**DATA TEST**

**Duration:** Half-day.

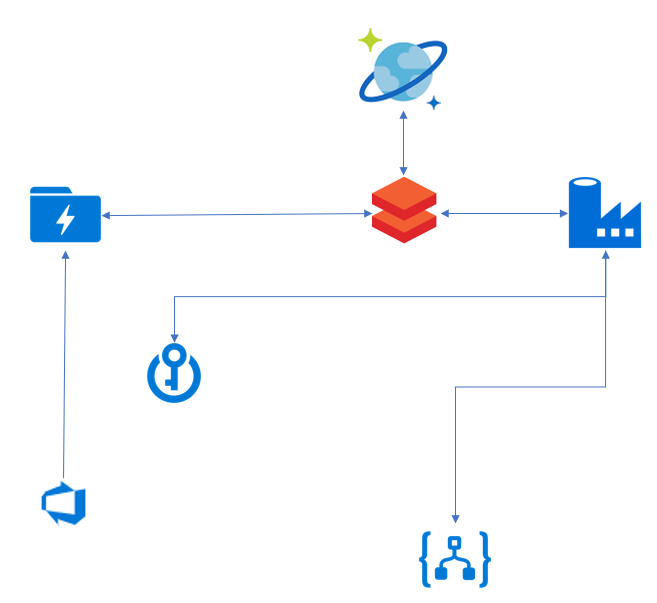
**Objective:** Test the candidate ability to deploy a project to production and run it, test his DevOps and Azure skills. No data science skills would be asses.

**Data:** Static geographical information of CityBike‘s stations in Brisbane (“*Brisbane\_CityBike.json*”, in raw-input)

**Code:** Scala Spark Job to build (with sbt clean compile), to write in CosmosDB some parameters should be set up (-Dendpoint="https://cosmos-example" -Dmasterkey="key1" -DsecondKey="key2")

**Instructions:**

The provided code performs a clustering based on either the location or characteristics of bike stations. This code need to be deployed in production and run Daily. In order to do that the whole Azure Architecture should be deployed (the logic app is a nice to have to send notifications):



It should be developed as close as a real industrialized process even if the same dummy sample data would be run every day in this test.

**Deliveries:** The delivery of this test should contain:

* Azure Platform: Provide a User with login and password (or token) to connect to your platform (you can create a free trial Azure [here](https://azure.microsoft.com/fr-fr/free/)).
* Documentation: A short description of the operationalization, a short description of how the dev evolution should be added without impacting the run.
* Backlog: The User Stories & Tasks (done or to do) related to deployment to production of this code.