Hash Agile Technology – Round 2

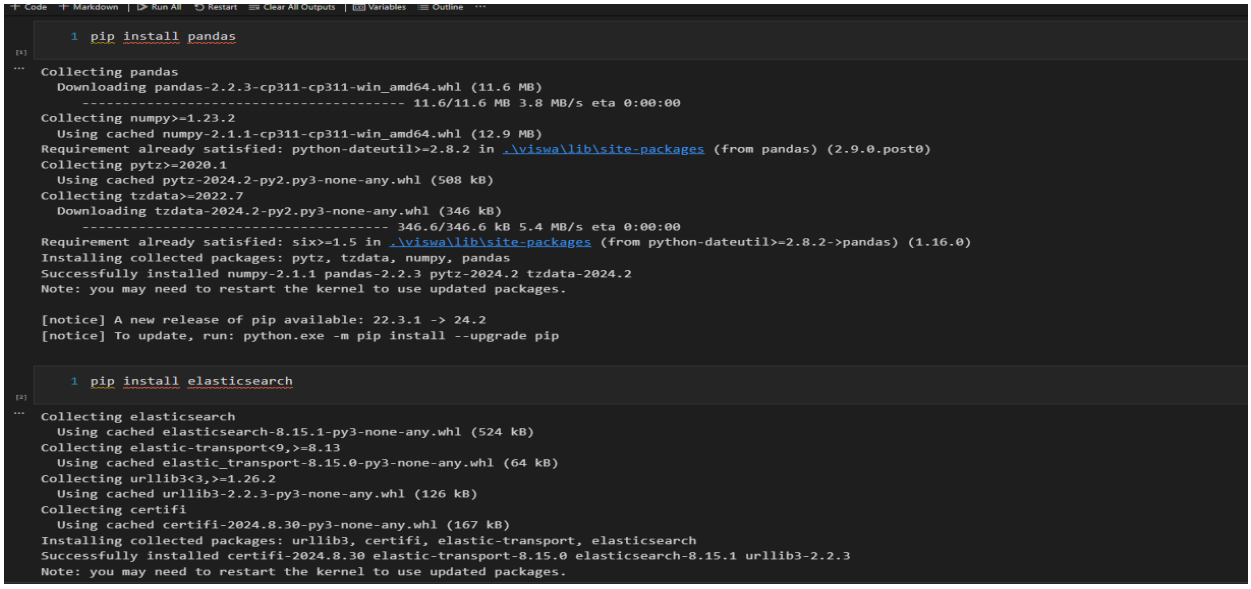
==========

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

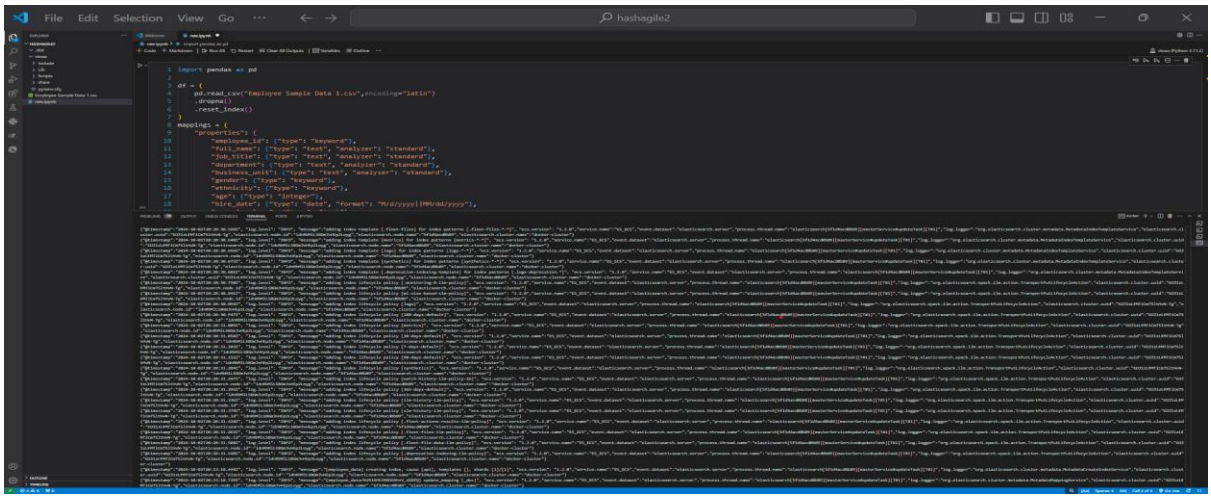
==========

Solution Screenshot:

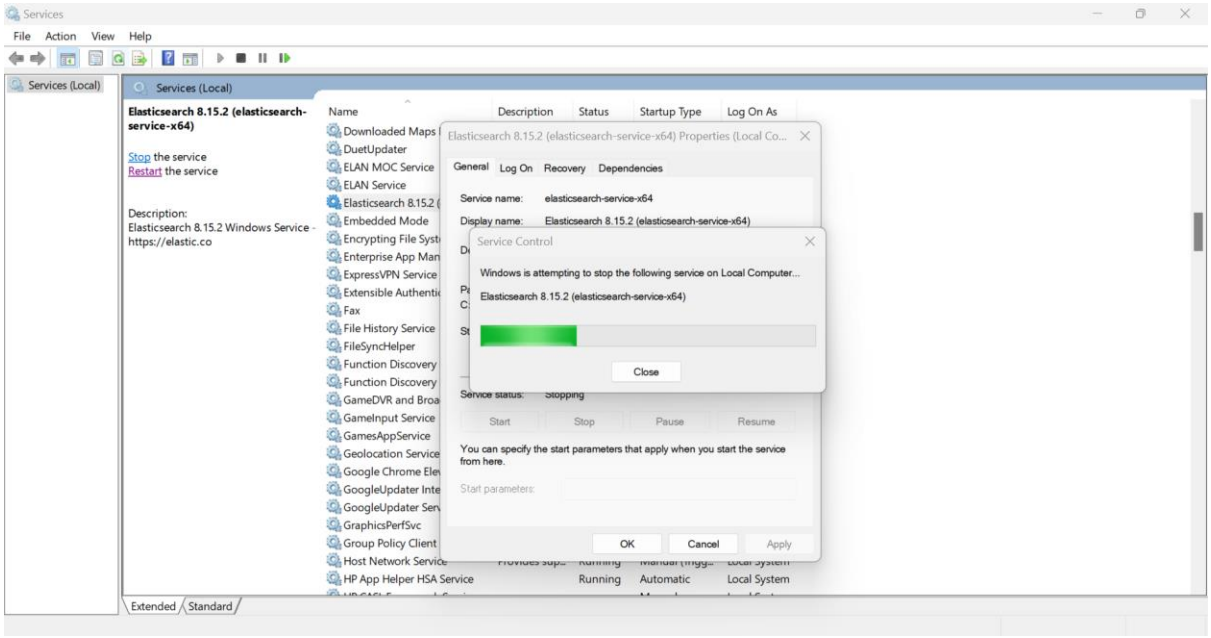
Step 1: Install necessary packages such as pandas,Elastic search etc.,



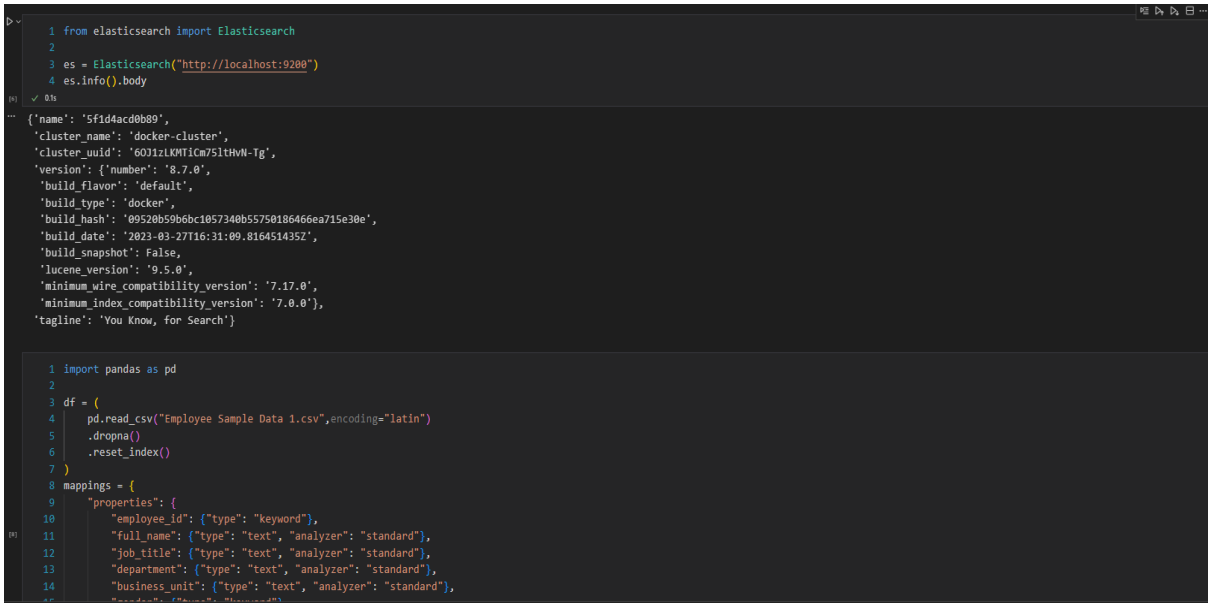
Step 2 : Download and connect the docker



Step 3: Start the Elasticsearch.bat



Step 4: Provide indices to the data set and upload the data set into VS code and also connect to the elastic search localhost:9200



```
from elasticsearch import Elasticsearch

es = Elasticsearch("http://localhost:9200")
es.info().body
```

0.1s

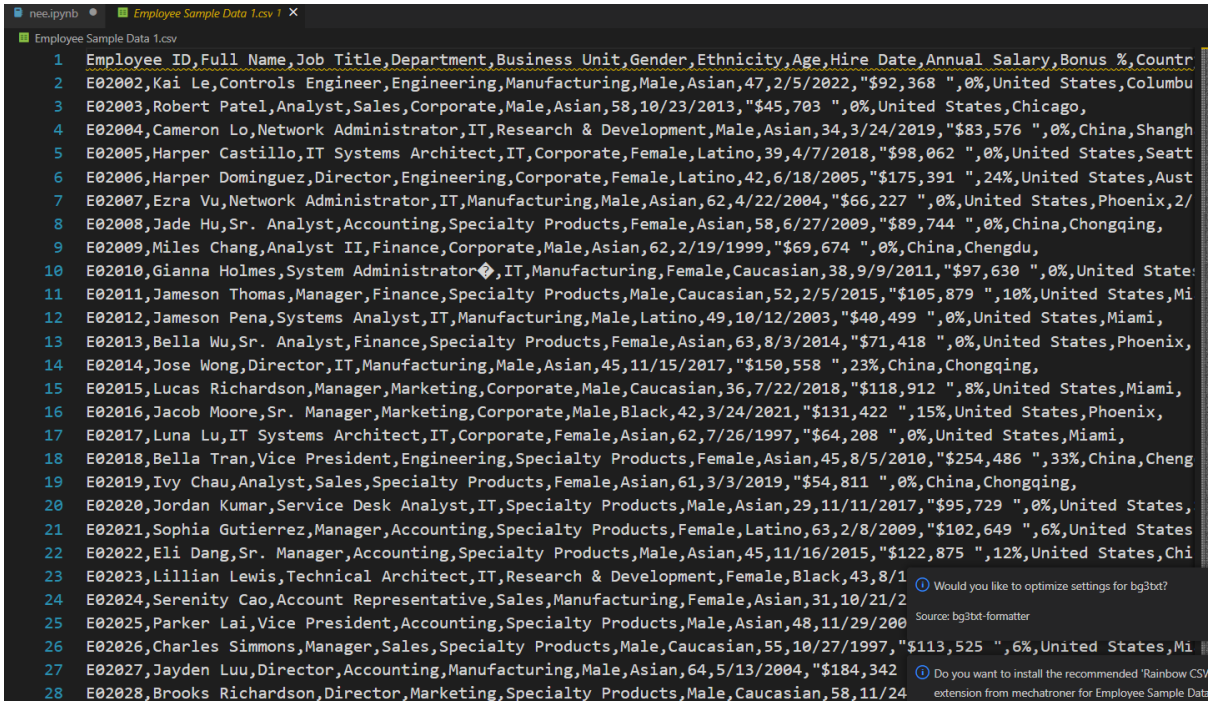
```
{'name': '5f1d4acd0b89',
 'cluster_name': 'docker-cluster',
 'cluster_uuid': '6OJ1zLKMTiCm75ltHvN-Tg',
 'version': {'number': '8.7.0',
  'build_flavor': 'default',
  'build_type': 'docker',
  'build_hash': '09520b59b6bc1057340b55750186466ea715e30e',
  'build_date': '2023-03-27T16:31:09.816451435Z',
  'build_snapshot': False,
  'lucene_version': '9.5.0',
  'minimum_wire_compatibility_version': '7.17.0',
  'minimum_index_compatibility_version': '7.0.0'},
 'tagline': 'You Know, for Search'}
```

```
import pandas as pd

df = (
    pd.read_csv("Employee Sample Data 1.csv",encoding="latin")
    .dropna()
    .reset_index()
)
mappings = {
    "properties": {
        "employee_id": {"type": "keyword"},
        "full_name": {"type": "text", "analyzer": "standard"},
        "job_title": {"type": "text", "analyzer": "standard"},
        "department": {"type": "text", "analyzer": "standard"},
        "business_unit": {"type": "text", "analyzer": "standard"},
```

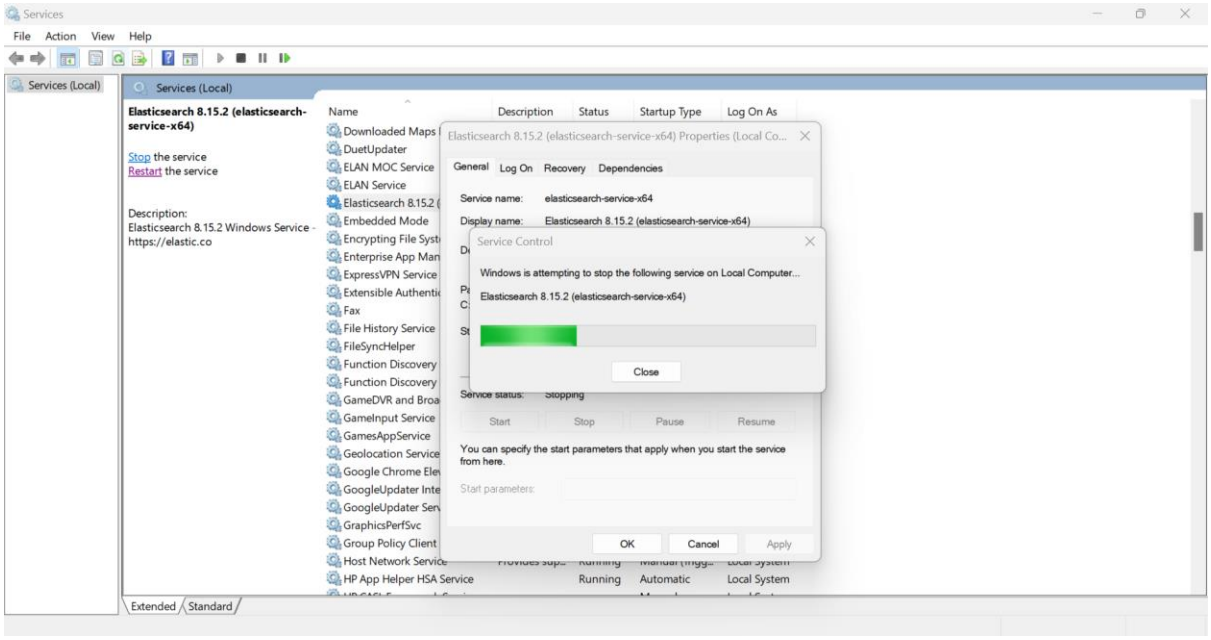
```
mappings = {
    "properties": {
        "employee_id": {"type": "keyword"},
        "full_name": {"type": "text", "analyzer": "standard"},
        "job_title": {"type": "text", "analyzer": "standard"},
        "department": {"type": "text", "analyzer": "standard"},
        "business_unit": {"type": "text", "analyzer": "standard"},
        "gender": {"type": "keyword"},
        "ethnicity": {"type": "keyword"},
        "age": {"type": "integer"},
        "hire_date": {"type": "date", "format": "M/d/yyyy||MM/dd/yyyy"},
        "annual_salary": {"type": "text"},
        "bonus_percent": {"type": "text"},
        "country": {"type": "keyword"},
        "city": {"type": "keyword"},
        "exit_date": {"type": "date", "format": "M/d/yyyy", "null_value": None}
    }
}
if not es.indices.exists(index="employee_data"):
    es.indices.create(index="employee_data", mappings=mappings)
for i, row in df.iterrows():
    doc = {
            "id": row["Employee ID"],
            "Full Name": row["Full Name"],
            "Job Title": row["Job Title"],
            "Department": row["Department"],
            "Business Unit": row["Business Unit"],
            "Gender": row["Gender"],
            "Ethnicity": row["Ethnicity"],
            "Age": row["Age"],
            "Hire Date": row["Hire Date"],
            "Annual Salary": row["Annual Salary"],
            "Bonus": row["Bonus %"],
            "Country": row["Country"],
            "City": row["City"]
        }
    if pd.notnull(row["Exit Date"]):
        try:
            exit_date = pd.to_datetime(row["Exit Date"], format="%m/%d/%Y")
            doc["Exit Date"] = exit_date.strftime('%Y-%m-%d')
        except ValueError:
```

```
            pass
    es.index(index="employee_data", id=i, document=doc)
es.indices.refresh(index="employee_data")
es.cat.count(index="employee_data", format="json")
resp = es.search(
    index="employee_data",
    query={
        "bool": {
            "must": {
                "match_phrase": {
                    "Job Title": "Network Administrator"
                }
            },
            "filter": {
                "bool": {
                    "must": {
                        "match": {"City": "Phoenix"}
                    },
                    "must_not": {
                        "match_phrase": {"Department": "Human Resources"}
                    }
                }
            }
        }
    }
)
for hit in resp['hits']['hits']:
    print(hit["_score"])
```

```
[8]  ✓ 6.2s

...  5.643914
```



```
Employee ID,Full Name,Job Title,Department,Business Unit,Gender,Ethnicity,Age,Hire Date,Annual Salary,Bonus %,Countr
E02002,Kai Le,Controls Engineer,Engineering,Manufacturing,Male,Asian,47,2/5/2022,"$92,368 ",0%,United States,Columbu
E02003,Robert Patel,Analyst,Sales,Corporate,Male,Asian,58,10/23/2013,"$45,703 ",0%,United States,Chicago,
E02004,Cameron Lo,Network Administrator,IT,Research & Development,Male,Asian,34,3/24/2019,"$83,576 ",0%,China,Shangh
E02005,Harper Castillo,IT Systems Architect,IT,Corporate,Female,Latino,39,4/7/2018,"$98,062 ",0%,United States,Seatt
E02006,Harper Dominguez,Director,Engineering,Corporate,Female,Latino,42,6/18/2005,"$175,391 ",24%,United States,Aust
E02007,Ezra Vu,Network Administrator,IT,Manufacturing,Male,Asian,62,4/22/2004,"$66,227 ",0%,United States,Phoenix,2/
E02008,Jade Hu,Sr. Analyst,Accounting,Specialty Products,Female,Asian,58,6/27/2009,"$89,744 ",0%,China,Chongqing,
E02009,Miles Chang,Analyst II,Finance,Corporate,Male,Asian,62,2/19/1999,"$69,674 ",0%,China,Chengdu,
E02010,Gianna Holmes,System Administrator,IT,Manufacturing,Female,Caucasian,38,9/9/2011,"$97,630 ",0%,United State
E02011,Jameson Thomas,Manager,Finance,Specialty Products,Male,Caucasian,52,2/5/2015,"$105,879 ",10%,United States,Mi
E02012,Jameson Pena,Systems Analyst,IT,Manufacturing,Male,Latino,49,10/12/2003,"$40,499 ",0%,United States,Miami,
E02013,Bella Wu,Sr. Analyst,Finance,Specialty Products,Female,Asian,63,8/3/2014,"$71,418 ",0%,United States,Phoenix,
E02014,Jose Wong,Director,IT,Manufacturing,Male,Asian,45,11/15/2017,"$150,558 ",23%,China,Chongqing,
E02015,Lucas Richardson,Manager,Marketing,Corporate,Male,Caucasian,36,7/22/2018,"$118,912 ",8%,United States,Miami,
E02016,Jacob Moore,Sr. Manager,Marketing,Corporate,Male,Black,42,3/24/2021,"$131,422 ",15%,United States,Phoenix,
E02017,Luna Lu,IT Systems Architect,IT,Corporate,Female,Asian,62,7/26/1997,"$64,208 ",0%,United States,Miami,
E02018,Bella Tran,Vice President,Engineering,Specialty Products,Female,Asian,45,8/5/2010,"$254,486 ",33%,China,Cheng
E02019,Ivy Chau,Analyst,Sales,Specialty Products,Female,Asian,61,3/3/2019,"$54,811 ",0%,China,Chongqing,
E02020,Jordan Kumar,Service Desk Analyst,IT,Specialty Products,Male,Asian,29,11/11/2017,"$95,729 ",0%,United States,
E02021,Sophia Gutierrez,Manager,Accounting,Specialty Products,Female,Latino,63,2/8/2009,"$102,649 ",6%,United States
E02022,Eli Dang,Sr. Manager,Accounting,Specialty Products,Male,Asian,45,11/16/2015,"$122,875 ",12%,United States,Chi
E02023,Lillian Lewis,Technical Architect,IT,Research & Development,Female,Black,43,8/1
E02024,Serenity Cao,Account Representative,Sales,Manufacturing,Female,Asian,31,10/21/2
E02025,Parker Lai,Vice President,Accounting,Specialty Products,Male,Asian,48,11/29/200
E02026,Charles Simmons,Manager,Sales,Specialty Products,Male,Caucasian,55,10/27/1997,"$113,525 ",6%,United States,Mi
E02027,Jayden Luu,Director,Accounting,Manufacturing,Male,Asian,64,5/13/2004,"$184,342
E02028,Brooks Richardson,Director,Marketing,Specialty Products,Male,Caucasian,58,11/24
```



Name: Viswanathan R

College: Dr.Mahalingam College of Engineering and Technology, Pollachi.

Department: Artificial Intelligence and Data Science

Mail Id: viswanathan092003@gmail.com