

## **DEPARTMENT OF CSE - DATA SCIENCE**

**Event: "Big Data Architecture and ETL Pipelines"** 

Venue: Room C-222

Time: 09:00AM to 11:00 AM



Department of Computer Science and Engineering (Data Science)



# Big Data Architecture and ETL Pipelines



24<sup>th</sup> April 2025



Room C-222

09:00 AM to 11:00 AM





Dr. Mounica

Sr. Technology Business Consultant Wells Fargo

Faculty Coordinator **Ms. Swati Sehgal** Assistant Professor - CSE(DS) Convenor **Dr. B Swathi** HOD - CSE(DS) On 24<sup>th</sup> of April 2025, the Department of Computer Science and Engineering (Data Science), hosted an Expert Talk on **"Big Data Architecture and ETL Pipelines."** The Department of Computer Science and Engineering (Data Science) organized an expert talk on **Big Data Architecture and ETL Pipelines** aimed at enhancing students' understanding of modern data processing techniques. The session was conducted by Dr. Mounica, a seasoned professional with vast experience in Technology Consulting at Wells Fargo.

The session began with an introduction to **Big Data** and its growing significance in today's datadriven world. Dr. Mounica explained the challenges associated with handling massive datasets and emphasized the importance of efficient architectures and data pipelines for business intelligence and analytics.



#### **Key Topics Covered**

• Big Data Concepts and Architecture

Dr. Mounica discussed the layered architecture of big data systems, focusing on data

ingestion, storage, processing, and visualization. She explained how scalable and distributed frameworks are critical in managing big data workloads.

#### • Python for Big Data:

She highlighted the use of **Python** in big data environments, detailing popular libraries and frameworks that are essential for data manipulation, preprocessing, and integration tasks.

#### • Introduction to PySpark

A major part of the talk was devoted to **PySpark**, an interface for Apache Spark in Python. Dr. Mounica demonstrated how PySpark simplifies distributed data processing and accelerates large-scale data analytics. She discussed RDDs (Resilient Distributed Datasets), DataFrames, key transformations, and actions used in PySpark.

### Data Visualization with Power BI

The session concluded with an insightful demonstration on **Power BI**, a business analytics service by Microsoft. Dr. Mounica illustrated how data processed through ETL pipelines can be effectively visualized using Power BI dashboards, making complex data understandable and accessible to business users.



The expert talk provided students with a strong foundational understanding of how big data solutions are designed and implemented in the Industry. Dr. Mounica's practical insights into **Python, PySpark,** and **Power BI** bridged the gap between theoretical concepts and real-world applications. The session was highly interactive, with students engaging actively during the Q&A segment.

The Department extends its heartfelt gratitude to Dr. Mounica for delivering such an informative session and inspiring students to explore the exciting domain of big data analytics.



**Faculty Coordinator** 

HOD

Prof. PALLAVI NAYAK

Dr. BASAVARAJU SWATHI