

# Kafka-based Data Extraction for flights information

---

## Introduction

This document outlines the configuration and execution steps for the data extraction pipeline which scrapes flights data from airport websites and uses Apache Kafka for data processing.

## Pre-requisites

- Knowledge of Python programming and pip.
- Understanding of Apache Kafka concepts.
- Python 3.8+, Apache Kafka, `confluent-kafka` library installed.

## System Requirements

- Apache Kafka 2.x
- Python 3.8 or above
- Confluent Kafka Python client installed (`pip install confluent-kafka`)

## Installation and Configuration

### Kafka Setup

#### 1. Zookeeper :

```
./bin/zookeeper-server-start.bat config/zookeeper.properties
```

#### 2. Kafka Broker :

```
./bin/kafka-server-start.bat config/server.properties
```

## Running the Scripts

### Running the Scripts

Execute the producer scripts to scrape data and send it to Kafka:

```
python weather_producer.py
python flights_producer.py
python reviews_producer.py
```

### Starting the Consumer Scripts

Run the consumer scripts to consume data from Kafka and save it to CSV files:

```
python weather_consumer.py  
python flights_consumer.py  
python reviews_consumer.py
```

## Monitoring and Maintenance

Monitor Kafka broker health and script execution regularly. Verify CSV outputs for correctness.