

# Data Publisher Framework

## Contents

Overview .....	2
Design Consideration.....	2
Requirements.....	2
Technologies .....	3

## Overview

There is a data publisher which publishes data on demand, there should be multiple consumers which should be able to consume data simultaneously. Consumers will pass Date range and Data format as input parameters. There should be JWT based / any alternate authentication between consumer and publisher.



## Design Consideration

- Only authorized consumers should be able to access data
- There should be data completeness checks using checksum – Nice to have
- The expected data volume published by publisher will be around 1MB (~100K records).
- The system should support and demo 3 consumers consuming data in parallel.
- The complete execution needs to happen within 120 seconds i.e. starting from consumer requesting data from publisher to data download at consumer end.
- The system needs to be scalable to adopt different file formats like xml, json, AVRO etc. from publishers
- Fault tolerant – The system should behave in adverse scenarios. Data Preparation & Ingestion

## Requirements

- The publisher REST service should always be ready to consume requests from consumer.
- All publisher service should first authenticate consumers based on username / password / key / JWT token
- **Story1:**
  - The REST end point should be able to take input request parameters - **Order Date Range** and **Output Format** requested (like CSV, AVRO, JSON, XML). In response to data request publisher should acknowledge the request and return a job ID
  - The publisher should start a job internally which should read Order data from database and generate consumer expected file format.
- **Story 2:**
  - The consumer should be able to check the Job Status using the JOB ID
  - Maintain different Job Status – In Progress, Aborted, Completed

- **Story 3:**
  - When JOB completes the consumer will call final service to download data using same JOB ID
  - Consumer should be able to download data in zipped format containing file in requested format.
- **Story 4**
  - There should be REST end point configured to provide status of all the consumer jobs for particular date.

## Technologies

- JDK 1.8
- Spring Boot
- MySQL / Mongo / Oracle
- Core Java
- Swagger/Postman – for demonstrating
- Git commit
- 80% Code Coverage
- Sonar cube analysis