# Big data project (Airbnb Prices in European Cities)

1. **Architecture**
   1. **Data ingestion:**

Data is first ingested into the system from CSV files from Kaggle.

* 1. **Data storage:**

Data is stored in (HDFS), which provides distributed storage and processing of large datasets.

* 1. **Batch processing:**

Batch processing can beperformed using Apache Spark to process the daily batch of Airbnb data stored in HDFS and calculate metrics such as prices, cities, room types...

* 1. **Real-time processing:**

Real-time processing of incoming data can be performed using Apache Spark Streaming and Apache Kafka.

* 1. **Dashboarding and reporting:**

Finally, data is presented to end-users through interactive dashboards (web application).