



Event Report

Workshop Title: Data Engineering & Analytics with PySpark

Speaker: Mr. Tushar Goyal, Sr. Data Engineer, REGex Software Services

Dates: 7th - 8th April 2023

Venue: NJR i3 Labs, Techno NJR Knowledge Campus, Udaipur

Introduction

A two-day workshop on "Data Engineering & Analytics with PySpark" was successfully organized, featuring Mr. Tushar Goyal, a seasoned Senior Data Engineer from REGex Software Services. The event was aimed at equipping participants with hands-on skills and advanced knowledge in data engineering and analytics using PySpark, a leading big data processing framework.

Overview

The workshop was divided into theory and practical sessions over two days, covering the following key topics:

1. Day 1 - Foundational Concepts

- Introduction to Big Data and PySpark
- Understanding RDDs (Resilient Distributed Datasets) and DataFrames
- Setting up PySpark and practical hands-on installation
- PySpark's architecture: Spark Core, Spark SQL, and Spark Streaming

2. Day 2 - Advanced Techniques and Real-world Applications

- Advanced transformations and actions in PySpark
 - Handling structured and unstructured data
 - Data cleaning and preprocessing with PySpark
 - Use cases in real-world analytics projects
 - Best practices in PySpark for scalable data engineering
-

Speaker's Highlights

Mr. Tushar Goyal brought a wealth of experience to the workshop, sharing insights from his professional journey and presenting real-world case studies. His teaching methodology included:

- Clear explanations of complex concepts
 - Live coding demonstrations
 - Interactive Q&A sessions to address participant queries
-

Participation and Engagement

The workshop witnessed enthusiastic participation from a diverse audience of students, data professionals, and researchers. Key highlights of participant engagement included:

- Hands-on exercises with live datasets
 - Group discussions to solve data engineering challenges
 - Personalized feedback from the speaker
-

Key Takeaways

Participants gained:

- A deep understanding of PySpark's architecture and components
 - Practical experience in processing large-scale datasets
 - Skills to implement scalable data pipelines
 - Knowledge of industry best practices for data engineering
-

Feedback

The workshop received overwhelmingly positive feedback. Attendees appreciated the clarity of explanations, the balance of theory and practicals, and the expertise of the speaker.

Conclusion

The two-day workshop was a resounding success, providing valuable learning opportunities for all attendees. It served as a platform to enhance skills in data engineering and analytics with PySpark and inspired participants to explore further applications of big data technologies in their respective domains.

Acknowledgment

Special thanks to Mr. Tushar Goyal for his exceptional contribution, and to the organizing committee for their efforts in ensuring a seamless event experience.