNIT - II: Interacting with NoSQL (09 Periods)

If NoSql Then What, Language Bindings For NoSQL Data Stores, Performing Crud Operations, Creating Records, Accessing Data, Updating And Deleting Data.

UNIT - III: NoSQL Storage Architecture (09 Periods)

Working With Column-Oriented Databases, Hbase Distributed Storage Architecture, Document Store Internals, Understanding Key/Value Stores In Memcached And Redis, Eventually Consistent Non Relational Databases.

UNIT - IV: NoSQL Stores (08 Periods)

Similarities Between SqlAndMongodb Query Features, Accessing Data From Column- Oriented Databases Like Hbase, Querying Redis Data Stores, Changing Document Databases, Schema Evolution In Column-Oriented Databases, Hbase Data Import And Export, Data Evolution In Key/Value Stores.

UNIT - V: Indexing and Ordering Data Sets (09 Periods)

ndexing and Ordering Data Sets: Essential Concepts Behind A Database Index, Indexing And Ordering In Mongodb, Creating and Using Indexes In Mongodb, Indexing And Ordering In Couchdb, Indexing In Apache Cassandra.

Total Periods: 45

TEXT BOOKS:

T1. Professional NoSQL, Shashank Tiwari, Wrox Press, Wiley, 2011, ISBN: 978-0-470-94224-6.
T2. NoSQL Distilled, Pramod Sadalage and Martin Fowler, Addison-Wesley Professional, 2012.

REFERENCE BOOKS:

R1. Making Sense of NoSQL, Dan McCreary and Ann Kelly, Manning Publications, 2013.
R2. Getting Started with NoSQL, Gaurav Vaish, Packt Publishing, 2013.

WEB RESOURCES:

1. https://www.coursera.org/learn/introduction-to-nosql-databases
2. https://www.coursera.org/specializations/nosql-big-data-and-spark-foundations
3. https://www.coursera.org/learn/nosql-databases