CS30202

Database Management Systems

Term Project

c,s,v

21CS10005 - Anit Mangal

21CS10026 – Gorantla Thoyajakshi

21CS10037 – Khimasia Krish

21CS30031 - Mihir Mallick

21CS30054 – Tanishq Prasad

Roman's Taxi Service

Optimizing Roman's Taxi Service Operations through a Distributed Database System

Abstract:

This project aims to develop a comprehensive, distributed database system for Roman's Taxi Service, leveraging the New York City Taxi and Limousine Commission (TLC) Trip Record Data. Our objective is to enhance the company's operational efficiency, customer satisfaction, and profitability through data-driven insights. By employing a distributed database architecture on Hadoop with Apache HBase and integrating data via Protocol Buffers, we intend to create a robust platform for analyzing vast datasets related to taxi operations. This system will support complex queries, map-reduce operations, and data visualization, providing Roman's Taxi Service with actionable insights into customer behavior, driver performance, and revenue trends. Through this project, we aim to empower Roman's Taxi Service to make informed decisions, optimize service delivery, and maintain a competitive edge in the taxi service sector.

Introduction:

Roman's Taxi Service operates as a premier taxi service provider, catering to the transportation needs of its clientele. Focused on enhancing efficiency, the company leverages technology and data-driven approaches to streamline its operations. To this end, Roman's Taxi Service has developed a bespoke web application, furnishing employees and administrators with a comprehensive interface for accessing and analyzing diverse datasets to facilitate informed decision-making.

The project centers on providing Roman's Taxi Service with a robust platform for data analysis and operational optimization. Through the web application, users

can glean valuable insights spanning customer behavior, revenue trends, driver performance, and more. This empowers personnel to make data-informed decisions and implement targeted strategies for improving overall efficiency and service quality.

By harnessing the capabilities of this application, Roman's Taxi Service's workforce can identify underperforming drivers, fine-tune route planning to enhance efficiency, and tailor services to meet customer preferences. The Roman's Taxi Service web application emerges as a pivotal tool for optimizing operations, enhancing customer satisfaction, and driving sustained profitability for the company.

Motivation:

The driving force behind the development of Roman's Taxi Service web application lies in empowering the company with data-centric insights to refine its operational efficiency and bolster profitability.

In the face of escalating competition within the taxi service sector, companies are compelled to continually elevate their operational standards to maintain a competitive edge. Recognizing this imperative, Roman's Taxi Service embarked on a technological journey to streamline its operations and elevate service standards for its clientele.

Through meticulous analysis of pertinent data encompassing customer behavior, driver performance metrics, and revenue trends, Roman's Taxi Service endeavors to pinpoint areas ripe for enhancement and execute informed strategies to optimize operational workflows. This concerted effort translates into heightened customer satisfaction levels, enhanced customer retention rates, and augmented profitability for the company.

Facilitated by the web application, Roman's Taxi Service's employees and administrators gain unfettered access to invaluable data insights. This enables proactive issue identification and swift remedial action implementation. For instance, in response to sudden spikes in demand within specific regions, the company can swiftly deploy additional drivers to ensure prompt service delivery, thereby enhancing customer experiences and bolstering its competitive stance.

Dataset:

The dataset utilized is the New York City Taxi and Limousine Commission (TLC) Trip Record Data, providing comprehensive insights into the operations of New York-based taxis.

Containing a wealth of information, this dataset encompasses details such as pickup and drop-off locations, trip distances, durations, fare amounts, payment methods, and passenger counts. Additionally, it offers insights into taxi types, including service classifications (e.g., medallion taxi, street hail livery), passenger capacities, and accepted payment modes.

Furthermore, the dataset includes supplementary attributes such as trip timestamps, weather conditions, and precise pickup and drop-off coordinates. These attributes are instrumental in facilitating nuanced analysis within the web application, enabling insights into customer demand dynamics, driver performance metrics, and traffic patterns.

Sourced from the New York City Taxi and Limousine Commission, this dataset undergoes regular updates and spans several years, comprising millions of meticulously recorded trip records. Its comprehensive nature renders it an invaluable resource for researchers, analysts, and developers engaged in various transportation-related projects.

Design:

The design of the web-based application, leveraging the New York City Taxi and Limousine Commission (TLC) Trip Record Data, for facilitating query execution and analysis by Roman's Taxi Service employees and administrators can be outlined as follows:

User Interface:
The application will boast a user-friendly interface, offering intuitive options for query execution and report generation via a comprehensive dashboard.

Query and Analysis Section:
Users can specify parameters such as date range, taxi type, passenger count, pickup/drop-off locations, and other relevant criteria within the query and analysis section. This functionality enables tailored queries about passenger demand, driver performance, revenue metrics, and other operational facets of the taxi service.

Visualization and Reporting:

Query results will be presented in easily interpretable formats, including tables, charts, and graphs, facilitating trend identification and pattern discernment within the data. This visualization and reporting feature empower users to derive actionable insights, fostering informed decision-making to enhance the taxi service.

Data Management:
The system will possess robust capabilities for handling extensive datasets and seamless integration with the web application, ensuring efficient data management and accessibility.

Security and Access Control:
Backend data updates will be seamlessly facilitated, while stringent access controls within the interface will prevent unauthorized data manipulation by web application users, thereby safeguarding against data contamination.

The web-based application will offer Roman's Taxi Service employees and administrators a complete interface for executing queries and accessing analyses pertinent to taxi service management. This includes user-friendly navigation, robust query capabilities, visual reporting tools, efficient data management, and stringent security measures.

Frameworks:

Our framework will make use of the following programs:

- Python
- Hadoop/Apache HBase
- Protocol Buffers
- HTML, CSS, JavaScript
- Flask

Weekly Work Plan

Week 1: Preparation and Initial Setup

- Install Hadoop/HBase on a chosen infrastructure (laptop/server cluster).
- Review the New York City TLC Trip Record Data and plan data integration strategies.

Week 2: Data Integration and System Design

- Load the initial datasets to nodes in the Hadoop cluster.
- Design the system architecture, focusing on the query and analysis section, data management, and security measures.
- Begin development on user interface components using HTML, CSS, JavaScript, and Flask.

Week 3: Development and Testing

- Implement map-reduce operations for processing and analyzing data.
- Pipe map-reduce outputs using Protocol Buffers to integrate multiple data sources effectively.
- Conduct initial testing on simple queries and map-reduce operations to ensure system functionality and data accuracy.

Week 4: Finalization and Presentation

- Finalize the user interface and visualization components of the web application.
- Perform comprehensive testing of the entire system, including data analysis capabilities, security, and user interface.
- Prepare a presentation and demo of the project, showcasing its features, capabilities, and potential impact on Roman's Taxi Service's operations.

This detailed plan aims to deliver a fully functional distributed database system for Roman's Taxi Service, addressing the company's need for a data-driven approach to operational optimization within a tight four-week timeline.