Atlan Backend Task

**Documentation**

## Problem Statement

**Atlan Collect** has a variety of long-running tasks that require time and resources on the servers. As it stands now, once we have triggered off a long-running task, there is no way to tap into it and pause/stop/terminate the task, upon realizing that an erroneous request went through from one of the clients (mostly web or pipeline).

### Tech stack used

* SpringBoot framework for creating the API endpoints
* Java 8.
* H2 in Memory Database for storing CSV file data & status of jobs.

### Dummy Data Structure

Dummy data CSVs contain the following attributes (created using Faker):

* Name
* Email
* Favorite Color
* Phone number
* Country

### Approach

In this to make assure that only one job fulfill at one time I use the concept of Synchronization locking technique. If one Job already in progress than JSON form message send back to user that one job is already running. For resuming the process I remember by storing in database that how many rows already read and stored in database. So, next time when user resume the job than I skip already read records by looking in database that how many lines in CSV file are already read.

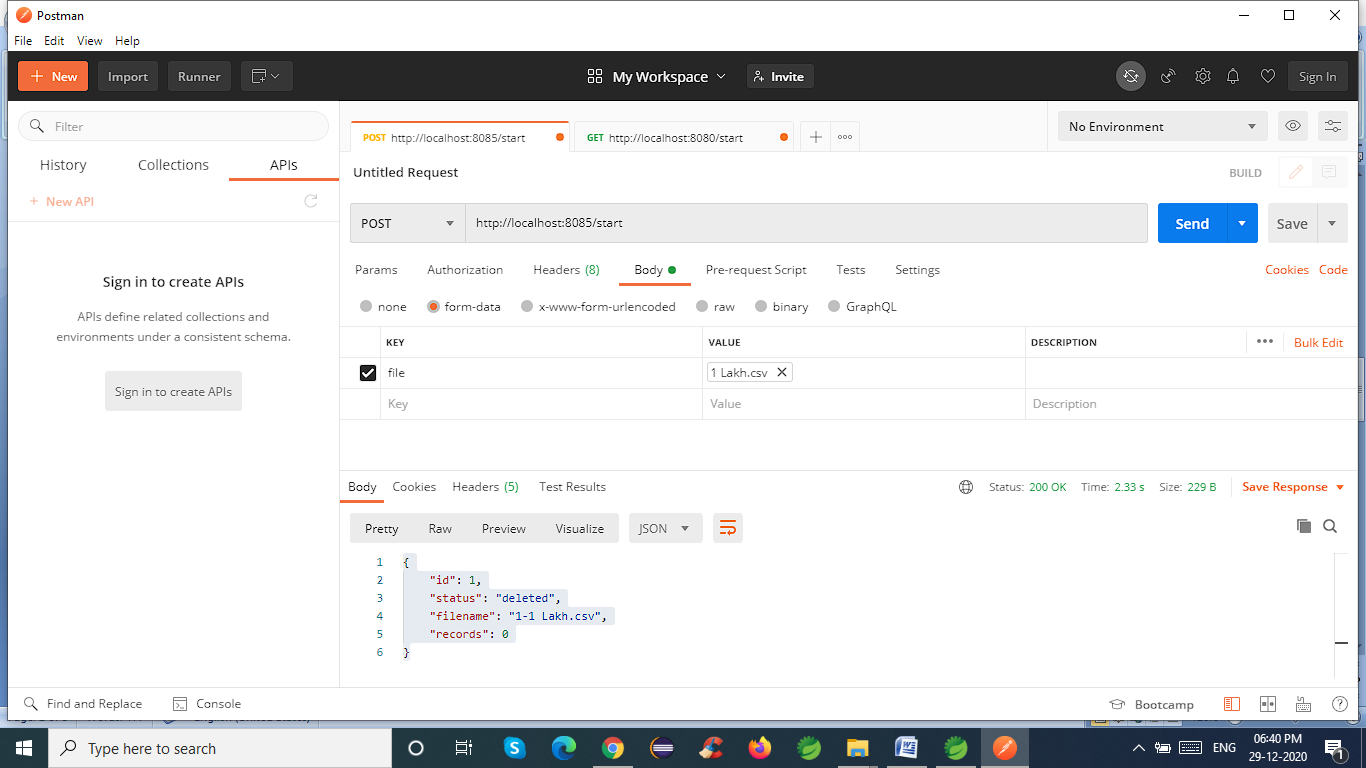
* **Processing**: In progress.
* **stopped**: Temporarily stopped via user query, but can be resumed.
* **terminated**: Permanently stopped via user query, and cannot be resumed.
* **Job Completed**: 100% progress.

ENDPOINTS

1. **EndPoint (For start new Job)-** {base\_URL}/start

**Type-** Post

**Note:** In this request you have to attach CSV file(Should have Six columns) in Form Body in Postman as key parameter name as file as below.



# Response- {

# "id": 1,

# "status": "processing",

# "filename": "1-1 Lakh.csv",

# "records": 0

# }

# 2. EndPoint (To pause currently processing job)- {base\_URL}/stop

**Type-** Get

# Response- {

# "id": 1,

# "status": "paused",

# "filename": "1-1 Lakh.csv",

# "records": 86535

# }

**3. EndPoint(For Resume Job)-** {base\_URL}/start/{id}

**Type-** Get

**Response-** {

"id": 1,

"status": "processing",

"filename": "1-1 Lakh.csv",

"records": 86535

}

4. EndPoint(For Get Data saved By job)- {base\_URL}/data/{id}

Type- Get

Response- All data In Jason Form [{},{},…]

5. EndPoint(For Terminate Paused or Processing Job)-

{base\_URL}/terminate/{id}

Response- {

   "id": 1,

  "status": "terminated",

   "filename": "1-1 Lakh.csv",

   "records": 0

}

6.Endpoint(For delete all data of completed Job)-

{base\_URL}/delete/{id}

Response- {

   "id": 1,

  "status": "deleted",

   "filename": "1-1 Lakh.csv",

   "records": 0

}

7. Endpoint(For check Status Of Job)- {base\_URL}/status/{id}

Response- {

"id": 1,

 "status": "processing" or "Paused" or "terminated",

  "filename": "1-1 Lakh.csv",

 "records": 0

}