VIGNAN'S NIRULA
INSTITUTE OF TECHNOLOGY & SCIENCE FOR WOMEN
(Approved by AICTE, New Delhi & Affiliated to JNTU Kakinada)
Vignan Avenue, Peda Palakaluru, Guntur - 522009
Department of Information Technology

Innovative Practice

Course Information:

Faculty Name	**:**	Mrs.T.Naga Navya,Mr.B.V.Suresh Kumar
Course Name	:	Big Data Analytics
Class	**:**	III B. Tech II Semester
Academic Year	**:**	2024-2025
Title of the Topic	**:**	HDFS design
Activity Name	**:**	Collaborative Learning

Objective:
Students will collaboratively explore and analyze the design and architecture of Hadoop Distributed File System (HDFS) in the context of Big Data Analytics. They will apply concepts related to data storage, fault tolerance, scalability, and performance optimization, gaining a deep understanding of how HDFS manages large datasets across distributed systems.

Steps to Implement Collaborative Learning:

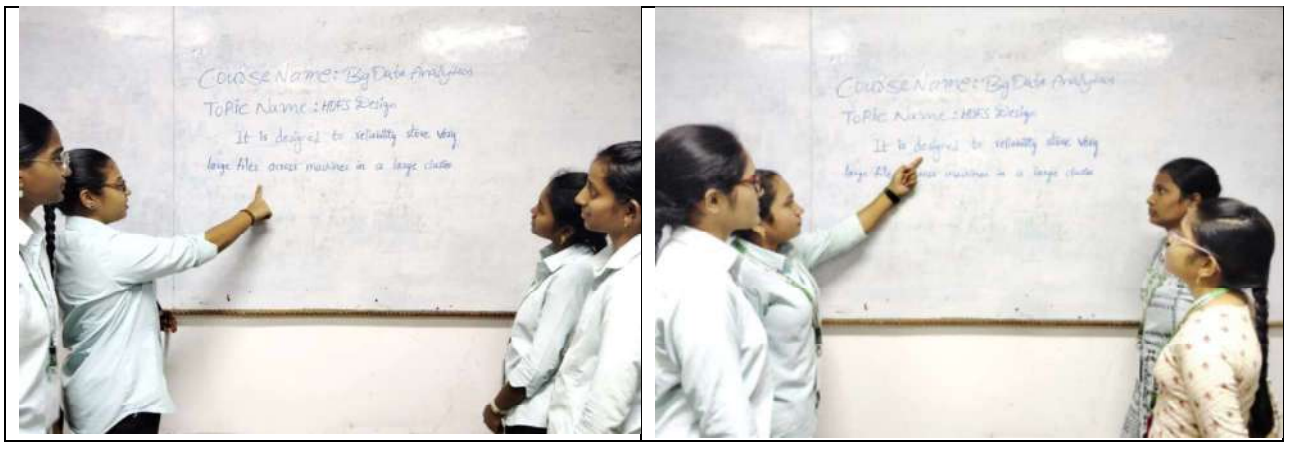
1. **Assign Problems:**
   - Divide the class into small groups.
   - Provide each group with a specific problem related to HDFS design and architecture.

Problems for Students Related to HDFS Design:

1. **HDFS Architecture and Data Storage:**
   - **Problem:** Imagine you are designing an HDFS-based system to store large amounts of data for a global e-commerce platform. The system needs to handle high-volume traffic and provide fast access to data.
   - **Task:** Explain how data is stored in HDFS, focusing on block size, data replication, and the role of the NameNode and DataNodes in the system.
   - **Question:** How would you configure the block size and replication factor to optimize performance for both storage capacity and fault tolerance?
2. **Fault Tolerance in HDFS:**
   - **Problem:**
     You are working on a Big Data project using HDFS where fault tolerance is crucial for the project's success. A failure occurs where one of the DataNodes becomes unavailable.
   - **Task:** Discuss how HDFS ensures fault tolerance when a DataNode fails and how data is recovered.
   - **Question:** How does HDFS handle multiple DataNode failures, and what strategies could be used to reduce the risk of data loss?

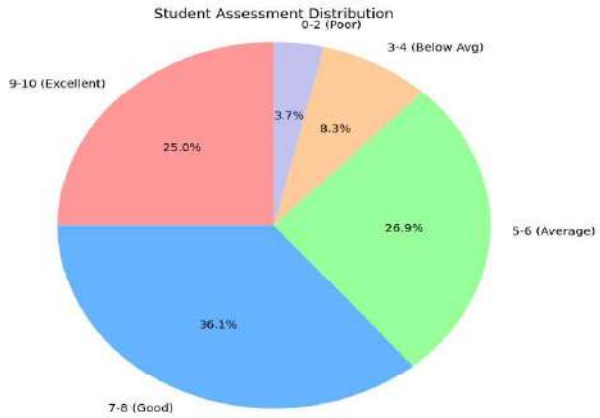
2. **Collaborative Activity:**
    - Students work in their groups to research and solve their assigned problem related to HDFS design.
    - Each group discusses the technical details of HDFS, including storage mechanisms, fault tolerance, scalability, and performance optimization.
    - Groups collaborate to come up with solutions and strategies based on real-world Big Data requirements.
3. **Phase 1 – Think (5-7 minutes):**
    - Students work individually or in pairs within their group to analyze the problem and come up with initial ideas for their solutions.
    - They write down their approach to solving the problem, focusing on key concepts such as block replication, NameNode/DataNode interaction, and fault tolerance strategies.
4. **Phase 2 – Pair (10-15 minutes):**
    - Students collaborate with their group members, share their ideas, and combine their findings to form a complete solution.
    - They discuss how different design choices might impact HDFS performance, scalability, and fault tolerance.
    - Each group prepares a brief explanation of their approach and prepares to share it with the class.
5. **Phase 3 – Share (10-12 minutes):**
    - Each group presents their findings to the class, explaining how their approach addresses the problem and the key decisions they made related to HDFS design.
    - A class-wide discussion follows, where the instructor facilitates comparisons of different HDFS strategies, highlighting how HDFS architecture can be optimized for different Big Data use cases.
    - Other groups can ask questions and offer suggestions, encouraging peer-to-peer learning.
6. **Wrap-Up (5 minutes):**
    - Reflect on the key concepts of HDFS design learned during the activity, such as fault tolerance, scalability, and data access optimization.
    - Students share insights on how they might apply HDFS design principles to real-world Big Data projects.
    - The instructor summarizes the discussion, emphasizing the importance of understanding HDFS architecture for handling large-scale distributed data storage efficiently.

Screenshot of the Practice



Assessment Summary

Marks Range	Number of Students	Percentage
9-10 (Excellent)	27	25%
7-8 (Good)	39	36.11%
5-6 (Average)	29	26.85%
3-4 (Below Avg)	9	8.33%
0-2 (Poor)	4	3.70%
Total	**108**	**100%**



Conclusion:

The design of Hadoop Distributed File System (HDFS) plays a fundamental role in enabling scalable, fault-tolerant, and efficient data storage in Big Data Analytics. HDFS is specifically optimized for storing large datasets across a distributed network, allowing businesses and organizations to manage and process data at massive scales.

Signature of the Faculty **Head of the Department**