

## Hands-On workshop on Big Data Processing with Hadoop, Pig, and Hive

On October 7, 2024, our university hosted an insightful workshop on Big Data, Hadoop, Hive, and Pig at the Magis Block Conference Room. The workshop targeted final-year MSc students and faculty members, focusing on the rapidly advancing landscape of big data, its broad applications, and the technical challenges it presents. The session commenced with a prayer by Mr. Melwil, followed by Mr. Reaven's introduction of the speaker, Mr. Puneeth—a seasoned industry professional. Mr. Puneeth then led the workshop.

### **Topics Covered:**

- **Data Engineering:** Emphasized the importance of data, addressing key questions about data's role and its increasing value.
- **Image Clarity in Phones:** Highlighted technological advancements in phone cameras, discussing resolutions up to 50MP and data processing requirements.
- **Origins of Big Data:** Introduced the concept of big data, referencing examples from communication and aviation industries.
- **Data Characteristics:** Explained big data's core attributes—Volume, Variety, and Velocity (the three V's).
- **Hadoop Infrastructure:** Discussed the master-slave architecture, including components such as the job tracker, worker nodes, name node, and block-level reporting.
- **Linux Compatibility:** Highlighted Hadoop's dependence on Linux environments.
- **Data Storage:** Explained Hadoop's data storage framework, including racks, job tracking services, and metadata organization.
- **Hadoop Ecosystem:** Provided an overview of Hadoop's structure, applications, and certain limitations.
- **Interactive Session:** Facilitated hands-on experience with tools like Hive, Pig, and Sqoop, where participants executed commands and explored big data tools in action.

**Conclusion:** The workshop effectively illuminated the importance of big data in driving innovation across sectors. Participants left with a strong foundational

understanding of big data's transformative potential and practical skills in using big data tools.

**Thanksgiving:** The session concluded with Gavin Livero delivering a vote of thanks, expressing appreciation to the speakers, organizers, and participants for their engagement and contributions. Dr. Annie Syrien moderated the session, ensuring a smooth and enriching experience for all attendees.

For the Images of the workshop on Big Data Processing with Hadoop, Pig, and Hive, kindly refer to the Photo Gallery:  
<https://www.sju.edu.in/gallery/113>