Thesis Project Portfolio

Test Data Generation for Large-Scale ETL Pipelines: Enhancing Data Integrity in Financial Services

(Technical Report)

Reframing Resistance and Engaging Community Perspectives in Urban Traffic Safety Planning

(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Arjun Trivedi

Fall, 2024

Department of Computer Science

Table of Contents

Sociotechnical Synthesis

Test Data Generation for Large-Scale ETL Pipelines: Enhancing Data Integrity in Financial Services

Reframing Resistance and Engaging Community Perspectives in Urban Traffic Safety Planning

Prospectus

Sociotechnical Synthesis

Introduction

The relationship between my technical work and STS research is rooted in the shared theme of data flow and its implications on decision-making processes. Although the contexts differ, with one focusing on financial services and the other on urban traffic safety, both projects highlight the importance of integrating technical efficiency with societal needs. My capstone project addresses the technical challenges of generating synthetic test data to enhance data integrity in ETL pipelines for financial services. In contrast, my STS research examines how community engagement can reshape urban traffic safety initiatives, exploring how data and stakeholder perspectives influence planning and policy decisions. By working on both projects simultaneously, I was able to draw parallels between the technical complexities of data processing and the human dynamics involved in decision-making, thereby enriching my understanding of socio-technical systems.

Summary of Capstone Project

In financial services, ensuring the accuracy and reliability of data processing systems poses a significant challenge when handling millions of daily customer transactions through Extract, Transform, Load (ETL) pipelines. To address this, I developed a synthetic test data generation system during my software engineering internship at JPMorgan Chase. The solution integrated web technologies, cloud services, and artificial intelligence to create and manage test data that closely simulated real-world scenarios. I then fed this synthetic data into the ETL pipeline at the extract stage, allowing for comprehensive testing of the entire process. Implementation involved leveraging cloud-based data cataloging and parallel processing techniques to enable efficient data generation. The system significantly reduced the time required for testing complex ETL pipelines, improving both the speed and accuracy of quality assurance processes. Future work could explore expanding the system's capabilities to handle a wider variety of financial data types and incorporating more advanced machine learning techniques.

Summary of STS Research Paper

Every 27 seconds, a life is lost to traffic accidents globally, underscoring the urgent need for innovative urban traffic safety measures. This research investigates how community resistance to such measures can be reframed as valuable input, enhancing both their technical efficacy and social acceptance. Utilizing Actor-Network Theory (ANT) complemented by the Multi-Level Perspective (MLP), the study explores the complex interactions between stakeholders in urban planning processes. The research question asks: How can community perspectives be effectively integrated into traffic safety initiatives? Through qualitative documentary analysis of academic literature, urban planning reports, and policy documents, this study expects to reveal that successful integration occurs through an interplay of regime-level changes, landscape pressures, and niche innovations. Preliminary findings suggest that cities adopting participatory approaches, such as Vision Zero policies, demonstrate improved safety outcomes and community satisfaction. This research contributes to STS by providing a nuanced understanding of urban planning's socio-technical nature. It offers practical insights for planners and policymakers on engaging communities in traffic safety measures, potentially leading to safer urban environments and a significant reduction in traffic-related fatalities.

Concluding Reflection

Working on both projects simultaneously provided a unique opportunity to bridge technical and societal perspectives, deepening my understanding of the interplay between data management and human decision-making. The technical project emphasized the importance of robust data systems, while the STS research illuminated the broader implications of how data-driven decisions are shaped by human and organizational factors. This dual approach underscored the necessity of integrating ethical considerations and stakeholder engagement into technical solutions. For instance, insights from my STS research on community engagement informed my understanding of how data scientists and engineers could design tools that better serve diverse stakeholders. Conversely, the technical project demonstrated the complexities of managing large-scale data, providing context for how processed data could be used in urban planning scenarios. Together, these projects emphasized that solving complex problems, whether in financial systems or urban planning, requires both technical expertise and a deep awareness of societal dynamics. This integrated perspective will continue to shape my approach to future challenges in engineering and society.