

# REAL-TIME DATA ANALYTICS FRAMEWORK FOR CLOUD-BASED DIGITAL MAP

<sup>1</sup>Park, Daedong\*, <sup>2</sup>Kim, Jungho, <sup>1</sup>Kim, Myungsun, <sup>1</sup>Hong, Seongsoo

<sup>1</sup>Dept. of Electrical and Computer Engineering, Seoul National University, Rep. of Korea

<sup>2</sup>Grad. School of Convergence Science and Technology, Seoul National University, Rep. of Korea

**KEYWORDS** – cloud connected vehicle, cloud-based digital map, real-time data analytics framework

## **Research objectives** (max. 100 words)

A cloud-based digital map is emerging as a software platform for future connected cars. A cloud infrastructure for a high definition digital map must orchestrate vehicles, a cloud datacenter and connectivity front/back-ends in collecting, analyzing, synthesizing and transferring a huge amount of data. Since most automotive services are safety-critical, the software platform for future connected cars inherits stringent requirements such as timeliness and fault-tolerance. Unfortunately, existing data analytics frameworks for cloud datacenters are incapable of guaranteeing such requirements. In this paper, we propose stream processing architecture for a data analytics framework that guarantees the timeliness requirement for supporting cloud-based digital maps.

## **Methodology** (max. 150 words)

Digital map data contains diverse types of location-related information. Some of such information is dynamic in nature and needs to be continuously updated with raw data collected thru abundant vehicle sensors. Thus, stream processing is a mandatory feature of a map engine. Unfortunately, existing data analytics frameworks are not suitable for the real-time streaming processing of a map engine due to uncontrolled data dispatching latency and performance degradation. We propose three mechanisms for real-time stream processing: an anticipatory pipeline creation mechanism to eliminate the unexpected data dispatching latency, a cgroup-based performance isolation mechanism to avoid unexpected performance degradation caused by CPU and memory contention, a network bandwidth reservation mechanism to control performance degradation in datacenter network. We have implemented the three mechanisms in a private cloud datacenter composed of 25 servers and conducted experiments using OpenStreetMap to validate the timeliness requirement.

## **Results** (max. 150 words)

The three mechanisms are specifically designed for real-time stream processing in a data analytics framework. The anticipatory pipeline creation mechanism reserves a sufficient number of pipelines in advance so that raw data can be dispatched into an available pipeline as soon as it arrives. To do so, the mechanism continuously monitors the utilization of pipelines and checks whether it exceeds a threshold. If so, it creates additional pipelines. The cgroup-based performance isolation mechanism ensures each stream processing pipeline to use a predefined portion of CPU and memory regardless of other pipelines. The network bandwidth reservation mechanism enables each stream processing pipeline to negotiate with network switches for reserving bandwidth before the pipeline works. As a result, the network traffic generated by a pipeline is not delayed by other network traffic. We have conducted experiments and validated that our solutions eliminate the unexpected latencies in updating dynamic data.

## **Limitations of this study** (max. 50 words)

The current study has two limitations. First, the proposed mechanisms do not address other requirements of a cloud-based digital map such as security and reliability. Second, it does not consider the extremely complex networking environment of a huge cloud datacenter since experiments are performed in a small private cloud.

## **What does the paper offer that is new in the field including in comparison to other work by the authors?** (max. 50 words)

Current data analytics frameworks have not yet considered the timeliness requirement of a cloud-based digital map even though it is very critical in the automotive domain. This paper proposes new mechanisms for a data analytics framework which support stream processing so that it guarantees the timeliness requirement.

## **Conclusions** (max. 50 words)

A data analytics framework for a cloud-based digital map must satisfy a timeliness requirement. In this paper, we propose three mechanisms for a data analytics framework that is capable of guaranteeing the timeliness requirement in stream data processing.