Guida per la riproducibilità del progetto

Per riprodurre interamente il nostro progetto si faccia riferimento alle seguenti direttive:

1. **Avviare un server locale di Kafka dalla command line:**

$ zookeeper -server -start /usr/local/etc/kafka/server.properties

1. **Avviare l’applicativo Kafka Zookeeper, il Producer e il Consumer**

$ kafka -topics --create --zookeeper localhost:2181 --replication -factor 1 --partitions 1 --topic “nome del topic che voglio creare”

$ kafka -console -producer --broker -list localhost:9092 --topic “nome del topic appena creato”

$ kafka -console -consumer --bootstrap -server localhost:9092 --topic “nome del topic creato” --from -beginning

1. **Creare in MongoDB un database suddiviso in due Collections. Apriamo una nuova scheda della command line e digitiamo i seguenti comandi:**

$ mongo #entro nella shell di mongoDB

$ use dataman #creo il database chiamato “dataman”

$ db.createCollection(‘covid) #creo la Collection che raccoglierà i dati del contagio

$ db.createCollection(‘alphavantage’) #creo la Collection che raccoglierà i dati finanziari

1. **Aprire ed eseguire i Notebook “Acquisizione Dati e Storage”, “Gold\_Silver\_Joint” , “Oil price” e “Coropleto\_dati” e i relativi notebook per eseguire il pre-processing e lo storage dei dati. I notebook possono essere lasciati in esecuzione per tutto il tempo necessario a raccogliere la mole desiderata di dati. A fine di ogni ciclo di raccolta i dati verranno inseriti in MongoDB**
2. **Importare i Dati raccolti i Tableau Desktop.**