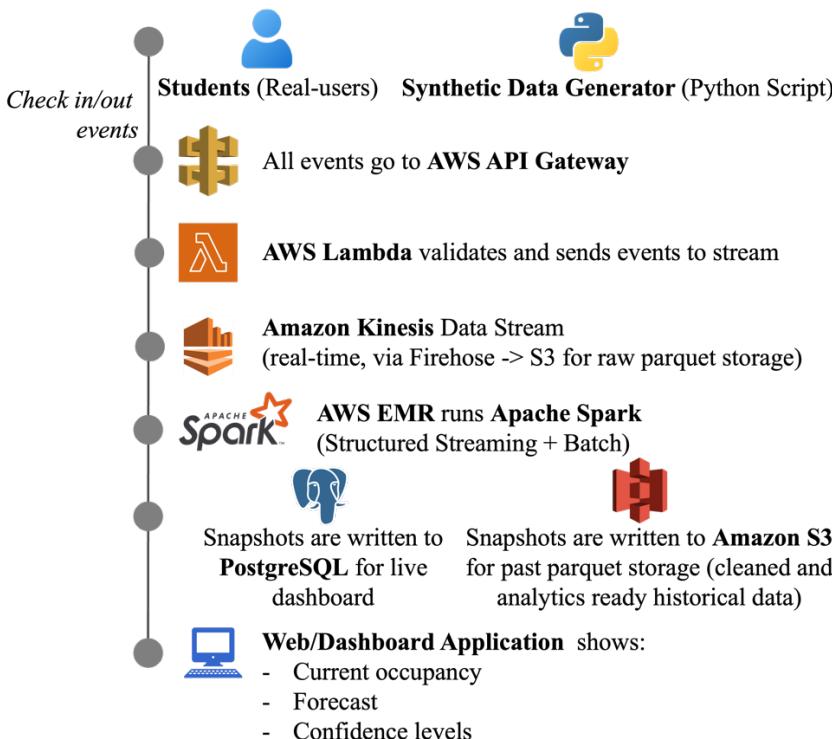


CMPT 732: Final Project Proposal

Student Name:	Diana Kozich, Manpreet Gill
Topic:	End-to-End Data Pipeline for Past, Present, and Future Parking Insights at SFU
Brief Description of a problem	Parking availability at SFU is often unpredictable, leading to frustration for students and staff. Currently there is no integrated system that analyzes historical trends, current occupancy, and future predictions. Without such insights drivers spend extra time searching for available spots, contributing to congestion.
Brief Description of the objectives	This project aims to address the problem by designing a big data workflow that can process and analyze parking data across different time dimensions: - Past (What are the peaks? How long do people stay? What drives spikes?) - Present (How many spots are occupied/free in each lot?) - Future (How many will be occupied in 15-30 mins?)
Technologies intended to use	Python, API Gateway, AWS Lambda, Amazon S3, Apache Spark, PostgreSQL, Amazon Kinesis, AWS EMR
Overview of the proposed workflow	 <pre> graph TD A[Students (Real-users)] -- "Check in/out events" --> B[AWS API Gateway] B --> C[AWS Lambda] C --> D[Amazon Kinesis Data Stream] D --> E["AWS EMR runs Apache Spark Structured Streaming + Batch"] E --> F[PostgreSQL for live dashboard] E --> G[Amazon S3 for past parquet storage cleaned and analytics ready historical data] F --> H[Web/Dashboard Application] G --> H </pre> <p>The diagram illustrates the data pipeline. It starts with 'Students (Real-users)' and 'Check in/out events'. These events go through 'AWS API Gateway' (represented by a yellow gear icon). From there, they pass through 'AWS Lambda' (represented by an orange lambda icon) which validates and sends events to 'Amazon Kinesis Data Stream' (represented by an orange cube icon). The stream then feeds into 'AWS EMR runs Apache Spark' (represented by a red spark icon), which performs 'Structured Streaming + Batch' processing. The output of Spark goes to two places: 'PostgreSQL for live dashboard' (represented by a blue monitor icon) and 'Amazon S3 for past parquet storage (cleaned and analytics ready historical data)' (represented by a red cube icon). Finally, a 'Web/Dashboard Application' (represented by a blue monitor icon) shows the results, including current occupancy, forecast, confidence levels, and historical trends.</p>