PES UNIVERSITY
Department of CSE
UE22CS343BB3 - DATABASE TECHNOLOGIES
Jan – May 2025

DBT PROJECT

Mode: TEAM

Professional Report Submission Due Date: **21.04.2025**
Project Report File Naming Convention: DBT ProjRpt SRN1-SRN2-SRN3-SRN4 (only last 3 digits of each SRN)
Eg.: *DBT ProjRpt 001-002-003-004*
Team Presentations Dates: **22.04.2025 - 24.04.2025**

[A] Technologies/Frameworks to be Utilized for Streaming:

1. Apache Spark Streaming and Spark SQL: Implement various workloads such as executing Spark SQL queries for actions, transformations, or aggregations on input data.
2. Apache Kafka Streaming: Utilize publishing/subscribing mechanisms or producer/consumer functionalities with a **minimum of three topics**.
3. Data Storage: Choose a DBMS (e.g., PostgreSQL, MySQL) for storing data after stream processing [will be utilized for Batch processing later].
4. Additional Tools: Employ other necessary tools like Zookeeper.

[B] Batch Mode Execution:

Run identical queries in batch mode on the entire dataset retrieved from the database mentioned in point 3.

[C] Comparison and Evaluation:

Evaluate and compare the results, accuracy, and performance between the streaming mode and batch mode of execution.

Language: Java/Python/C

Example of Streaming Input Data: Twitter feeds (tweets), newsfeeds, etc.

Computation Examples:

- Counting tweets within a specified window
- Grouping tweets by hashtags
- Applying aggregate functions (e.g., min, max, avg) on numerical data within tumbling windows

Consumption:

- Store tweets into a database for further processing, such as batch mode processing.

Note:

- The window size should be significant and relevant to the chosen domain problem (e.g., 15-30 minutes of tweets as one window).
- Provide a detailed explanation of the selected problem and its corresponding solution, including techniques applied such as sliding or tumbling windows.
- Document all details including code (SQL, Java/Python), input and output data snapshots, etc., in the project report.
- **Ensure that the background color of screens in the snapshots is white before capturing them.**
- Strictly follow the naming conventions specified.
- Use the project report template provided. **The professional report should be submitted in a PDF file.**

Good Luck!