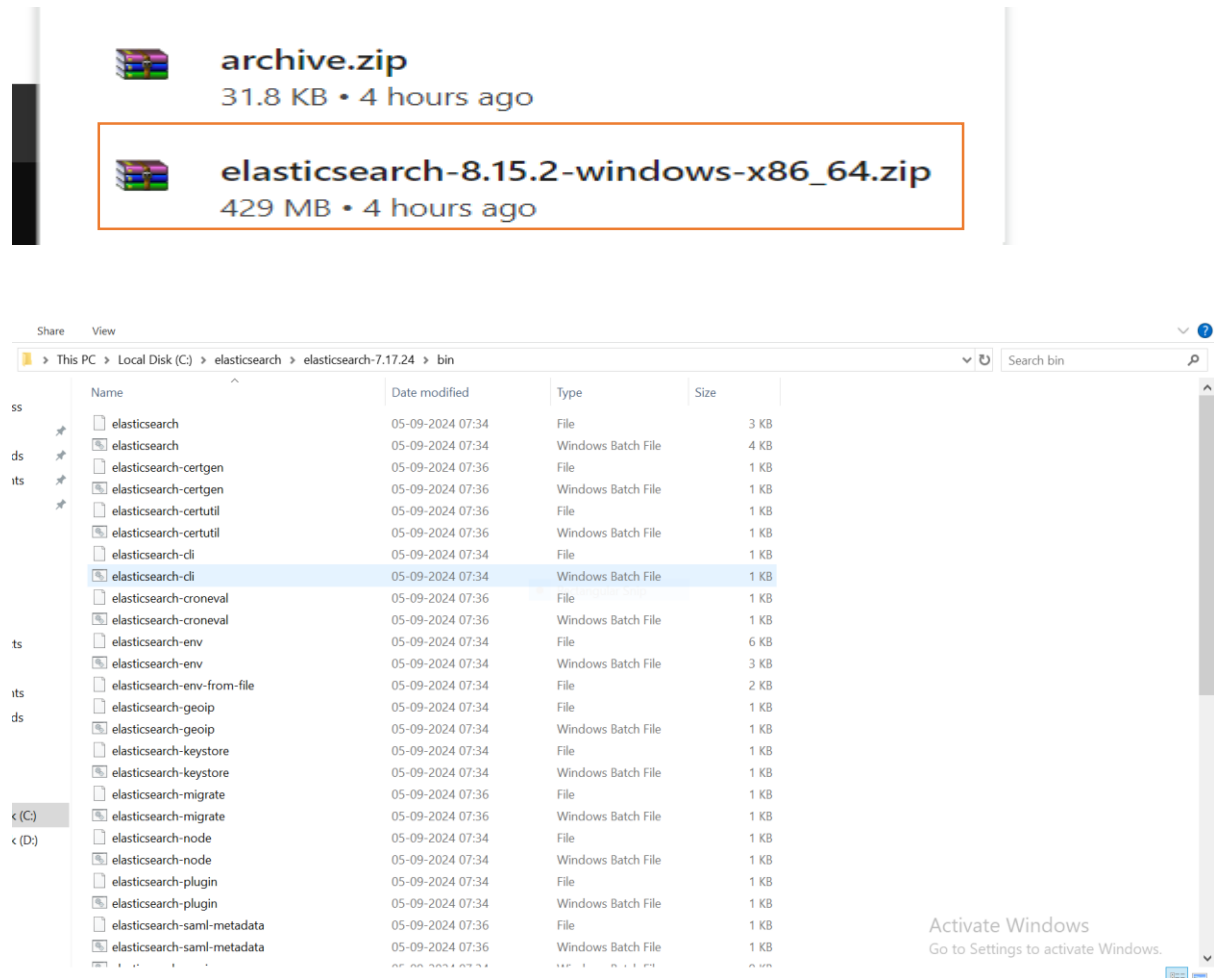


ROUND2 ASSIGNMENT

1. Install Elasticsearch on your local machine.
2. Create an index (collection) in Elasticsearch.
3. Index the Employee data
from <https://www.kaggle.com/datasets/williamlucas0/employee-sample-data>

SOLUTION

Install Elasticsearch in your local machine.




`cd path\to\elasticsearch\bin`


To start elasticsearch—`elasticsearc.bat`

```
Command Prompt - C:\Windows\System32\cmd.exe - elasticsearch.bat
C:\elasticsearch\elasticsearch-7.17.24\bin>elasticsearch.bat
Oct 02, 2024 10:06:57 PM sun.util.locale.provider.LocaleProviderAdapter <clinit>
WARNING: COMPAT locale provider will be removed in a future release
[2024-10-02T22:07:01.713][INFO][o.e.n.Node] [DESKTOP-OPECQ09] version[7.17.24], pid[1368], build[default/zip/fcf25fff740db6ab3ed5d145c58d70e4c3528ea7/2024-09-08T07:34:51.812485320Z], OS[Windows 10/10.0/amd64], JVM[Oracle Corporation/OpenJDK 64-Bit Server VM/22.0.1/22.0.1+8-16]
[2024-10-02T22:07:01.722][INFO][o.e.n.Node] [DESKTOP-OPECQ09] JVM home [C:\elasticsearch\elasticsearch-7.17.24\jdk], using bundled JDK [true]
[2024-10-02T22:07:01.722][INFO][o.e.n.Node] [DESKTOP-OPECQ09] JVM arguments [-Des.networkaddress.cache.ttl=60, -Des.networkaddress.cache.negative.ttl=10, -X
X:+AlwaysPreTouch, -Xss1m, -Djava.awt.headless=true, -Dfile.encoding=UTF-8, -Djna.nosys=true, -XX:-OmitStackTraceInFastThrow, -XX:+ShowCodeDetailsInExceptionMessages, -Dio
netty.noUnsafe=true, -Dio.netty.noKeySetOptimization=true, -Dio.netty.recycler.maxCapacityPerThread=0, -Dio.netty.allocator.numDirectArenas=0, -Dlog4j.shutdownHookEnabled=
false, -Dlog4j2.disable.jmx=true, -Dlog4j2.formatMsgNoLookups=true, -Djava.locale.providers=SPI,COMPAT, --add-opens=java.base/java.io=ALL-UNNAMED, -Djava.security.manager=al
low, -XX:+UseG1GC, -Djava.io.tmpdir=C:\Users\Admin\AppData\Local\Temp\elasticsearch, -XX:+HeapDumpOnOutOfMemoryError, -XX:+ExitOnOutOfMemoryError, -XX:HeapDumpPath=data, -X
X:ErrorFile=logs/hs_err_pid%p.log, -Xlog:gc*,gc+age=trace,safepoint:file=logs/gc.log:utctime,pid,tags:filecount=32,filesize=64m, -XX:+UnlockDiagnosticVMOptions, -XX:G1NumCo
llectionsKeepPinned=10000000, -Xms4033m, -Xmx4033m, -XX:MaxDirectMemorySize=2114977792, -XX:G1HeapRegionSize=4m, -XX:InitiatingHeapOccupancyPercent=30, -XX:G1ReservePercent
=15, -Delasticsearch, -Des.path.home=C:\elasticsearch\elasticsearch-7.17.24, -Des.path.conf=C:\elasticsearch\elasticsearch-7.17.24\config, -Des.distribution.flavor=default,
-Des.distribution.type=zip, -Des.bundled_jdk=true]
[2024-10-02T22:07:08.747][INFO][o.e.p.PluginsService] [DESKTOP-OPECQ09] loaded module [aggs-matrix-stats]
[2024-10-02T22:07:08.747][INFO][o.e.p.PluginsService] [DESKTOP-OPECQ09] loaded module [analysis-common]
[2024-10-02T22:07:08.747][INFO][o.e.p.PluginsService] [DESKTOP-OPECQ09] loaded module [constant-keyword]
[2024-10-02T22:07:08.755][INFO][o.e.p.PluginsService] [DESKTOP-OPECQ09] loaded module [frozen-indices]
[2024-10-02T22:07:08.755][INFO][o.e.p.PluginsService] [DESKTOP-OPECQ09] loaded module [ingest-common]
[2024-10-02T22:07:08.755][INFO][o.e.p.PluginsService] [DESKTOP-OPECQ09] loaded module [ingest-geoip]
[2024-10-02T22:07:08.755][INFO][o.e.p.PluginsService] [DESKTOP-OPECQ09] loaded module [ingest-user-agent]
[2024-10-02T22:07:08.755][INFO][o.e.p.PluginsService] [DESKTOP-OPECQ09] loaded module [kibana]
[2024-10-02T22:07:08.755][INFO][o.e.p.PluginsService] [DESKTOP-OPECQ09] loaded module [lang-expression]
[2024-10-02T22:07:08.755][INFO][o.e.p.PluginsService] [DESKTOP-OPECQ09] loaded module [lang-mustache]
[2024-10-02T22:07:08.755][INFO][o.e.p.PluginsService] [DESKTOP-OPECQ09] loaded module [lang-painless]
[2024-10-02T22:07:08.763][INFO][o.e.p.PluginsService] [DESKTOP-OPECQ09] loaded module [legacy-geo]
[2024-10-02T22:07:08.763][INFO][o.e.p.PluginsService] [DESKTOP-OPECQ09] loaded module [mapper-extras]
[2024-10-02T22:07:08.763][INFO][o.e.p.PluginsService] [DESKTOP-OPECQ09] loaded module [mapper-version]
[2024-10-02T22:07:08.763][INFO][o.e.p.PluginsService] [DESKTOP-OPECQ09] loaded module [parent-join]
[2024-10-02T22:07:08.763][INFO][o.e.p.PluginsService] [DESKTOP-OPECQ09] loaded module [percolator]
[2024-10-02T22:07:08.772][INFO][o.e.p.PluginsService] [DESKTOP-OPECQ09] loaded module [rank-eval]
[2024-10-02T22:07:08.772][INFO][o.e.p.PluginsService] [DESKTOP-OPECQ09] loaded module [reindex]
[2024-10-02T22:07:08.772][INFO][o.e.p.PluginsService] [DESKTOP-OPECQ09] loaded module [repositories-metering-api]
[2024-10-02T22:07:08.780][INFO][o.e.p.PluginsService] [DESKTOP-OPECQ09] loaded module [repository-encrypted]
[2024-10-02T22:07:08.780][INFO][o.e.p.PluginsService] [DESKTOP-OPECQ09] loaded module [repository-url]
[2024-10-02T22:07:08.780][INFO][o.e.p.PluginsService] [DESKTOP-OPECQ09] loaded module [runtime-fields-common]
[2024-10-02T22:07:08.780][INFO][o.e.p.PluginsService] [DESKTOP-OPECQ09] loaded module [search-business-rules]
[2024-10-02T22:07:08.780][INFO][o.e.p.PluginsService] [DESKTOP-OPECQ09] loaded module [searchable-snapshots]
[2024-10-02T22:07:08.780][INFO][o.e.p.PluginsService] [DESKTOP-OPECQ09] loaded module [snapshot-repo-test-kit]
```

Verify installation - <http://localhost:9200> – visit this page to ensure the details of elasticsearch

Download dataset from Kaggle

**archive.zip**
31.8 KB • 4 hours ago

**elasticsearch-8.15.2-windows-x86_64.zip**
429 MB • 4 hours ago

Activate Windows
Go to Settings to activate Windows.

Downloaded dataset

Employee Sample Data 1 - Excel																		
File Home Insert Page Layout Formulas Data Review View Tell me what you want to do... Sign in Share																		
Clipboard Font Alignment Number Styles Cells Editing																		
Employee ID																		
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Employee	Full Name	Job Title	Departme	Business U	Gender	Ethnicity	Age	Hire Date	Annual Sa	Bonus %	Country	City	Exit Date				
2	E02002	Kai Le	Controls E	Engineer	Manufact	Male	Asian		47	#####	\$92,368	0%	United Sta	Columbus				
3	E02003	Robert Pa	Analyst	Sales	Corporate	Male	Asian		58	10/23/2015	\$45,703	0%	United Sta	Chicago				
4	E02004	Cameron I	Network A	IT	Research i	Male	Asian		34	3/24/2015	\$83,576	0%	China	Shanghai				
5	E02005	Harper Ca	IT Systems	IT	Corporate	Female	Latino		39	#####	\$98,062	0%	United Sta	Seattle				
6	E02006	Harper De	Director	Engineer	Corporate	Female	Latino		42	6/18/2005	\$175,391	24%	United Sta	Austin				
7	E02007	Ezra Vu	Network A	IT	Manufact	Male	Asian		62	4/22/2004	\$66,227	0%	United Sta	Phoenix	2/14/2014			
8	E02008	Jade Hu	Sr. Analyst	Accountin	Specialty	F Female	Asian		58	6/27/2005	\$89,744	0%	China	Chongqing				
9	E02009	Miles Chai	Analyst II	Finance	Corporate	Male	Asian		62	2/19/1995	\$69,674	0%	China	Chengdu				
10	E02010	Gianna Hc	System Ad	IT	Manufact	Female	Caucasian		38	#####	\$97,630	0%	United Sta	Seattle				
11	E02011	Jameson T	Manager	Finance	Specialty	F Female	Caucasian		52	#####	\$105,879	10%	United Sta	Miami				
12	E02012	Jameson F	Systems A	IT	Manufact	Male	Latino		49	#####	\$40,499	0%	United Sta	Miami				
13	E02013	Bella Wu	Sr. Analyst	Finance	Specialty	F Female	Asian		63	#####	\$71,418	0%	United Sta	Phoenix				
14	E02014	Jose Wong	Director	IT	Manufact	Male	Asian		45	11/15/2015	\$150,558	23%	China	Chongqing				
15	E02015	Lucas Rich	Manager	Marketing	Corporate	Male	Caucasian		36	7/22/2018	\$118,912	8%	United Sta	Miami				
16	E02016	Jacob Moc	Sr. Manag	Marketing	Corporate	Male	Black		42	3/24/2021	\$131,422	15%	United Sta	Phoenix				
17	E02017	Luna Lu	IT Systems	IT	Corporate	Female	Asian		62	7/26/1997	\$64,208	0%	United Sta	Miami				
18	E02018	Bella Tran	Vice Presi	Engineering	Specialty	F Female	Asian		45	#####	\$254,486	33%	China	Chengdu				
19	E02019	Ivy Chau	Analyst	Sales	Specialty	F Female	Asian		61	#####	\$54,811	0%	China	Chongqing				
20	E02020	Jordan Ku	Service De	IT	Specialty	F Female	Asian		29	#####	\$95,729	0%	United Sta	Seattle				
21	E02021	Sophia Gu	Manager	Accountin	Specialty	F Female	Latino		63	#####	\$102,649	6%	United Sta	Austin				
22	E02022	Eli Dang	Sr. Manag	Accountin	Specialty	F Female	Asian		45	11/16/2015	\$122,875	12%	United Sta	Chicago				

Python code

```
import requests

import pandas as pd

from requests.auth import HTTPBasicAuth

# Replace these with your Elasticsearch details

es_url = 'http://localhost:9200'

username = 'elastic' # Your username

password = '*****' # Your password

# Function to index documents

def index_documents(index_name, dataframe):

    print(f"Indexing documents into '{index_name}'...")

    for index, document in dataframe.iterrows():

        url = f"{es_url}/{index_name}/_doc/{document['EmployeeID']}" # Use EmployeeID as the document ID

        response = requests.put(url, auth=(username, password), json=document.to_dict())

        if response.status_code in [200, 201]:

            print(f"Document {index} indexed successfully.")

        else:

            print(f"Failed to index document {index}. Response: {response.json()}")

# Load employee data from CSV
```

try:

```
employee_data = pd.read_csv('employee_data.csv') # Update this path  
print("Employee data loaded successfully.")
```

except Exception as e:

```
print(f"Error loading employee data: {e}")  
exit()
```

Create index and index documents

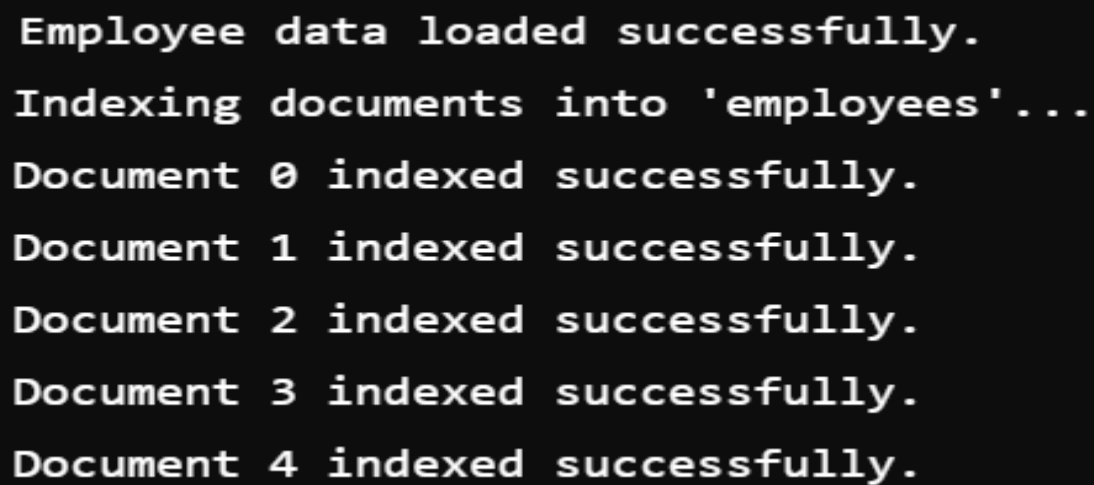
```
index_name = 'employees'
```

```
index_documents(index_name, employee_data)
```

*In terminal- **python index_employees.py***

*Verify data indexing - **curl -X GET "localhost:9200/employees/_search?pretty"***

Final Output screenshot



```
Employee data loaded successfully.  
Indexing documents into 'employees'...  
Document 0 indexed successfully.  
Document 1 indexed successfully.  
Document 2 indexed successfully.  
Document 3 indexed successfully.  
Document 4 indexed successfully.
```