

```
In [30]: from elasticsearch import Elasticsearch
```

```
In [31]: from elasticsearch import Elasticsearch
```

```
# Replace 'elastic' and 'your_password' with your actual username and password
es = Elasticsearch(
    [{"host": "localhost", "port": 9200, "scheme": "http"}], # Specify scheme here
    basic_auth=('elastic', 'GFahX5h4A0m-m7F2Q9Qv'),
    verify_certs=False # Disable SSL certificate verification for localhost
)

# Test connection
print(es.info())
```

```
{'name': 'DESKTOP-PJIL76A', 'cluster_name': 'elasticsearch', 'cluster_uuid': 'lx_ghzATSPiTVAneu8LXqA', 'version'
: {'number': '8.15.2', 'build_flavor': 'default', 'build_type': 'zip', 'build_hash': '98adf7bf6bb69b66ab95b761c9
e5aadb0bb059a3', 'build_date': '2024-09-19T10:06:03.564235954Z', 'build_snapshot': False, 'lucene_version': '9.1
1.1', 'minimum_wire_compatibility_version': '7.17.0', 'minimum_index_compatibility_version': '7.0.0'}, 'tagline'
: 'You Know, for Search'}
```

```
In [32]: index_name = 'employee_office'
```

```
# Check if the index already exists
if not es.indices.exists(index=index_name):
    # Create the index if it does not exist
    es.indices.create(index=index_name)
    print(f"Index '{index_name}' created.")
else:
    print(f"Index '{index_name}' already exists.")
```

Index 'employee_office' already exists.

```
In [33]: es.indices.exists(index="employee_office")
```

```
Out[33]: HeadApiResponse(True)
```

```
In [34]: pip install pandas elasticsearch
```

```
Requirement already satisfied: pandas in c:\users\acer\appdata\local\programs\python\python312\lib\site-packages
(2.2.3)
Requirement already satisfied: elasticsearch in c:\users\acer\appdata\local\programs\python\python312\lib\site-p
ackages (8.15.1)
Requirement already satisfied: numpy>=1.26.0 in c:\users\acer\appdata\local\programs\python\python312\lib\site-p
ackages (from pandas) (2.1.1)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\acer\appdata\local\programs\python\python312\l
ib\site-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in c:\users\acer\appdata\local\programs\python\python312\lib\site-pa
ckages (from pandas) (2024.2)
Requirement already satisfied: tzdata>=2022.7 in c:\users\acer\appdata\local\programs\python\python312\lib\site-
packages (from pandas) (2024.2)
Requirement already satisfied: elastic-transport<9,>=8.13 in c:\users\acer\appdata\local\programs\python\python3
12\lib\site-packages (from elasticsearch) (8.15.0)
Requirement already satisfied: urllib3<3,>=1.26.2 in c:\users\acer\appdata\local\programs\python\python312\lib\s
ite-packages (from elastic-transport<9,>=8.13->elasticsearch) (1.26.20)
Requirement already satisfied: certifi in c:\users\acer\appdata\local\programs\python\python312\lib\site-package
s (from elastic-transport<9,>=8.13->elasticsearch) (2024.8.30)
Requirement already satisfied: six>=1.5 in c:\users\acer\appdata\local\programs\python\python312\lib\site-packag
es (from python-dateutil>=2.8.2->pandas) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

```
In [37]: import pandas as pd
import datetime
import os
import sys
```

```
In [38]: import os
directory = "D:/"
for filename in os.listdir(directory):
    if filename.endswith('.csv'):
        print(filename)
```

Employee Sample Data.csv

```
In [39]: df = pd.read_csv("D:/Employee Sample Data.csv", encoding='ISO-8859-1')
df.head()
```

Out[39]:	Employee ID	Full Name	Job Title	Department	Business Unit	Gender	Ethnicity	Age	Hire Date	Annual Salary	Bonus %	Country
0	E02002	Kai Le	Controls Engineer	Engineering	Manufacturing	Male	Asian	47.0	2/5/2022	\$92,368	0%	United States
1	E02003	Robert Patel	Analyst	Sales	Corporate	Male	Asian	58.0	10/23/2013	\$45,703	0%	United States
2	E02004	Cameron Lo	Network Administrator	IT	Research & Development	Male	Asian	34.0	3/24/2019	\$83,576	0%	China
3	E02005	Harper Castillo	IT Systems Architect	IT	Corporate	Female	Latino	39.0	4/7/2018	\$98,062	0%	United States
4	E02006	Harper Dominguez	Director	Engineering	Corporate	Female	Latino	42.0	6/18/2005	\$175,391	24%	United States

```
In [40]: import json
from elasticsearch import Elasticsearch, helpers
from elasticsearch.exceptions import NotFoundError, RequestError
```

```
In [41]: df = df.where(pd.notnull(df), None)

# Convert the DataFrame to JSON format
json_data = df.to_json(orient='records')

# Load JSON data
data_to_index = json.loads(json_data)

# Generate data for bulk indexing
def generate_data(df):
    for record in df.to_dict(orient='records'):
        # Replace NaN with None for numeric fields (like Age)
        record['Age'] = record['Age'] if pd.notnull(record['Age']) else None
        record['Annual Salary'] = record['Annual Salary'].replace('$', '').strip() if pd.notnull(record['Annual Salary']) else None
        yield {
            "_index": "employee_office",
            "_id": record.get("Employee ID"), # Use Employee ID as document ID
            "_source": record
        }

# Perform bulk indexing
try:
    response = helpers.bulk(es, generate_data(df))
    print("Data indexed successfully. Indexed documents count:", response[0])
except Exception as e:
    print("An unexpected error occurred:", str(e))
    if hasattr(e, 'errors'):
        print("Failed to index documents:", e.errors)
```

Data indexed successfully. Indexed documents count: 1262

```
In [42]: # Define the collections as global variables
collections = {}
```

```
In [43]: def createCollection(p_collection_name):
    return df.copy()
```

```
In [44]: def indexData(df, p_exclude_column):
    indexed_df = df.drop(columns=[p_exclude_column])
    return indexed_df
```

```
In [45]: def searchByColumn(df, p_column_name, p_column_value):
    result = df[df[p_column_name] == p_column_value]
    return result
```

```
In [46]: def getEmpCount(df):
    return df.shape[0]
```

```
In [47]: def delEmpById(df, p_employee_id):
    return df[df['Employee ID'] != p_employee_id]
```

```
In [48]: def getDepFacet(df):
    return df['Department'].value_counts()
```

```
In [49]: v_nameCollection = 'Hash_Ajaykumar'
v_phoneCollection = 'Hash_1234'
```

```
In [50]: name_collection = createCollection(v_nameCollection)
phone_collection = createCollection(v_phoneCollection)
```

```
In [51]: print("Employee Count in Name Collection:", getEmpCount(name_collection))
```

Employee Count in Name Collection: 1262

```
In [52]: name_collection_indexed = indexData(name_collection, 'Department')
phone_collection_indexed = indexData(phone_collection, 'Gender')
```

```
In [53]: name_collection = delEmpById(name_collection, 'E02003')
print("Employee Count after Deletion:", getEmpCount(name_collection))
```

Employee Count after Deletion: 1259

```
In [54]: # Search by Column
it_employees = searchByColumn(name_collection, 'Department', 'IT')
male_employees = searchByColumn(name_collection, 'Gender', 'Male')
it_employees_phone = searchByColumn(phone_collection, 'Department', 'IT')
```

```
In [55]: # Get Department Facets
department_facet_name = getDepFacet(name_collection)
department_facet_phone = getDepFacet(phone_collection)
```

```
In [56]: # Output Results
print("Employees in IT Department:", it_employees)
print("Male Employees:", male_employees)
print("Employees in IT Department (Phone Collection):", it_employees_phone)
print("Department Facet (Name Collection):\n", department_facet_name)
print("Department Facet (Phone Collection):\n", department_facet_phone)
```

Employees	in IT Department:	Employee ID	Full Name	Job Title	Department \
2	E02004	Cameron Lo	Network Administrator	IT	
3	E02005	Harper Castillo	IT Systems Architect	IT	
5	E02007	Ezra Vu	Network Administrator	IT	
8	E02010	Gianna Holmes	System Administrator	IT	
10	E02012	Jameson Pena	Systems Analyst	IT	
...
1253	E02246	Landon Reyes	Director	IT	
1255	E02248	Lucas Alexander	Director	IT	
1258	E02251	Genesis Herrera	Manager	IT	
1259	E02252	Olivia Vazquez	Network Engineer	IT	
1260	E02253	Leilani Ng	Systems Analyst	IT	

	Business Unit	Gender	Ethnicity	Age	Hire Date \
2	Research & Development	Male	Asian	34.0	3/24/2019
3	Corporate	Female	Latino	39.0	4/7/2018
5	Manufacturing	Male	Asian	62.0	4/22/2004
8	Manufacturing	Female	Caucasian	38.0	9/9/2011
10	Manufacturing	Male	Latino	49.0	10/12/2003
...
1253	Research & Development	Male	Latino	44.0	8/23/2014
1255	Manufacturing	Male	Black	41.0	9/20/2013
1258	Research & Development	Female	Latino	34.0	10/3/2015
1259	Specialty Products	Female	Latino	53.0	4/13/2020
1260	Corporate	Female	Asian	48.0	9/19/2011

	Annual Salary	Bonus %	Country	City	Exit Date
2	\$83,576	0%	China	Shanghai	None
3	\$98,062	0%	United States	Seattle	None
5	\$66,227	0%	United States	Phoenix	2/14/2014
8	\$97,630	0%	United States	Seattle	None
10	\$40,499	0%	United States	Miami	None
...
1253	\$171,823	27%	United States	Columbus	None
1255	\$192,944	22%	United States	Miami	None
1258	\$126,898	10%	Brazil	Manaus	None
1259	\$93,053	0%	Brazil	Sao Paulo	None
1260	\$50,513	0%	United States	Seattle	10/30/2019

[347 rows x 14 columns]

Male Employees:	Employee ID	Full Name	Job Title	Department \
0	E02002	Kai Le	Controls Engineer	Engineering
2	E02004	Cameron Lo	Network Administrator	IT
5	E02007	Ezra Vu	Network Administrator	IT
7	E02009	Miles Chang	Analyst II	Finance
9	E02011	Jameson Thomas	Manager	Finance
...
1251	E02244	Connor Howard	Systems Analyst	IT
1253	E02246	Landon Reyes	Director	IT
1254	E02247	Noah Ma	Vice President	Accounting
1255	E02248	Lucas Alexander	Director	IT
1256	E02249	Henry Vo	Controls Engineer	Engineering

	Business Unit	Gender	Ethnicity	Age	Hire Date	Annual Salary \
0	Manufacturing	Male	Asian	47.0	2/5/2022	\$92,368

2	Research & Development	Male	Asian	34.0	3/24/2019	\$83,576
5	Manufacturing	Male	Asian	62.0	4/22/2004	\$66,227
7	Corporate	Male	Asian	62.0	2/19/1999	\$69,674
9	Specialty Products	Male	Caucasian	52.0	2/5/2015	\$105,879
...
1251	Specialty Products	Male	Caucasian	42.0	3/15/2021	\$44,265
1253	Research & Development	Male	Latino	44.0	8/23/2014	\$171,823
1254	Manufacturing	Male	Asian	28.0	11/6/2018	\$201,013
1255	Manufacturing	Male	Black	41.0	9/20/2013	\$192,944
1256	Manufacturing	Male	Asian	60.0	9/13/1996	\$81,699

	Bonus %	Country	City	Exit Date
0	0%	United States	Columbus	None
2	0%	China	Shanghai	None
5	0%	United States	Phoenix	2/14/2014
7	0%	China	Chengdu	None
9	10%	United States	Miami	None
...
1251	0%	United States	Austin	6/27/2022
1253	27%	United States	Columbus	None
1254	31%	United States	Phoenix	None
1255	22%	United States	Miami	None
1256	0%	China	Beijing	None

[578 rows x 14 columns]

Employees in IT Department (Phone Collection):				Employee ID	Full Name	Job Title	Departme
nt \							
2	E02004	Cameron Lo	Network Administrator	IT			
3	E02005	Harper Castillo	IT Systems Architect	IT			
5	E02007	Ezra Vu	Network Administrator	IT			
8	E02010	Gianna Holmes	System Administrator	IT			
10	E02012	Jameson Pena	Systems Analyst	IT			
...
1253	E02246	Landon Reyes	Director	IT			
1255	E02248	Lucas Alexander	Director	IT			
1258	E02251	Genesis Herrera	Manager	IT			
1259	E02252	Olivia Vazquez	Network Engineer	IT			
1260	E02253	Leilani Ng	Systems Analyst	IT			

	Business Unit	Gender	Ethnicity	Age	Hire Date	\
2	Research & Development	Male	Asian	34.0	3/24/2019	
3	Corporate	Female	Latino	39.0	4/7/2018	
5	Manufacturing	Male	Asian	62.0	4/22/2004	
8	Manufacturing	Female	Caucasian	38.0	9/9/2011	
10	Manufacturing	Male	Latino	49.0	10/12/2003	
...
1253	Research & Development	Male	Latino	44.0	8/23/2014	
1255	Manufacturing	Male	Black	41.0	9/20/2013	
1258	Research & Development	Female	Latino	34.0	10/3/2015	
1259	Specialty Products	Female	Latino	53.0	4/13/2020	
1260	Corporate	Female	Asian	48.0	9/19/2011	

	Annual Salary	Bonus %	Country	City	Exit Date
2	\$83,576	0%	China	Shanghai	None
3	\$98,062	0%	United States	Seattle	None
5	\$66,227	0%	United States	Phoenix	2/14/2014
8	\$97,630	0%	United States	Seattle	None
10	\$40,499	0%	United States	Miami	None
...
1253	\$171,823	27%	United States	Columbus	None
1255	\$192,944	22%	United States	Miami	None
1258	\$126,898	10%	Brazil	Manaus	None
1259	\$93,053	0%	Brazil	Sao Paulo	None
1260	\$50,513	0%	United States	Seattle	10/30/2019

[347 rows x 14 columns]

Department Facet (Name Collection):

Department	
IT	347
Sales	185
Engineering	159
Marketing	136
Accounting	132
Finance	129
Human Resources	124

Name: count, dtype: int64

Department Facet (Phone Collection):

Department	
IT	347
Sales	188
Engineering	159
Marketing	136
Accounting	132

Finance 129
Human Resources 124
Name: count, dtype: int64

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js