Real-Time Traffic Optimization for Taxi and Ride-Sharing for Smart Cities

Sai Bhargav K

Giridhar Sriram
Sriram J

Abhinav Varma
Varma V

Team 3

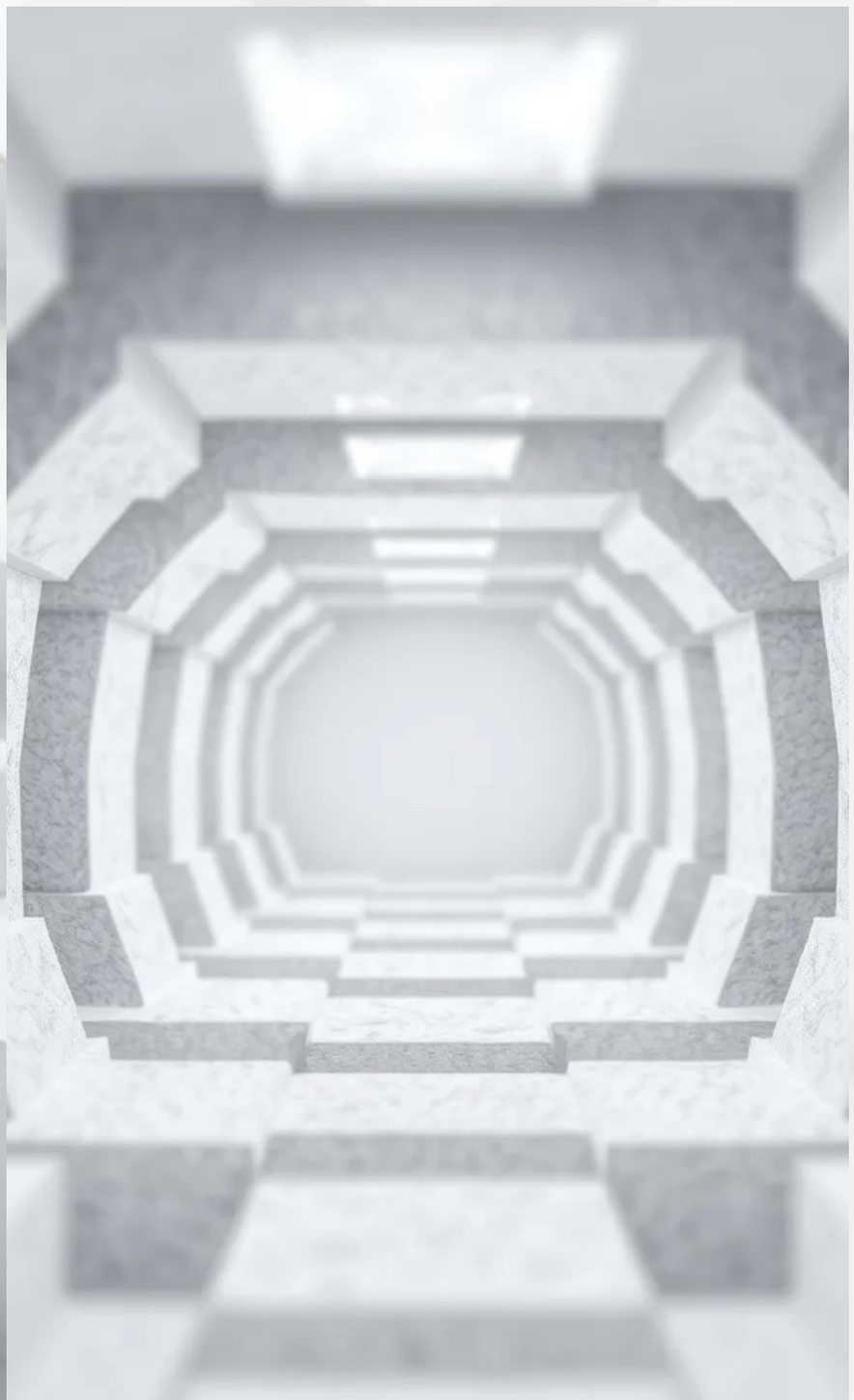
CONTENTS

1. Introduction

Why is this project important?

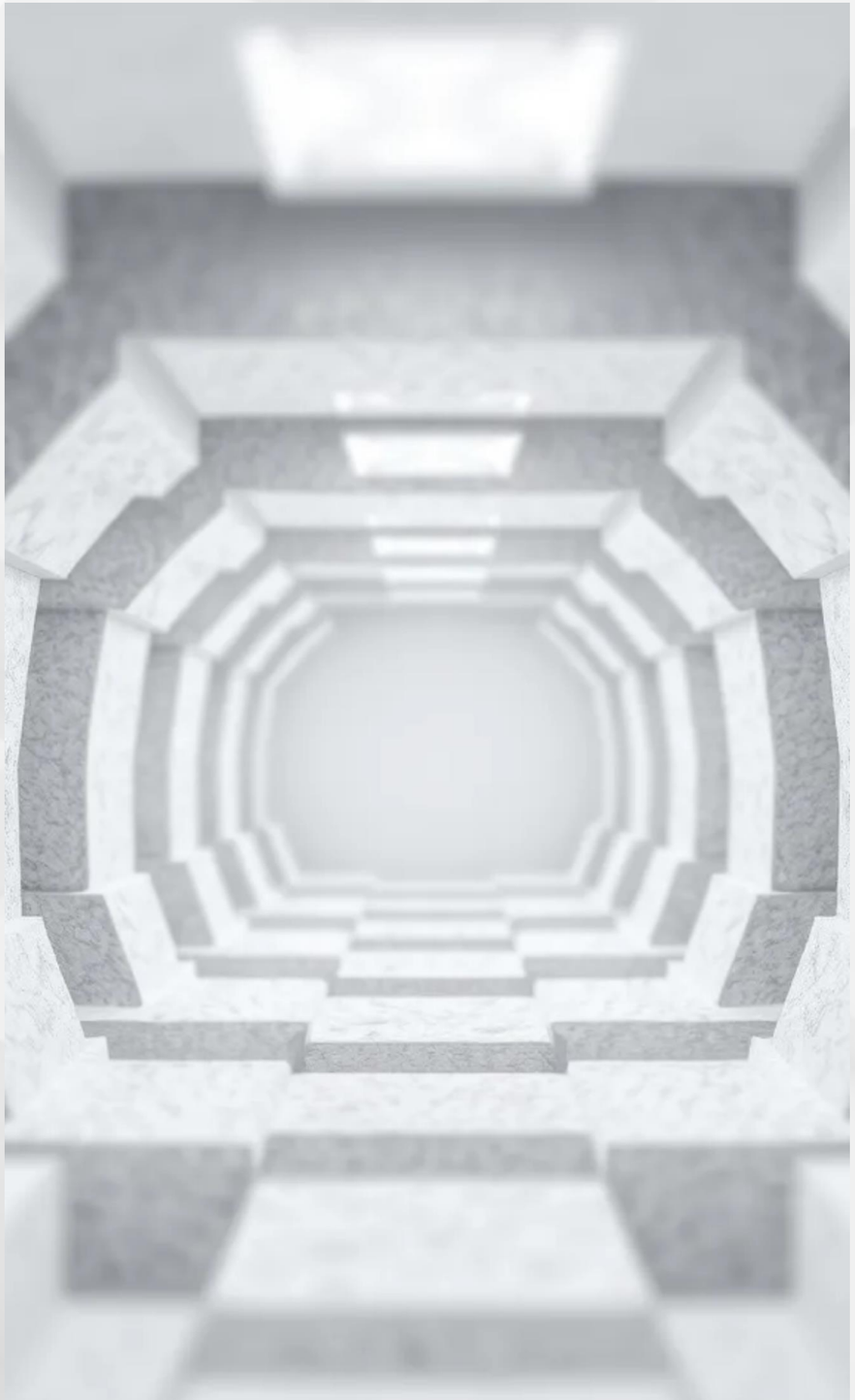
•Traffic congestion leads to longer travel times, more fuel use, and higher pollution.
•Our project analyzes real-time and historical traffic data to improve city traffic.
•It helps ride-hailing services like Uber, Lyft, and taxis optimize their routes.
•Uses AI and big data to provide better route suggestions and predict congestion.

Project Objectives



Objectives Overview

- Analyze real-time traffic patterns to identify congestion hotspots
- Optimize ride-hailing services by predicting demand and suggesting efficient driver allocation
- Predict future traffic congestion using historical data and machine learning models
- Provide real-time route optimization suggestions to reduce travel delays
- Develop a data visualization dashboard for city traffic authorities

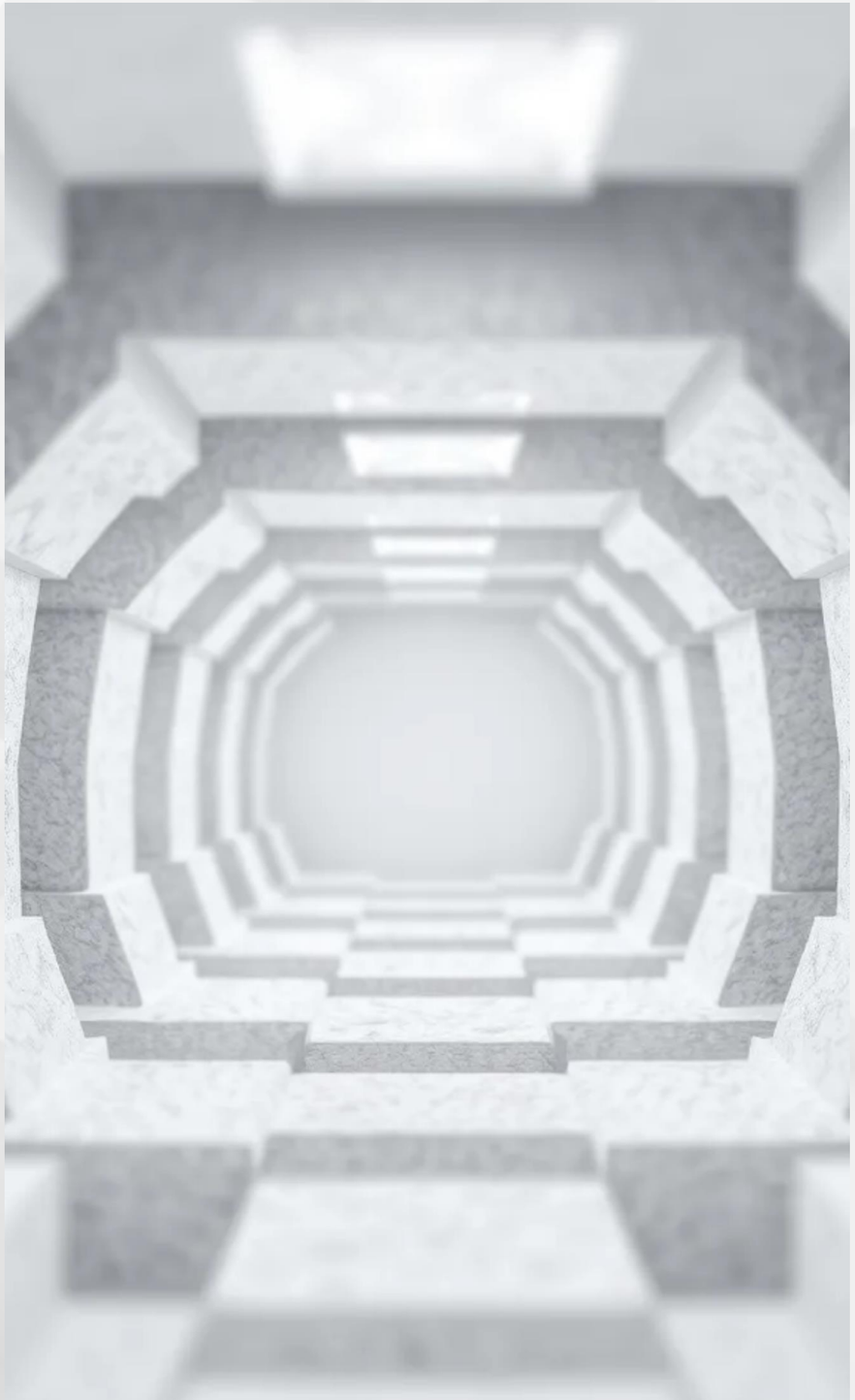


Key Problems Addressed

- How can we predict traffic congestion using real-time data?
- How can we reduce wait times for ride-hailing services?
- How do weather and peak hours impact traffic flow?
- How can big data handle large-scale traffic datasets efficiently?

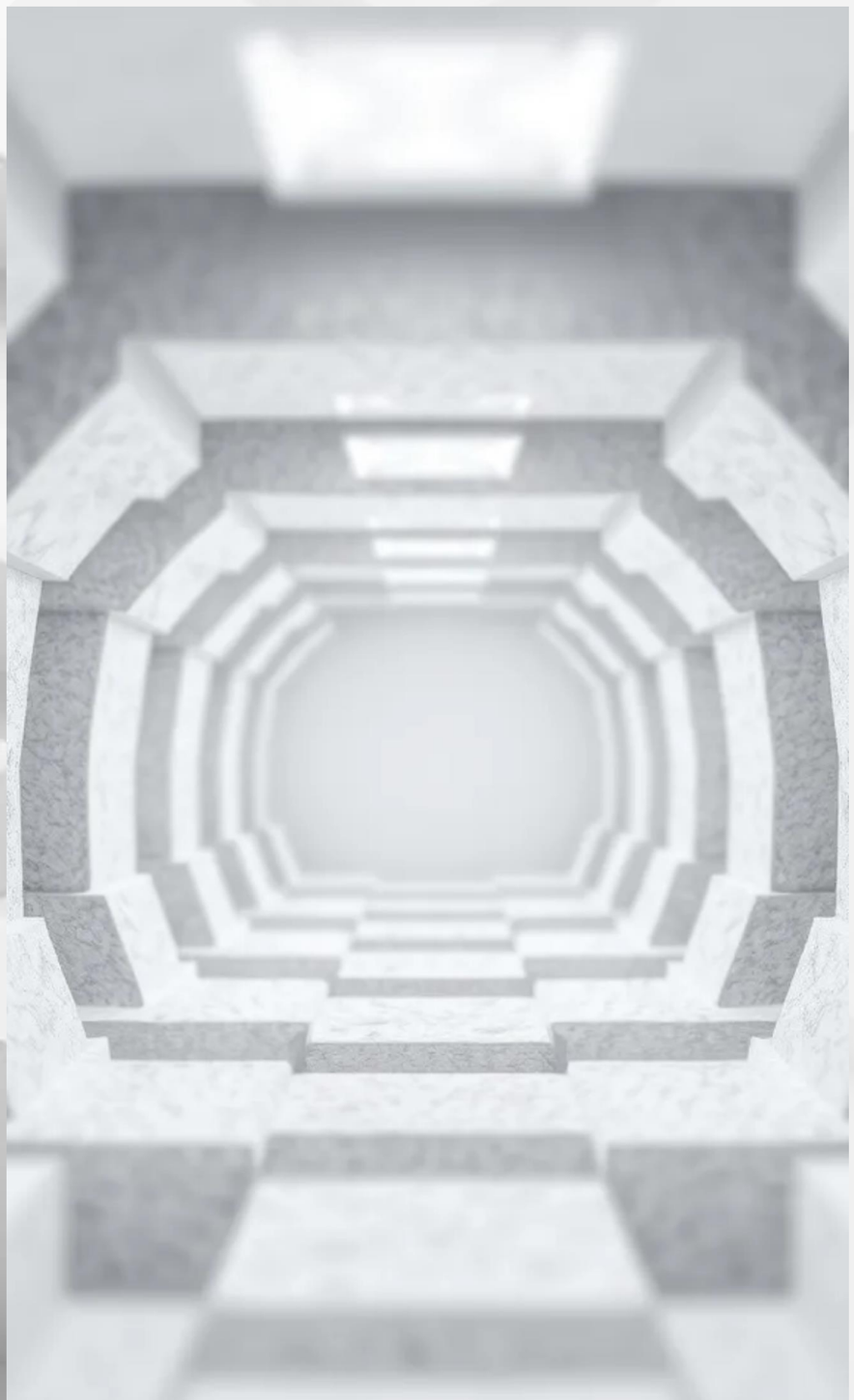
Data Sources and Collection

DATA
COLLECTION



Data Sources

1. **NYC Taxi & Uber Data:** Trip records, timestamps, pickup/drop-off locations. (TLC Dataset)
2. **Google Maps API:** Real-time traffic and congestion data.
3. **OpenStreetMap (OSM):** Road networks for better route planning. (OSM Data)
4. **NOAA Weather Data:** Weather impact on traffic flow. (Weather API)



Data Collection and Challenges

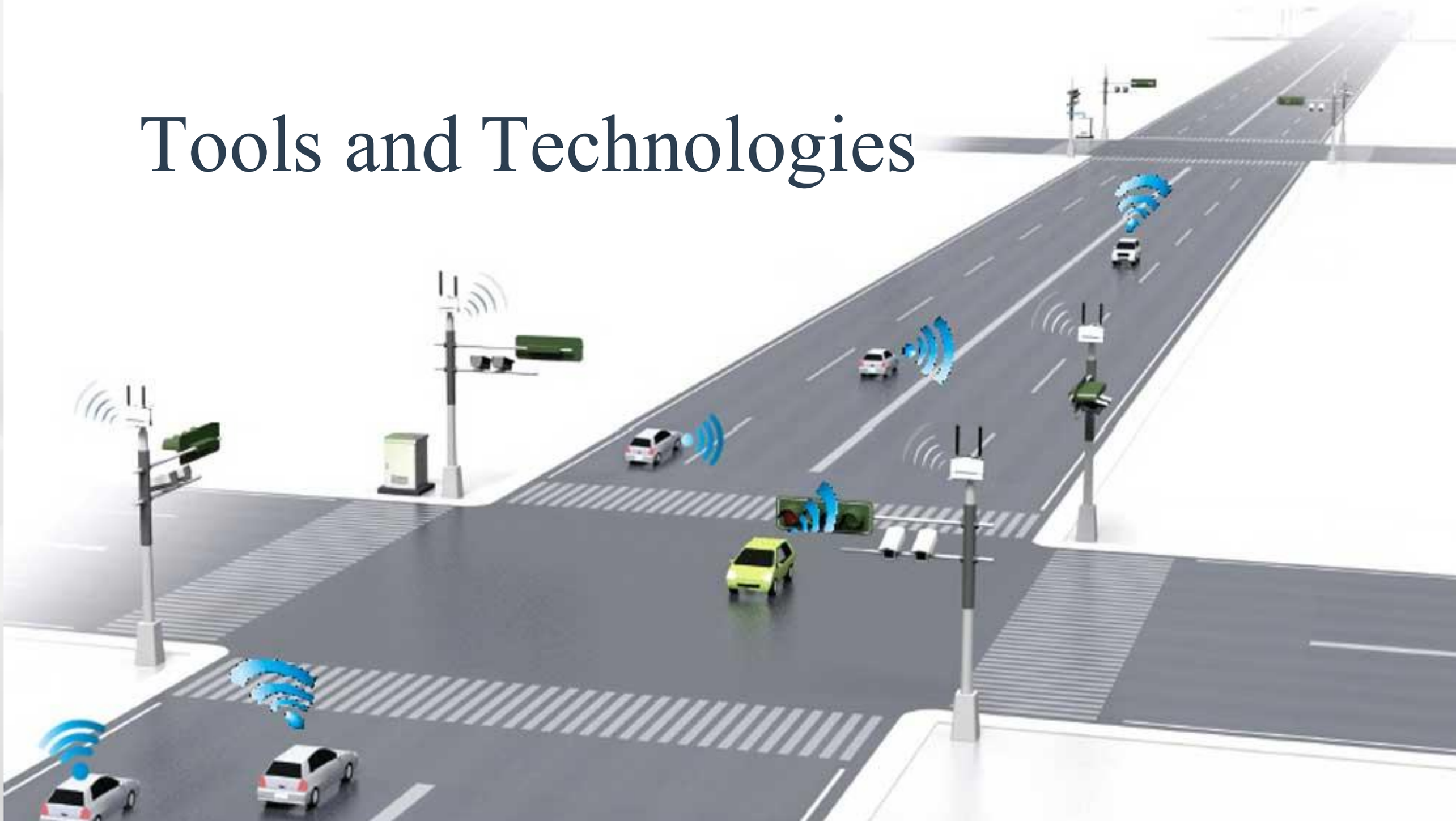
How We Collect the Data:

1. Real-time streaming via Kafka & Spark Streaming.
2. Historical storage in HDFS & AWS S3 for long-term analysis.
3. Geospatial processing using OpenStreetMap data.

Challenges in Data Collection:

1. Large data streams → Use Kafka & Spark for real-time processing.
2. Data privacy issues → Use anonymized datasets.
3. Different data formats → Standardization via Spark & Python.

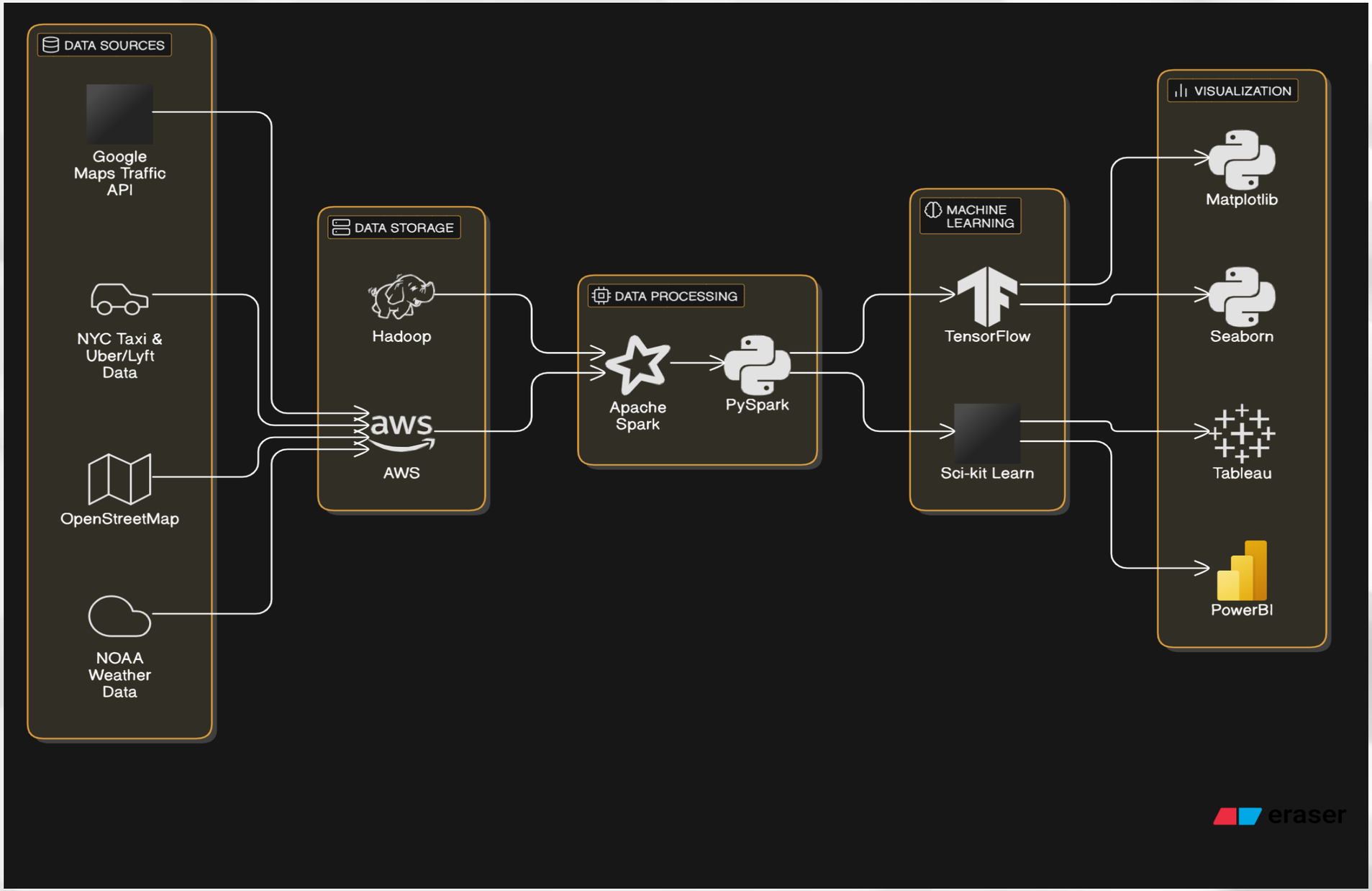
Tools and Technologies



Tools & Technologies

Component	Technology	Why We Use It?
Data Ingestion	Kafka, Apache NiFi	Streaming real-time traffic data
Data Storage	Hadoop HDFS, AWS S3, Google BigQuery	Handling large-scale traffic and ride data
Data Processing	Apache Spark, PySpark, Dask	Distributed processing of real-time and historical data
Machine Learning	Spark MLlib, Scikit-learn, TensorFlow	Predicting traffic congestion and optimizing routes
Visualization	Tableau, Power BI, Google Maps API	Creating insights with interactive dashboards

Data Stack Diagram



Timeline and Milestones

Weekly Milestones

Week	Milestone	Task Details
Week 1-3	Data Collection & Preprocessing	Set up APIs, collect real-time/historical data, clean & preprocess data
Week 4-6	Traffic Pattern Analysis	Use Spark MLlib to identify congestion zones and high-demand areas
Week 7-9	Predictive Modeling	Train machine learning models (LSTM, ARIMA) for traffic prediction
Week 10-12	Optimization & Dashboard	Develop real-time route recommendations and interactive dashboards

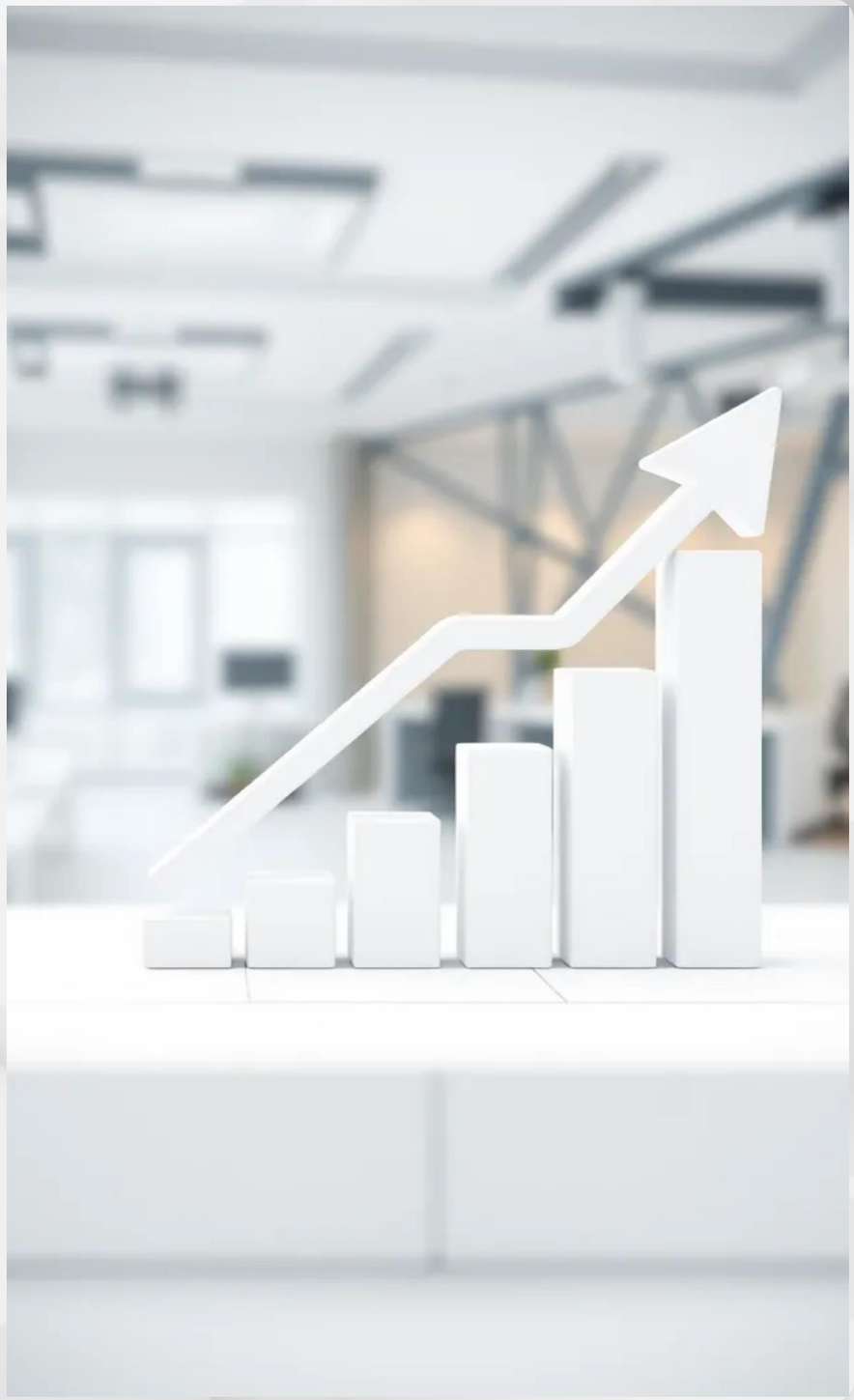
Task Allocation

Data Collection & Storage → Abhinav Varma V

Big Data Processing & Analysis (→ Giridhar Sriram J

Visualization & Dashboard → Sai Bhargav K

Project Documentation & Presentation → All Members




Thank You