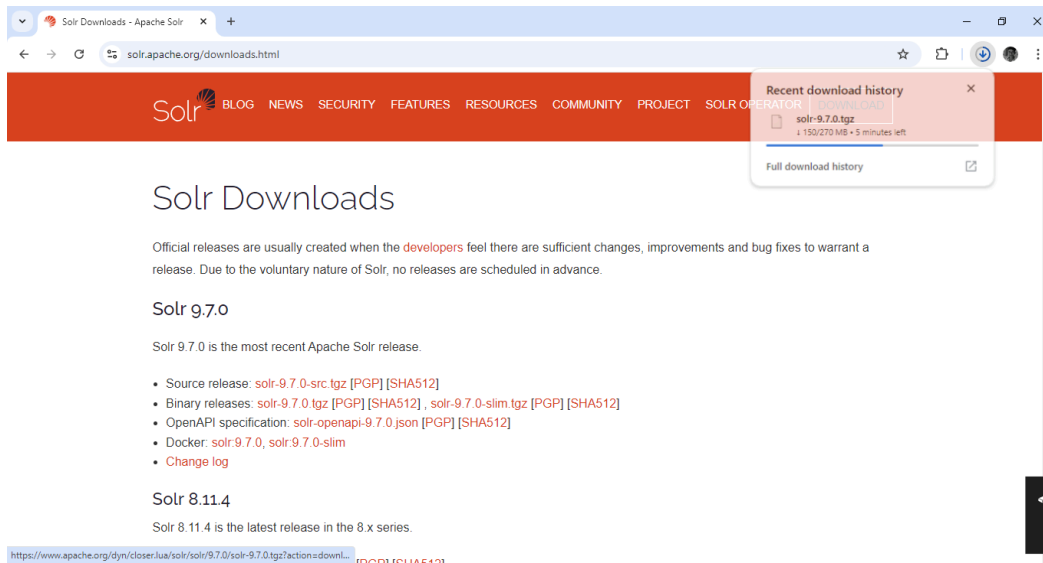


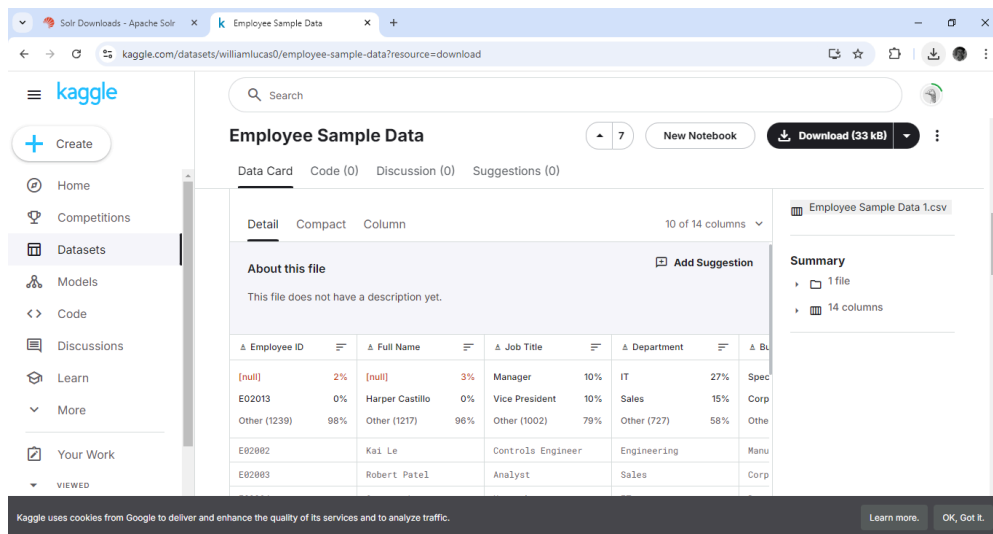
Tasks:

1. Read the Apache Solr.
2. Install Solr on your local machine.
3. Create a collection in Solr.
4. Index the Employee data from Kaggle datasets.

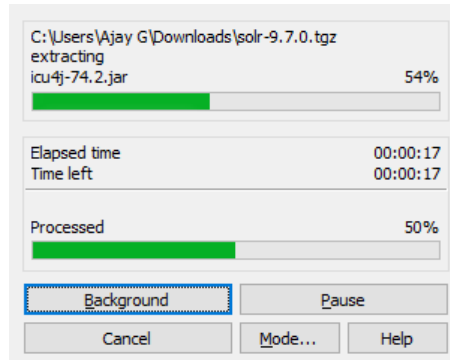
Step 1: Install Apache Solr on local computer



Download the Kaggle Datasets



Extract the Downloaded Solr.tgz



Run the Apache Solra using below commands:

- Solra.cmd start

```
Microsoft Windows [Version 10.0.16299.1087]
(c) 2017 Microsoft Corporation. All rights reserved.

C:\Users\Ajay G\Desktop\Agile Test Test\solr-9.7.0\bin>solr.cmd start
Java 19 detected. Enabled workaround for SOLR-16463
Java HotSpot(TM) 64-Bit Server VM warning: JVM cannot use large page memory because it does not have enough privilege to
lock pages in memory.
WARNING: A command line option has enabled the Security Manager
WARNING: The Security Manager is deprecated and will be removed in a future release
Failed to parse command-line arguments due to: Unrecognized option: --max-wait-secs

usage: bin/solr status [-maxWaitSecs <SECS>] [--solr-url <URL>]

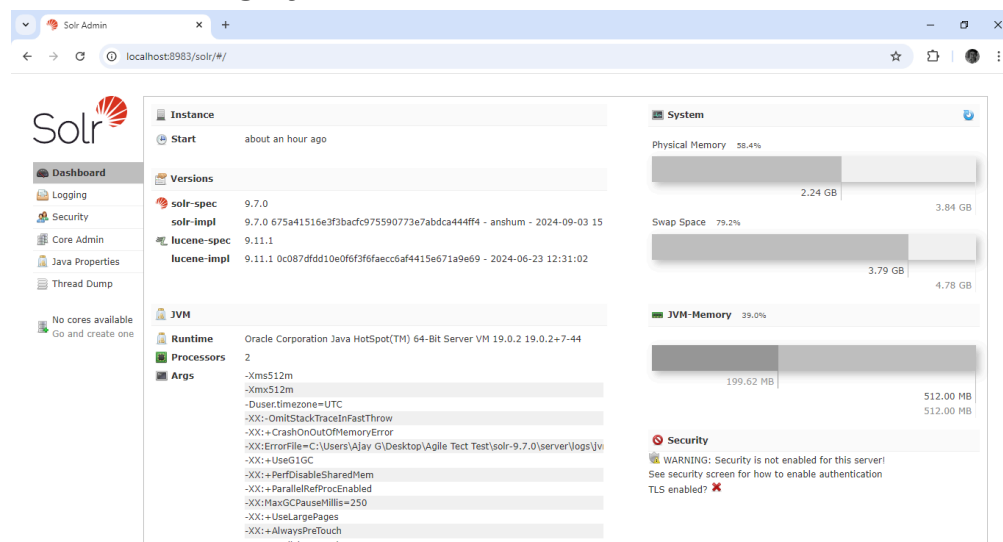
List of options:
-maxWaitSecs <SECS>  Wait up to the specified number of seconds to see Solr running.
--solr-url <URL>      Address of the Solr Web application, defaults to: http://localhost:8983.

Please see the Reference Guide for more tools documentation:
https://solr.apache.org/guide/solr/latest/deployment-guide/solr-control-script-reference.html

ERROR: Solr did not start or was not reachable. Check the logs for errors.

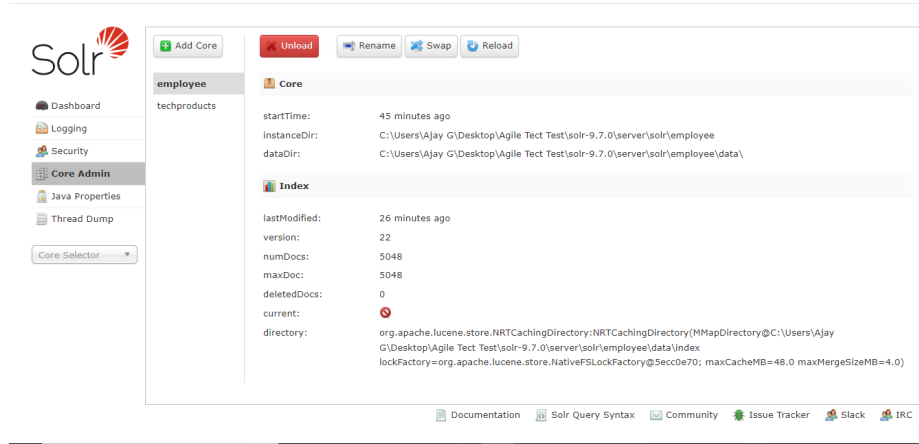
C:\Users\Ajay G\Desktop\Agile Test Test\solr-9.7.0\bin>
```

Verify Solra is running by chrome localhost:8983

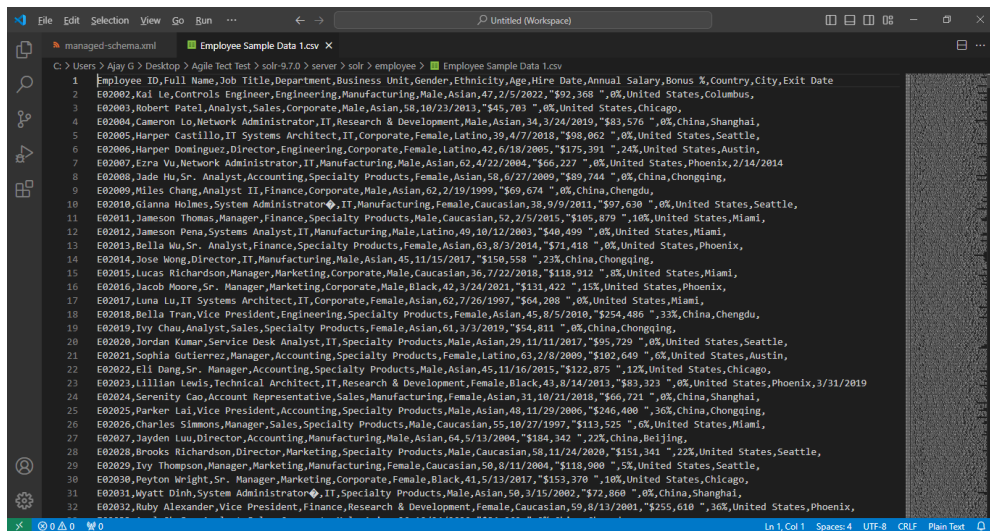
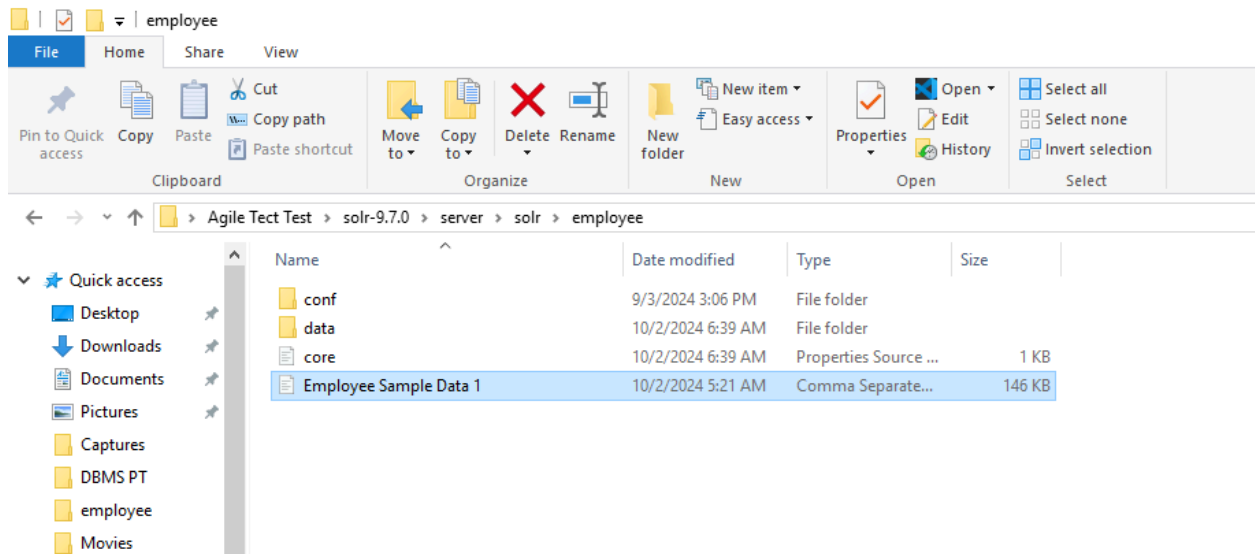


Create new Collection in Solra

I created an employee Collection using **bin\solra.cmd create -c employee**



Index the Employee Data from kaggle:



```
File Edit Selection View Go Run ... Untitled (Workspace)
managed-schema.xml solrconfig.xml managed-schema.xml C:\Employee... X
C:\Users\Ajay G> Desktop> Agile Test Test> solr-9.7.0> server> solr> employee> conf> managed-schema.xml
41 <schema name="default-config" version="1.7">
102
103 <!-- In this default configset, only four fields are pre-declared:
104 id, version, and text_and_root. All other fields will be type guessed and added via the
105 "add-unknown-fields-to-the-schema" update request processor chain declared in solrconfig.xml.
106
107 Note that many dynamic fields are also defined - you can use them to specify a
108 field's type via field naming conventions - see below.
109
110 WARNING: The text_catch-all field will significantly increase your index size.
111 If you don't need it, consider removing it and the corresponding copyField directive.
112
113 -->
114
115 <field name="id" type="string" indexed="true" stored="true" required="true" multiValued="false" />
116 <!-- docValues are enabled by default for long type so we don't need to index the version field -->
117 <field name="_version_" type="plong" indexed="false" stored="false"/>
118
119 <!-- If you don't use child/nested documents, then you should remove the next two fields: -->
120 <!-- for nested documents (minimal; points to root document) -->
121 <field name="_root_" type="string" indexed="true" stored="false" />
122 <!-- for nested documents (relationship tracking) -->
123 <field name="_nest_path_" type="_nest_path_" /><field type="name" _nest_path_" class="solr.NestPathField" />
124
125 <field name="text" type="text_general" indexed="true" stored="false" multiValued="true"/>
126
127 <!-- This can be enabled, in case the client does not know what fields may be searched. It isn't enabled by default
128 because it's very expensive to index everything twice. -->
129 <!-- <copyfield source="*" dest="text"/> -->
130
131 <!-- Dynamic field definitions allow using convention over configuration
132 for fields via the specification of patterns to match field names.
133 EXAMPLE: name="*" will match any field ending in _ (like myid1, _e1)
```

Verify the Indexed data from dataset: Using the Query "."

Solr Admin

localhost:8983/solr/#/employee/query?q=*:*&q.op=OR&indent=true&useParams=

employee

start, rows
0 10

fl

df

paramset(s)

wt

indent on

debugQuery

defType

hl

facet

spatial

spellcheck

Raw Query Parameters

JSON Query

```
"response":{
  "numFound":5048,
  "start":0,
  "numFoundExact":true,
  "docs":[
    {
      "Department":["Engineering"],
      "Gender":["Male"],
      "Ethnicity":["Asian"],
      "Age":[47],
      "Country":["United States"],
      "City":["Columbus"],
      "id":["66e0c7db-4fcf-4a83-8afe-5ca50f5f26f4"],
      "Employee_ID":["E02002"],
      "Full Name":["Kai Le"],
      "Job Title":["Controls Engineer"],
      "Business Unit":["Manufacturing"],
      "Hire Date":["2/25/2022"],
      "Annual Salary":["$92,368 "],
      "Bonus _":["0$"],
      "_version_":["101810796350472192"],
      "_root _":["66e0c7db-4fcf-4a83-8afe-5ca50f5f26f4"]
    },
    {
      "Department":["Sales"],
      "Gender":["Male"],
      "Ethnicity":["Asian"],
      "Age":[58],
      "Country":["United States"],
      "City":["Chicago"],
      "id":["554cd126-ce42-4a84-b26a-fd7c9ee219e7"],
      "Employee_ID":["E02003"],
```

```
File Edit Selection View Go Run ... Untitled (Workspace)
managed-schema.xml Employee Sample Data 1.csv X
C:\Users\Ajay G> Desktop> Agile Test Test> solr-9.7.0> server> solr> Employee Sample Data 1.csv
1 Employee_ID,Full Name,Job Title,Department,Business Unit,Gender,Ethnicity,Age,Hire Date,Annual Salary,Bonus %,Country,City,Exit Date
2 E02002,Kai Le,Controls Engineer,Engineering,Manufacturing,Male,Asian,47,2/25/2022,"$92,368 ",0$,United States,Columbus,
3 E02003,Robert Patel,Analyst,Sales,Corporate,Male,Asian,58,10/23/2018,"$45,703 ",0$,United States,Chicago,
4 E02004,Cameron Lo,Network Administrator,IT,Research & Development,Male,Asian,34,3/24/2019,"$83,526 ",0$,China,Shanghai,
5 E02005,Harper Castillo,IT Systems Architect,IT,Corporate,Female,Latino,39,4/7/2018,"$98,062 ",0$,United States,Seattle,
6 E02006,Harper Dominguez,Director,Engineering,Corporate,Female,Latino,42,6/18/2005,"$175,391 ",24%,United States,Austin,
7 E02007,Ezra Wu,Network Administrator,IT,Manufacturing,Male,Asian,62,4/22/2004,"$66,227 ",0$,United States,Phoenix,2/14/2014
8 E02008,Jade Wu,Sr. Analyst,Accounting,Specialty Products,Female,Asian,59,6/27/2009,"$89,744 ",0$,China,Chongqing,
9 E02009,Miles Chang,Analyst II,Finance,Corporate,Male,Asian,62,2/19/1999,"$69,674 ",0$,China,Chengdu,
10 E02010,Gianna Holmes,System Administrator,IT,Manufacturing,Female,Caucasian,38,9/9/2011,"$97,630 ",0$,United States,Seattle,
11 E02011,Jameson Thomas,Manager,Finance,Specialty Products,Male,Caucasian,52,2/5/2015,"$185,879 ",10%,United States,Miami,
12 E02012,Jameson Pena,Systems Analyst,IT,Manufacturing,Male,Latino,49,10/12/2003,"$48,499 ",0$,United States,Miami,
13 E02013,Bella Wu,Sr. Analyst,Finance,Specialty Products,Female,Asian,61,8/3/2014,"$71,418 ",0$,United States,Phoenix,
14 E02014,Jose Wang,Director,IT,Manufacturing,Male,Asian,45,11/15/2017,"$150,558 ",22%,China,Chongqing,
15 E02015,Lucas Richardson,Manager,Marketing,Corporate,Male,Caucasian,36,7/22/2018,"$118,912 ",0$,United States,Miami,
16 E02016,Jacob Moore,Sr. Manager,Marketing,Corporate,Male,Black,42,3/24/2021,"$131,422 ",15%,United States,Phoenix,
17 E02017,Luna Lu,IT Systems Architect,IT,Corporate,Female,Asian,62,7/26/1997,"$64,288 ",0$,United States,Miami,
18 E02018,Bella Tran,Vice President,Engineering,Specialty Products,Female,Asian,45,8/5/2018,"$254,486 ",33%,China,Chengdu,
19 E02019,Ivy Chau,Analyst,Sales,Specialty Products,Female,Asian,61,3/3/2019,"$54,811 ",0$,China,Chongqing,
20 E02020,Jordan Kumar,Service Desk Analyst,IT,Specialty Products,Male,Asian,28,11/11/2017,"$95,729 ",0$,United States,Seattle,
21 E02021,Sophia Gutierrez,Manager,Accounting,Specialty Products,Female,Latino,63,2/8/2009,"$182,649 ",0$,United States,Austin,
22 E02022,Eli Dang,Sr. Manager,Accounting,Specialty Products,Male,Asian,45,11/16/2015,"$122,875 ",12%,United States,Chicago,
23 E02023,Lillian Lewis,Technical Architect,IT,Research & Development,Female,Black,43,8/14/2013,"$83,323 ",0$,United States,Phoenix,3/31/2019
24 E02024,Serenity Cao,Account Representative,Sales,Manufacturing,Female,Asian,31,10/21/2018,"$66,721 ",0$,China,Shanghai,
25 E02025,Parker Lei,Vice President,Accounting,Specialty Products,Male,Asian,48,11/29/2006,"$146,488 ",30%,China,Chongqing,
26 E02026,Charles Simmons,Manager,Sales,Specialty Products,Male,Caucasian,55,10/27/1997,"$113,525 ",0$,United States,Miami,
27 E02027,Jayden Luu,Director,Accounting,Manufacturing,Male,Asian,64,5/13/2004,"$184,342 ",22%,China,Beijing,
28 E02028,Brooks Richardson,Director,Marketing,Specialty Products,Male,Caucasian,58,11/24/2020,"$151,341 ",22%,United States,Seattle,
29 E02029,Ivy Thompson,Manager,Marketing,Manufacturing,Female,Caucasian,50,8/1/2004,"$118,900 ",5%,United States,Seattle,
30 E02030,Peyton Wright,Sr. Manager,Marketing,Corporate,Female,Black,41,5/13/2017,"$153,378 ",10%,United States,Chicago,
31 E02031,Wyatt Dinh,System Administrator,IT,Specialty Products,Male,Asian,50,3/15/2002,"$72,866 ",0$,China,Shanghai,
32 E02032,Ruby Alexander,Vice President,Finance,Research & Development,Female,Caucasian,59,8/13/2001,"$255,610 ",36%,United States,Phoenix,
```

