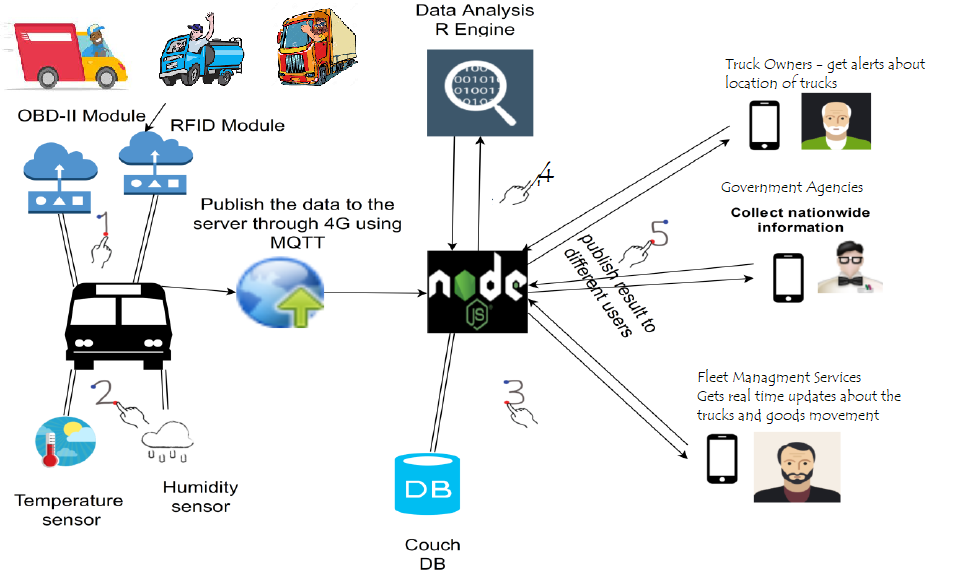
# STREAM PROCESSING and analytics Assignment 1 – Data Processing with Apache Kafka

Submission Date: ~~13 April 11.55 PM~~

Weightage: 15%

The number of accidents is shooting upward daily along with the rising safety concerns while travelling. A solution for most of these problems can be provided with an Intelligent Vehicle Monitoring System Using Global Positioning System along with Google Maps and Cloud Computing which collects useful information about a vehicle. There are also various sensors which relay information like fuel level, driver conditions and tire pressure. The vital information like the vehicle location, speed is gathered by the GPS which is fitted in the vehicle and transmitted in near-real-time to a centralized server maintained in the cloud network over MQTT protocol. This information is then available for the authorized users in real time and each licensed vehicle owner can access the data in cloud using a web portal anytime anywhere. This system thus provides an accurate positioning of the vehicle, speed, driver's condition and provides an intelligent monitoring of the vehicle remotely.



Your group is selected to prepare a working prototype of this IVMS using open source messaging platform Apache Kafka. A working prototype should mimic the following requirements -

1. Capturing the real time truck movement data from the sensors fitted in the trucks
2. Moving the running truck data over MQTT protocol to a centralized location
3. Moving data from centralized location to messaging store for intermittent storage (may put it in the persistent storage as well)
4. Preprocessing of the data received from the trucks for quality checks and for other required transformations
5. Doing the processing of data to identify the drivers exceeding the speed limits
6. Providing a mechanism to flag out the details of drivers exceeding the speed limits
7. Providing a way to maintain the count of over speeding incidents over the period of time, on particular routes, for particular trucks etc.

**You are supposed to carry out following tasks programmatically to help to roll out the solution.**

**Task 1: Architecture diagram for the whole solution**

**Task 2: Database schema and implementation for Truck driver data storage**

**Task 3: Simulator program for the truck data movement over the period of time**

**Task 4: Data Transfer program moving the data from the truck to central server like Mosquito broker through MQTT protocol**

**Task 5: Data transfer program from Mosquito broker to Kafka Topic and a raw data storage**

**Task 6: Data preprocessing / filtering program for identifying over speeding cases**

**Task 7: Program to keep statistics about over speeding cases over the period of time, for different routes, for different trucks etc.**

**Task 8: A simple interface for showing over speeding statistics to the end consumers**

References:

1. [Intelligent Vehicle Monitoring Using Global Positioning System and Cloud Computing](https://core.ac.uk/download/pdf/82497573.pdf)
2. [Real-Time Fleet Management Using Confluent Cloud and MongoDB](https://www.confluent.io/blog/fleet-management-gps-tracking-with-confluent-cloud-mongodb/)
3. [Track Transportation Assets in Real Time with Apache Kafka and Kafka Streams](https://www.confluent.io/blog/real-time-asset-tracking-software-with-kafka-elasticsearch-kibana/)
4. [IoT vehicle tracking demo](https://github.com/gschmutz/iot-vehicle-tracking-demo)

**Notes:**

* This is a take-home assignment to be carried out by each group of learners independently.
* This is programming exercise - requiring the generation of datasets to be used – on Mosquito + Kafka + Database environment using Python/Java language.
* You may consult / discuss with other learners peripheral aspects such as the environment but not on solving the specific problems in terms of design or implementation.
* You have to write the appropriate Python/Java code in to support you answers and submit with following nomenclature

Final document - SPA\_Assignment2\_<Group\_ID>.zip

* Prepare a demo video showcasing your solution working end to end and upload it on [**this**](https://drive.google.com/drive/folders/1JO7yTPFlGEzFT7LHLHa7VDGQABwA72Rf?usp=sharing) drive link following nomenclature

Demo\_SPA\_Assignment2\_<Group\_ID>.mpeg

* Provide appropriate justification when processing the data or arriving at the conclusions.
* In case of any further queries, if those are generic once, learners are encouraged to use discussion forums, otherwise they can reach out to me at [ppawar@wilp.bits-pilani.ac.in](mailto:ppawar@wilp.bits-pilani.ac.in).
* Manage your efforts properly as there is no scope to shift the deadlines announced above.