Assignment: Ingest json to retrieve results as per need.

Purpose:

The purpose of this utility is to extract information from a larger json file and to ingest into storage/querying system to acquire results as required. This utility will export the result into a CSV file.

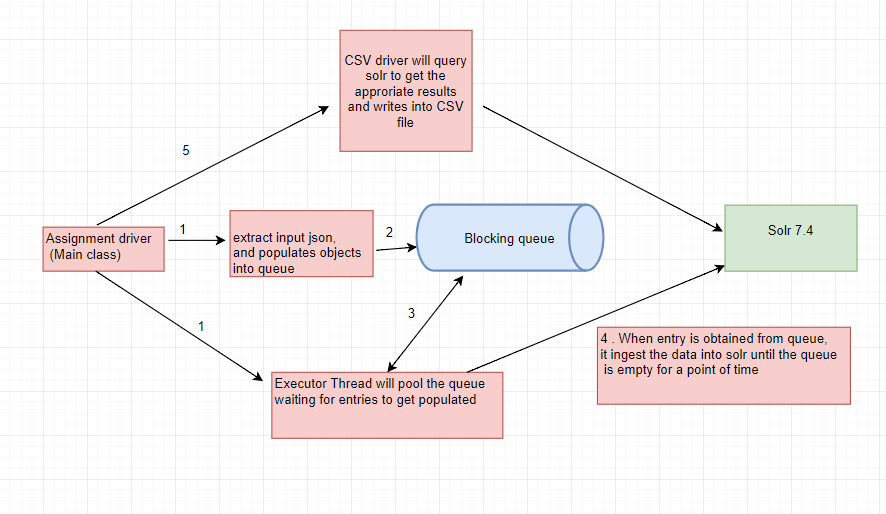
Technology used:

* Java 1.8
* Maven 3.2
* Log4j2
* Solr 7.4

Other dependencies used:

* Open-CSV reader
* GSON
* JACKSON-json
* Apache-commons libraries

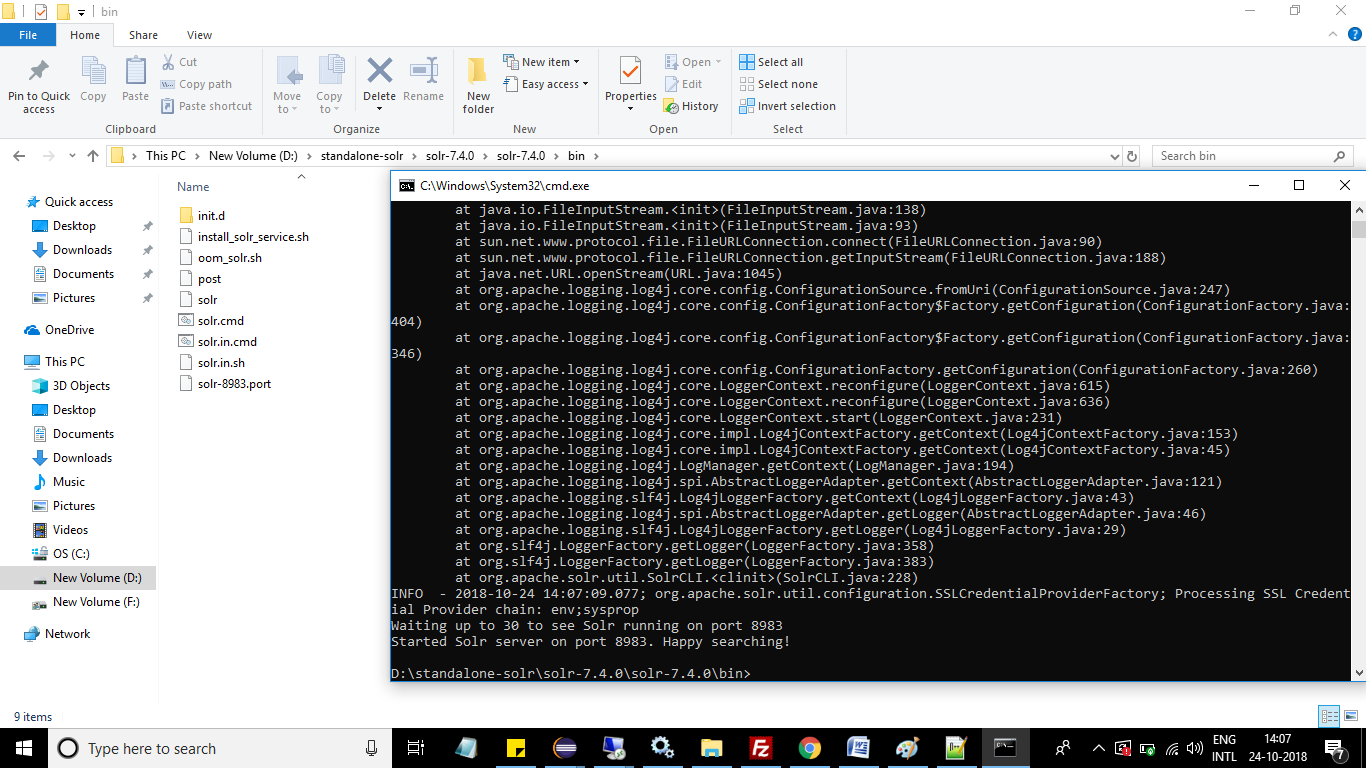
Architecture:



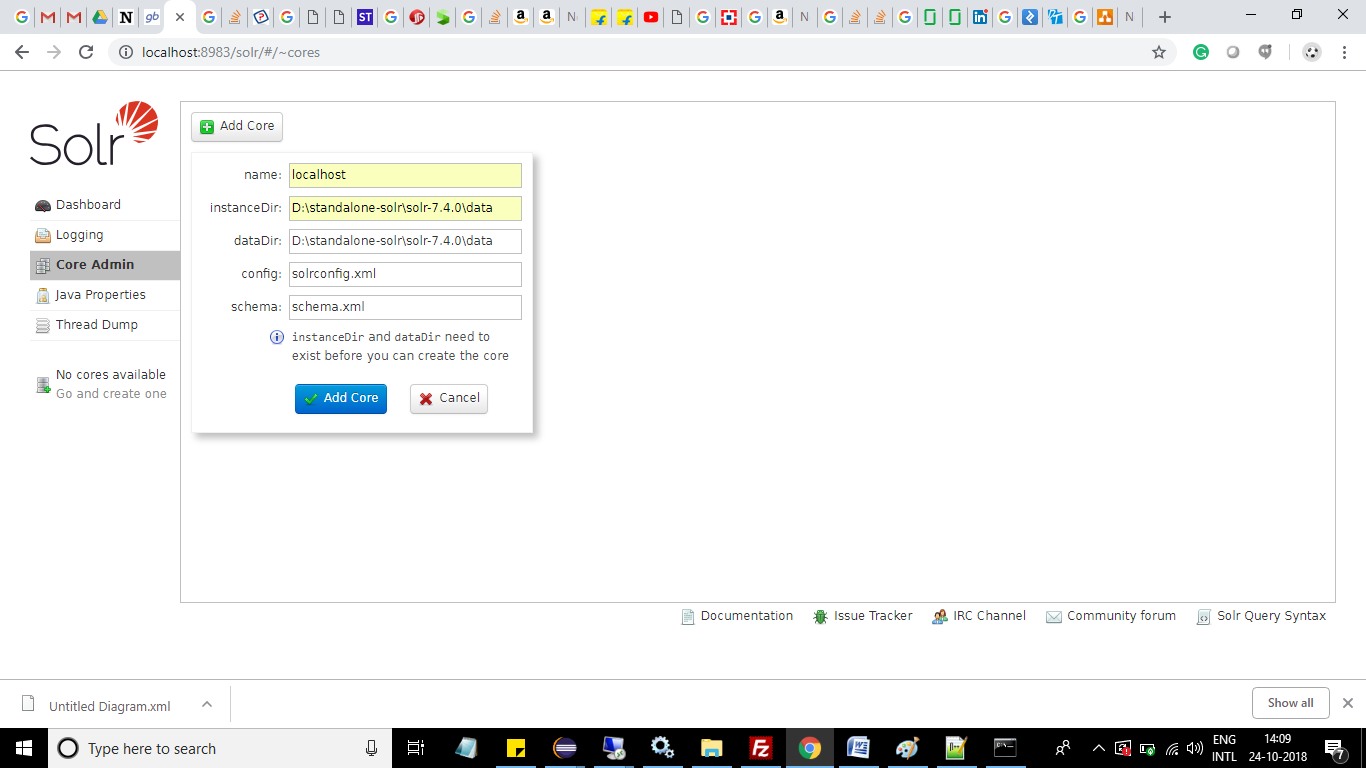
#### Process:

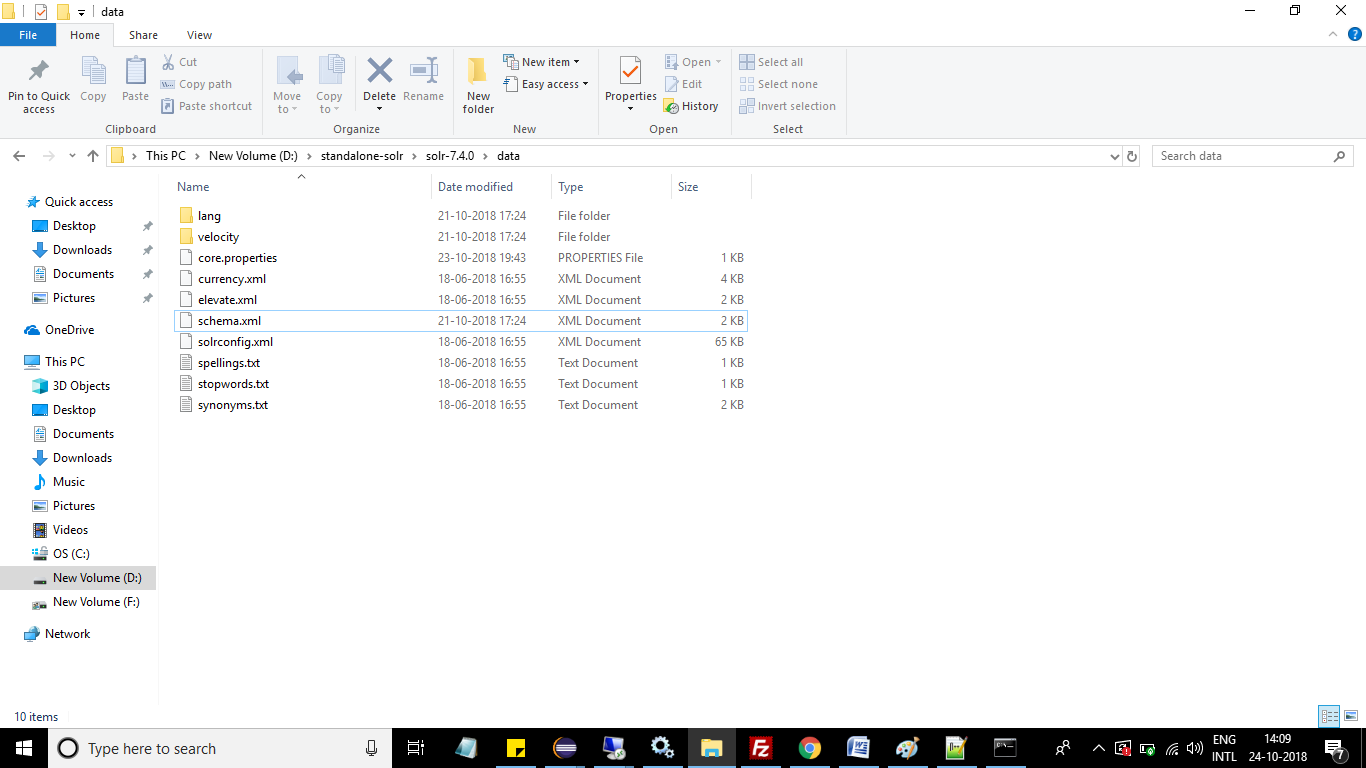
###### Pre-requisites:

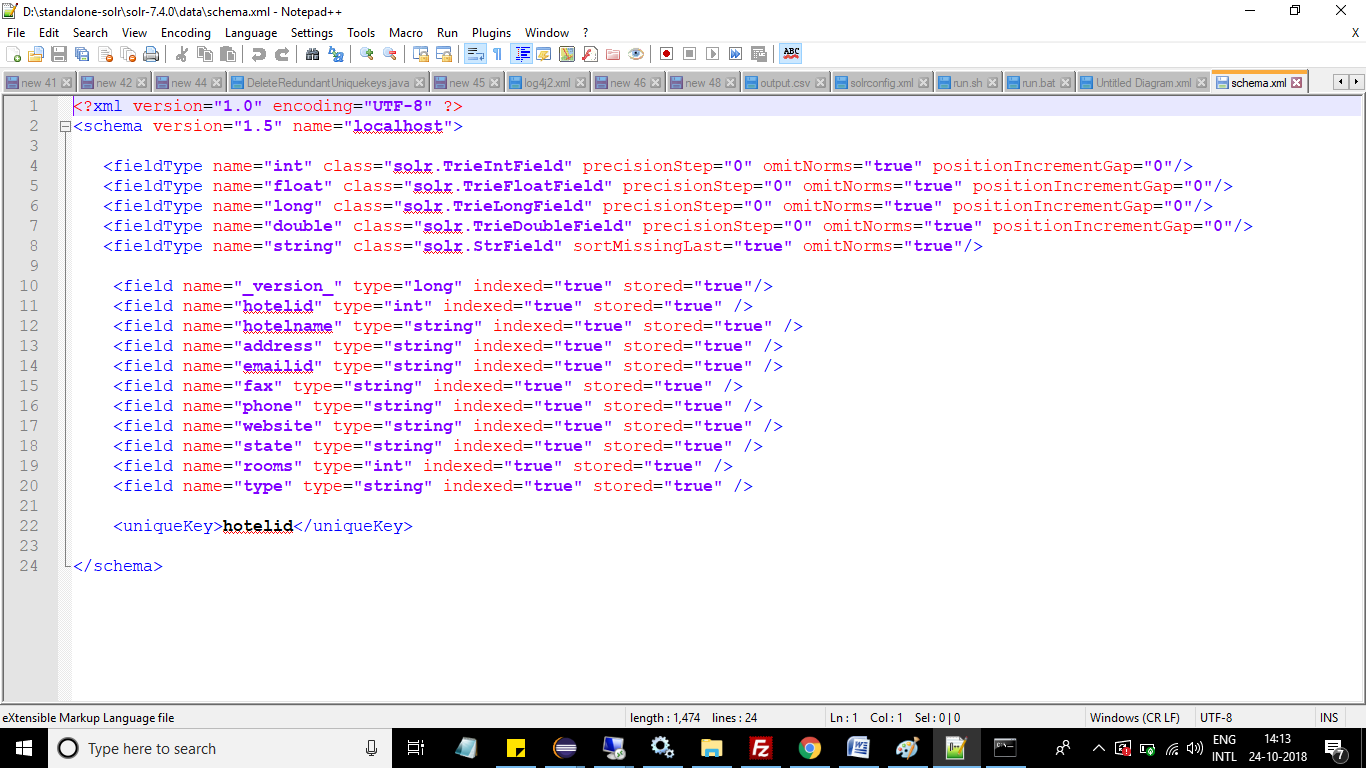
* Download Solr 7.4 version from solr website. Start solr server from bin from cmd as solr start.



* Open browser and hit the url <http://localhost:8983/solr>
* Create a empty directory as data in local file system and copy solr’s default config files that are in examples folder in solr directory which ships along.
* Edit the schema.xml file as per our need
* Click on the add core button to create core.







Ingestion:

* The utility will read the input json file as stream using a feature present in Google’s GSON to read JSON objects and arrays as streams. So this feature enabled in this utility proves it can handle larger JSON files with minimal memory.
* These streams are converted into POJO objects and are pushed inside a queue.
* A thread that will be invocated in the startup , will poll the queue for objects to be available.
* This thread will convert the objects into SolrInputDocuments and index into solr.

Querying & Result export:

* The utility will query the solr server to obtain the type of rooms and will obtain the hotel details based on the highest number of rooms present in the respective type.
* This query response is exported into the output.csv file in the configured output directory.

Sample output: 