METCS777

Big Data Analytics

Term Paper Proposal

Title of the Paper

Visualization of Big Data in Cloud Environments

A Comparative Study of Spark SQL and Google Data Studio Dashboards

Team Member Names

1. Tanvi Thopte-U35489362
2. Mehul Bisht-U93099406

Abstract

Our term paper focuses on visualisation of big data in cloud environments, exploring how modern cloud-based systems enable large-scale, interactive, and efficient data visualisation for analytics and decision-making.

We chose this topic because visualization is one of the most crucial aspects of data analytics; it allows organisations to turn complex, large-scale data into clear, actionable insights. With the rapid growth of cloud computing and distributed systems, visualizing massive datasets has become both a challenge and an opportunity. Understanding how different platforms handle this process is essential for anyone pursuing a career in data analytics or business intelligence.

In this project, we plan to study and compare multiple approaches for big data visualization.

First, we will use:

- Apache Spark SQL on Databricks to process and visualize large datasets using Python-based libraries such as Matplotlib or Plotly.

Then, we will use:

- Google BigQuery integrated with Google Data Studio (Looker Studio) to create interactive, cloud-hosted dashboards.
- We also plan to explore additional visualization tools like Power BI and Tableau Public to understand their integration capabilities and performance when connected to cloud-based datasets.

By comparing these different tools and workflows, we aim to evaluate their:

- Performance
- Scalability
- Interactivity
- Overall ease of use for analytical reporting

- This research will help identify best practices for visual analytics in cloud-based environments and provide insights into which platforms are most effective for large-scale data storytelling.

List of Papers / Articles / Resources

Databricks Documentation – “Visualizing Data with Spark SQL.”

Google Cloud Documentation – “Connecting BigQuery to Data Studio.”

Tableau Whitepaper – “Visual Analytics for Big Data.”

Microsoft Power BI Documentation – “Connecting Power BI to Google BigQuery.”

IEEE Access (2022) – “Visualization of Big Data: Tools and Techniques.”

Article – “Big Data Visualization: Tools, Challenges, and Techniques,” Towards Data Science (2023).

Kaggle Dataset – NYC Taxi Trips Dataset (for large-scale data visualization).

Zaharia, M. et al. (2016). “Apache Spark: A Unified Engine for Big Data Processing,” Communications of the ACM.

Google BigQuery Whitepaper – “Google BigQuery: Scalable Data Warehousing for Analytics.”

-