**Baltimore-311 Public Grievances Data**

**Final Project**

**Presented By:**

**Akhil Mittapalli**

**Nallamalli Vishnu Vikas**

**Gaurav Dali**

**Aravind Kumar Yaleti**

**University of Maryland, Baltimore County**

**Department of Data Science**

**Date: 05/14/2024**

**Project Advisor: Mesfin Abate**

**Contents**

**1. Business Process and Requirements**

**1.1 Database System Overview**

**1.1.1 Customers**

**1.1.2 Benefits**

**1.2 Requirements**

**1.2.1 Data to be Captured**

**1.2.2 Data Types**

**1.2.3 Data Entry and Capture**

**1.2.4 Reports Needed**

**1.3 Project Risks**

**2. Design Specifications**

**2.1 Relational Concepts**

**2.1.1 ER Diagram and Technology**

**2.2 Big Data Concepts**

**2.2.1 Big Data Architecture and Technology**

**2.3 Cost Estimates**

**2.3.1 Hardware**

**2.3.2 Software**

**2.3.3 Staff Resources**

**2.4 Roles and Work Performed by Each Team Member**

**2.4.1 Akhil Mittapalli - Project Manager**

**2.4.2 Nallamalli Vishnu Vikas - Data Engineer**

**2.4.3 Gourav Dali - Data Analyst**

**2.4.4 Aravind Kumar Yaleti - Visualization Specialist**

**Baltimore-311 Public Grievances Data**

1. **Business Process and Requirements**
   1. **Database System Overview**

The project has the goal of designing a complete system for the analysis and visualization of 311 data which comprises of information related to service requests, complaints, and inquiries of citizens. The main recipients of this system are city officials, urban planners, policymakers, community leaders, and researchers.

* + 1. **Customers and Benefits:**

1. **City Officials and Administrators:**

**Role:** These are the main users who will utilize the system to observe city services and promptly settle citizen complaints.

**Benefits:**

**Real-time Monitoring:** The process of notification of the problems gives us the chance to recognize the emerging issues immediately, which, in turn, allows for the quicker response times.

**Resource Allocation**: Helps in the process of prioritizing tasks based on the insights that are gathered from the data, thus ensuring the best utilization of the resources.

**Performance Metrics:** The feature will be able to give the service performance data which will thus be used to evaluate the service delivery and will help in the improvement of the service delivery.

1. **Urban Planners and Policymakers:**

**Role:** Scheme the system to plan and deliver the urban development projects and policies.

**Benefits:**

**Data-Driven Planning:** Collection of the full data enables the making of the right-decision for the urban development.

**Trend Analysis**: It detects the long-term trends and patterns, thereby, helps in the formulation of the proactive policies and planning.

1. **Community Leaders and Activists:**

**Role:** Advocate for community needs and monitor government responsiveness.

**Benefits:**

**Transparency:** Provides a clear picture of government actions and responsiveness, fostering accountability.

**Engagement:** Empowers leaders with data to engage with residents and push for necessary changes.

1. **Researchers and Analysts:**

**Role:** Conduct studies on urban dynamics and the effectiveness of city services.

**Benefits:**

**Comprehensive Data:** Offers a rich dataset for academic and practical research, supporting urban studies and innovations.

**Collaboration:** Facilitates collaboration across different fields to address urban issues.

* 1. **Requirements:**
     1. **Data to be Captured:**

**Service Requests:** Includes details such as type of service, request date, resolution date, and current status. This helps track the lifecycle of each service request.

**Complaints:** Captures category, detailed description, location, date, and status. Essential for identifying common issues and areas requiring attention.

**Inquiries:** Logs types of inquiries, response times, and resolution statuses, providing insights into frequently asked questions and information needs.

**Feedback and Suggestions:** Records type, description, location, and date, which are crucial for understanding public sentiment and areas for improvement.

* + 1. **Data Types:**

**Text:** Through the use of detailed descriptions of service requests, complaints, and feedback, it is possible to capture the essence of the situation and get an idea of what is going on.

**Numeric:** Unique request IDs, timestamps, and response times are the keys for tracking the essential features of request tracking.

**Geospatial:** Latitude and longitude are the primary factors used for mapping service requests and for the analysis of geographic trends.

**Categorical:** Service type, complaint category, and status are the means to classify the data, thereby, making it easy to sort and filter.

* + 1. **Data Entry and Capture:**

**Automatic:** Integration with existing 311 systems and city databases ensures real-time data ingestion and minimizes manual errors.

**Manual:** Allows for administrative adjustments and data corrections, ensuring accuracy and completeness.

**APIs:** Facilitates seamless data exchange with external sources, enhancing scalability and integration capabilities.

* + 1. **Reports Needed:**

**Operational Reports:** Provide daily and weekly updates on the status and resolution of service requests, helping managers track performance.

**Analytical Reports:** Offer in-depth monthly and quarterly analyses of trends, anomalies, and performance metrics.

**Geospatial Reports:** Utilize maps to show the distribution of requests and complaints, aiding in geographic analysis and planning.

* 1. **Project Risks:**

**Data Quality Issues:** The highlighted issue of data inconsistencies and missing values will be solved by the data cleaning and validation processes which are the main components of rigorous data cleaning and validation.

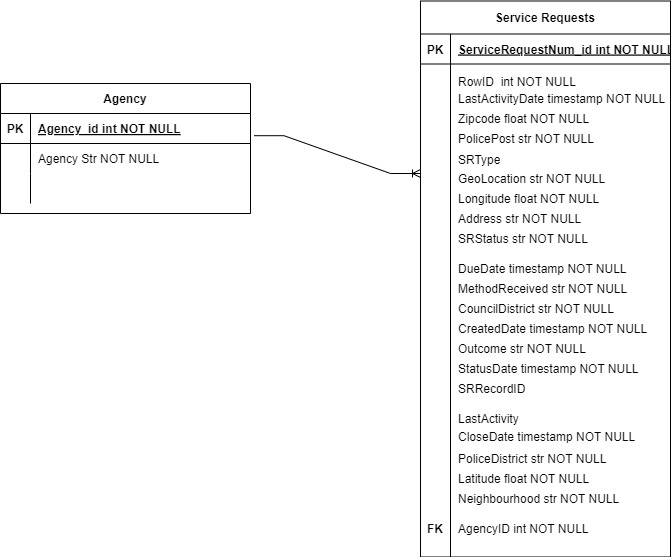
**Scalability Challenges:** The mechanism should be able to handle large datasets effectively, for example, via cloud-based solutions.

**Data Security:** Strictly enforcing stringent security measures to keep citizens' sensitive information safe.

**Technical Integration:** The provision of compatibility with the existing city systems and the ease of interaction of data are the requirements.

1. **Design Specifications**
   1. **Relational Concepts**
      1. **ER Diagram and Technology:**

**Entity-Relationship Diagram:** Illustrates the relationships between entities like ServiceRequest and Agency. Attributes such as request ID, date, status, type, and location define these relationships.



**Technology:**

**Google BigQuery:** Selected for its scalability and robust data processing capabilities, allowing for efficient storage and analysis of large datasets.

* 1. **Big Data Concepts**
     1. **Diagram and Technology:**

**Big Data Architecture:** The system is based on the ETL (Extract, Transform, Load) process that is responsible for data flow:

**Extract:** Information is collected from different sources, and the Open Baltimore website is one of them.

**Transform:** Data cleaning (vacancy of dublicates, missing values, handling of missing values), transformation (data type normalization, calculations, creating of new fields) and validation are the processes that make sure the data quality.

**Load:** The transformed data is then imported into Google BigQuery for analysis.

**Technology:**

**Google BigQuery:** Through its massive hardware capability, it stores and processes huge amounts of data.

**Power BI:** The software is used for the interactive visualizations and dashboards creation.

* 1. **Cost Estimates**
     1. **Hardware:**

**Cloud Infrastructure:** Estimated at $500 per month for Google BigQuery services, which includes storage and processing costs.

* + 1. **Software:**

**Power BI Licenses:** Estimated at $20 per user per month, providing access to powerful visualization tools.

* + 1. **Staff Resources:**

**Data Engineers:** Responsible for managing ETL processes, with an estimated annual cost of $100,000.

**Data Analysts:** Conducting data analysis and generating insights, estimated at $80,000 annually.

**Project Manager:** Overseeing the project execution and team coordination, estimated at $120,000 annually.

**Total Annual Cost:** Approximately $300,000, covering essential personnel and infrastructure.

* 1. **Roles and Work Performed by Each Team Member**
     1. **Akhil Mittapalli:**

**Role:** Project Manager

**Responsibilities:** Overseeing project execution, coordinating team activities, managing stakeholder communications, and ensuring project milestones are met.And also worked on performing ETL tasks.

* + 1. **Nallamalli Vishnu Vikas:**

**Role:** Data Engineer

**Responsibilities:** Handling data extraction, transformation, and loading processes. Ensuring data quality and integration with Google BigQuery.

* + 1. **Gaurav Dali:**

**Role:** Data Analyst

**Responsibilities:** Conducting thorough data analysis, identifying trends and patterns, generating insights, and preparing analytical reports.

* + 1. **Aravind Kumar Yaleti:**

**Role:** Visualization Specialist

**Responsibilities:** Creating interactive dashboards and visualizations in Power BI, ensuring data is presented in an accessible and insightful manner.