

## CMPT 732: Final Project Proposal

<b>Student Name:</b>	Diana Kozich, Manpreet Gill
<b>Topic:</b>	End-to-End Data Pipeline for Past, Present, and Future Parking Insights at SFU
<b>Brief Description of a problem</b>	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.
<b>Brief Description of the objectives</b>	<p>This project aims to address the problem by designing a big data workflow that can process and analyze parking data across different time dimensions:</p> <ul style="list-style-type: none"> <li>- Past (What are the peaks? How long do people stay? What drives spikes?)</li> <li>- Present (How many spots are occupied/free in each lot?)</li> <li>- Future (How many will be occupied in 15-30 mins?)</li> </ul>
<b>Technologies intended to use</b>	Python, API Gateway, AWS Lambda, Amazon S3, Apache Spark, PostgreSQL, Amazon Kinesis, AWS EMR
<b>Overview of the proposed workflow</b>	<pre> graph TD     A[Check in/out events] --&gt; B[Students Real-users / Synthetic Data Generator Python Script]     B --&gt; C[All events go to AWS API Gateway]     C --&gt; D[AWS Lambda validates and sends events to stream]     D --&gt; E[Amazon Kinesis Data Stream real-time, via Firehose -&gt; S3 for raw parquet storage]     E --&gt; F[AWS EMR runs Apache Spark Structured Streaming + Batch]     F --&gt; G[PostgreSQL for live dashboard]     F --&gt; H[Amazon S3 for past parquet storage cleaned and analytics ready historical data]     G --&gt; I[Web/Dashboard Application shows: Current occupancy, Forecast, Confidence levels, Historical trends]     H --&gt; I   </pre>