Introduction:

*General Bikeshare Feed Specification, known as **GBFS**, is the open data standard for bikeshare. GBFS will make real-time data feeds publicly available online in a uniform format so that map and transportation based apps can easily incorporate this data into their platforms*

***BIXI** Montréal is an organization created in 2014 by the city of Montreal to manage its bike-sharing system. The BIXI network has more than 6,000 bikes and 610 stations.BIXI is a hybrid between "**Bi**cycle" and "Ta**xi**" to underline the concept of being able to use a bicycle just like a taxi. **Bixi** is using General Bikeshare Feed Specification (**GBFS**).*

Project Overview:

One of the products is a trip planning application and the other one is a consultancy to Bixi. You are responsible to create a data warehouse for the company to achieve the goal.

You need to get an understanding of data feeds and implement appropriate batch or stream processing pipelines to load data in a data warehouse. This gives an opportunity to Analytics team and also the Reporting team to implement the best trip planning algorithm and report to senior managers to make strategic decisions.

Sprint 1: Data understanding and modelling

- *Analyzed the data source in order to better understand the quality and the content of the data.*
- *Created UML diagrams to show the relationship between files.*
- *Extracted a feed of data (JSON files) and loaded it into staging tables on Hive. Implemented a Scala program that automates this process.*

***Sprint 2:** Batch Processing Pipeline, **Skills:** Hive, Hadoop, Scala.*

- *Implemented a program to run ETL for transforming JSON files to CSV.*
- *Enriching transformed files and loading them in Hive tables.*

***Sprint 3**: Stream Processing, **Skills:** Hive, Hadoop, Kafka, Spark, Tableau.*

- *Implemented a Scala application to create Kafka topics and query Hive tables.*
- *Implemented a Spark streaming application that streams the "Enriched data" topic.*
- *Created Dashboards in Tableau connecting to Hive to perform data analytics.*