

A distributed open source big data platform for real-time geolocation tracking

Pedro Moreno-Bernal¹, Carlos Alan Cervantes-Salazar¹, Sergio Nesmachnow², Juan Manuel Hurtado-Ramírez³ and José Alberto Hernández-Aguilar¹

Transportation geolocation problem

Mobility in high-density cities is complex for traffic, the different transportation modes, and its multiples origins and destinations[1]. Urban mobility needs to provide intelligent mobility services, for this reason, transportation systems require effective mobility techniques to reduce traffic incidents and congestion.

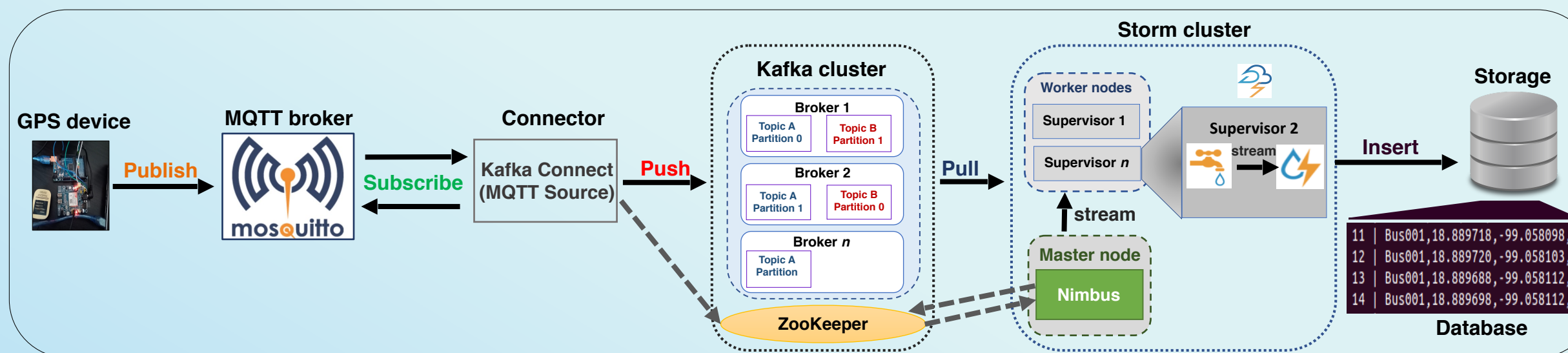


Intelligent Transportation Systems (ITS) is a mobility system that uses technological advances, methods, and applications of the Information and Communication Technologies (ICT) field, with high technology and mobility data[2].



Distributed open-source big data platform for real-time processing

ITS integrate Big Data technologies, computational intelligence techniques, and transport systems engineering to improve transportation mobility services. Big Data technologies integrate high technology and a large amount of data to develop new digital systems and applications in real-time to analyze, simulate, and process mobility data. The proposed open-source big data platform for real-time geolocation tracking is implemented using MQTT, Kafka, Storm, and MySQL.

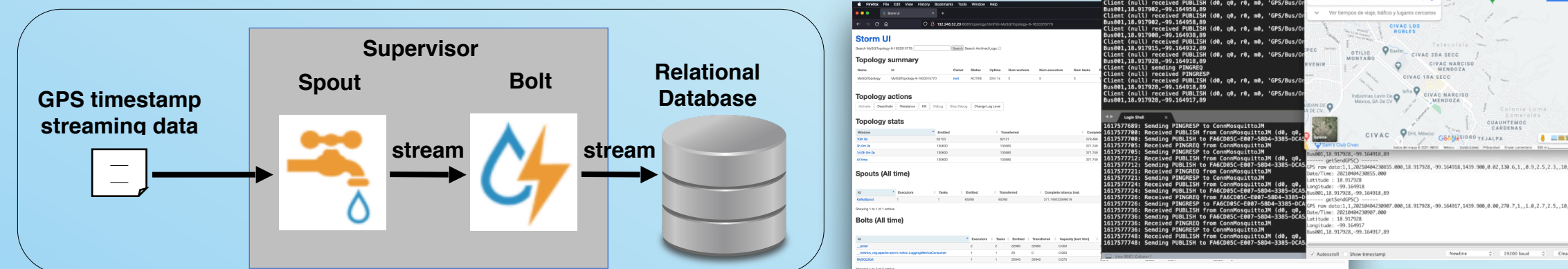


Goal: mobility data processing in real-time for geolocation tracking.

How: using a distributed open-source big data technologies such as MQTT, Kafka, Storm, and MySQL.

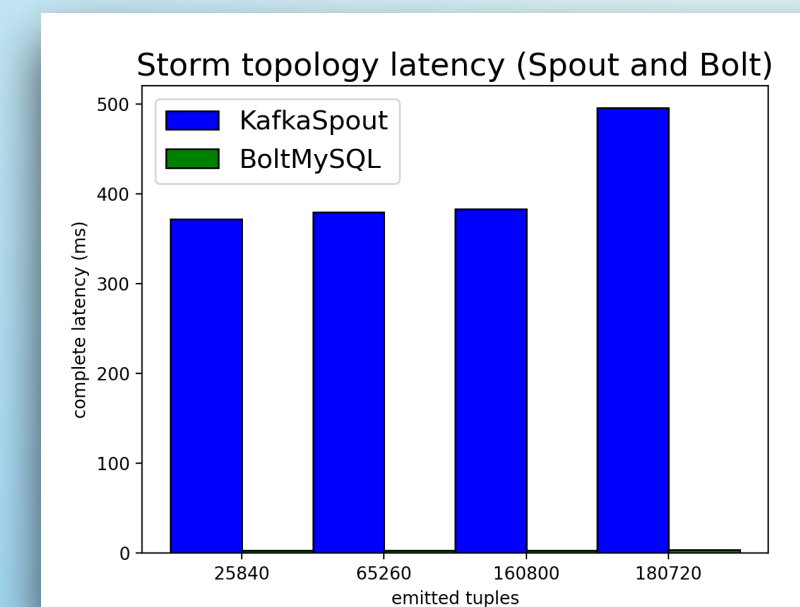
Our proposal: the proposed platform collects location information by an IoT device over MQTT protocol to Apache Kafka (publish-subscribe messaging), and then it is processed by Apache Storm, which guarantees fault tolerance, horizontal scalability, and low latency, and finally, information is stored on a MySQL database. The Spout/Bolt workflow in the Storm topology accepts tuples for processing from KafkaSpout to BoltMySQL, and then, data is stored on a relational database.

Experimental analysis: performed on GPS timestamps of 700 txt files from the T-Drive trajectories benchmark [3], up to one million timestamps in less than 57 minutes.



Results and conclusions

The distributed open-source big data platform for ITS in smart cities demonstrates the efficiency of the proposed solution in the 700 trajectories tested, processing up to one million GPS timestamps in less than 57 minutes. The Spout/Bolt workflow increments its latency (ms) as the number of tuples grows. The main lines for future work are oriented to extend the proposed system to compute in parallel topologies of Apache Storm.



References

- [1] Rodriguez, J.P.: The Geography of Transport Systems. Routledge, 5 edn. (2020).
- [2] Commission, E.: European commission mandate m/453 en (2009)
- [3] Zheng, Y.: T-drive trajectory data sample (2011)

Contact information

¹Universidad Autónoma del Estado de Morelos, Cuernavaca, Morelos, México, {[pmoreno](mailto:pmoreno@uaem.mx), [jose_hernandez](mailto:jose_hernandez@uaem.mx)}@uaem.mx

²Universidad de la República, Uruguay, sergion@fing.edu.uy

³Instituto de Biotecnología, Universidad Nacional Autónoma de México, jmhurt@ibt.unam.mx