ROBERT GORDON UNIVERSITY ABERDEEN

SCHOOL OF COMPUTING SCIENCE AND DIGITAL MEDIA

MSc Big Data Analytics

| Student Name: ABERATHNE SIRIWARDANE THANUJA NISHADI	Matriculation Number: 2016997
Supervisor: MR. SAJITH RAVINDRA	Second Marker:
Project Title: REAL-TIME ROAD TRAFFIC MONITORING USING APACHE KAFKA AND SPARK	
	Start Date:01/03/2022
	Submission Date: 27/12/2022

CONSENT

I agree   ☒
I do not agree ☐

That the University shall be entitled to use any results, materials or other outcomes arising from my project work for the purposes of non-commercial teaching and research, including collaboration.

DECLARATION

I confirm:

- **That the work contained in this document has been composed solely by myself and that I have not made use of any unauthorised assistance.**

- **That the work has not been accepted in any previous application for a degree.**

- **All sources of information have been specifically acknowledged and all verbatim extracts are distinguished by quotation marks.**

Student Signature:	Date Signed: 27/12/2022

RGU SYMBOL OF SUCCESS
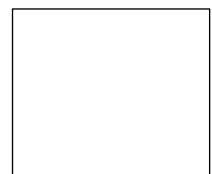
ABSTRACT

Traffic congestion is enormously increasing thus management of traffic flows especially in big cities are more challenging. Further, deploying and maintaining ground-based detection devices are prone to failures and expensive. In addition to that, many of the current devices are unable to access disaster or conflict areas furthermore they are not covering large geospatial areas at once. Therefore, one of the key enablers for having smooth traffic flows and better mobility is to rely on proper analysis of real-time traffic monitoring systems. The study will monitor and control traffic congestion by evaluating batch processing and streaming image data. Thus, the study covered two stages including Phase-I is for quantitative and qualitative methods of data gathering and analysis that targeting for road users and stakeholders, and Phase-II of the study covered the analysis of real-time image data using Apache Kafka, Spark Streaming, Spring Boot, and KSQL.

Analysis of real-time data is fairly challenging when compared to batch processing. In order to avoid dependencies among the components, it has implemented event-driven architecture including a multi-module maven project in IntelliJ Idea. Real-time sensor image data is ingested into the Kafka producer and then sent into the Kafka topics. Consumers are consumed the data from topics and stored them for transaction purposes. It has further, produced the aggregations of real-time traffic data through the KSQL interactive queries.

 Furthermore, it has verified the enables of reliability and scalability checking's of the KSQLDB cluster when 'sync' across restarts and scale-ups during the testing phase.

Keywords: stream processing, big data analytics, Kafka streaming, real-time road congestion, remote sensing.