# Professional Report: Event-Driven Data Processing for Taxi Trips

Author: Data Engineering Team

Date: January 2025

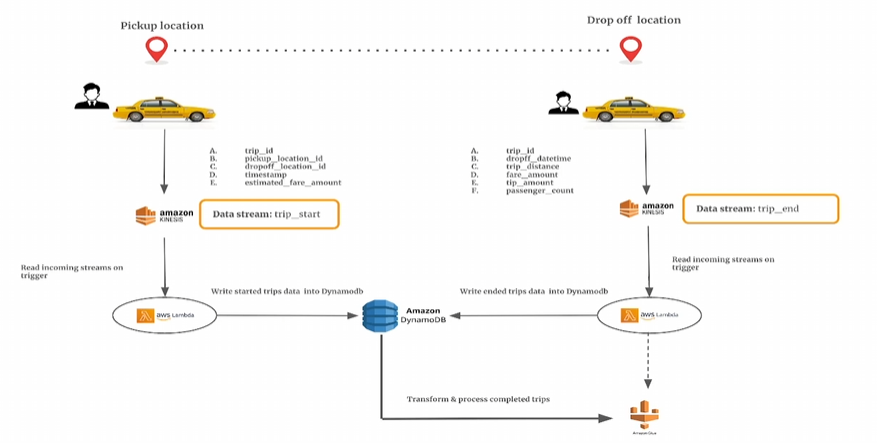
# Introduction

This report presents an event-driven data processing system for a taxi service company using AWS services such as Kinesis, Lambda, Glue, and DynamoDB. The system captures real-time trip data, processes it dynamically, and stores it efficiently for analytics and reporting.

# System Architecture

The architecture consists of the following key components:  
1. \*\*Amazon Kinesis Data Streams\*\*: Captures real-time trip data.  
2. \*\*AWS Lambda Functions\*\*: Processes trip start and end events.  
3. \*\*Amazon DynamoDB\*\*: Stores trip details.  
4. \*\*AWS Glue\*\*: Aggregates trip data for analytics.  
5. \*\*Amazon S3\*\*: Stores processed data for further use.

Below is the system architecture diagram:



# Data Flow and Processing

1. \*\*Trip Start Event:\*\*  
 - A customer initiates a trip.  
 - Trip details (trip ID, pickup location, fare estimate, etc.) are sent to Kinesis.  
 - A Lambda function listens to this data stream and writes it to DynamoDB.  
  
2. \*\*Trip End Event:\*\*  
 - When the trip is completed, additional details (drop-off location, fare amount, trip distance, etc.) are sent to Kinesis.  
 - A second Lambda function updates the same trip record in DynamoDB.  
  
3. \*\*Data Processing:\*\*  
 - AWS Glue processes completed trips.  
 - The processed data is stored in S3 for analytics.

# Deployment Process

1. \*\*Create Kinesis Data Streams\*\*  
 - Create two data streams: 'trip\_start' and 'trip\_end'.  
2. \*\*Develop and Deploy Lambda Functions\*\*  
 - Implement Lambda functions for processing trip start and end events.  
3. \*\*Set up DynamoDB Table\*\*  
 - Create a table to store trip details.  
4. \*\*Deploy AWS Glue Job\*\*  
 - Configure Glue for processing completed trips.

# Key Benefits

1. \*\*Scalability\*\*: The system can handle thousands of real-time events.  
2. \*\*Cost-Effectiveness\*\*: Serverless architecture optimizes cost.  
3. \*\*Real-Time Processing\*\*: Insights are available as soon as trips are completed.

# Conclusion

This event-driven system is an efficient solution for managing taxi trip data in real time. By leveraging AWS cloud services, the system provides seamless data ingestion, processing, and storage for improved analytics and business insights.