ELASTICSEARCH

LOCAL HOST


```
{
  "name" : "node-1",
  "cluster_name" : "my-application",
  "cluster_uuid" : "1dDQnDPiQ3-GzclWuYtwoA",
  "version" : {
    "number" : "8.15.2",
    "build_flavor" : "default",
    "build_type" : "zip",
    "build_hash" : "98adf7bf6bb69b66ab95b761c9e5aadb0bb059a3",
    "build_date" : "2024-09-19T10:06:03.564235954Z",
    "build_snapshot" : false,
    "lucene_version" : "9.11.1",
    "minimum_wire_compatibility_version" : "7.17.0",
    "minimum_index_compatibility_version" : "7.0.0"
  },
  "tagline" : "You Know, for Search"
}
```


CODE

```
from elasticsearch import Elasticsearch, helpers
import pandas as pd
import os
from datetime import datetime

es = Elasticsearch(
    "http://localhost:9200",
    basic_auth=("elastic", "elastic")
)


index_name = "my_csv_index"

index_settings = {
    "settings": {
        "number_of_shards": 1,
        "number_of_replicas": 1
    },
    "mappings": {
        "properties": {
            "Employee ID": {"type": "text"},
            "Full Name": {"type": "text"},
            "Job Title": {"type": "text"},
            "Department": {"type": "text"},
            "Business Unit": {"type": "text"},
            "Gender": {"type": "text"},
            "Ethnicity": {"type": "text"},
            "Age": {"type": "integer"},
            "Hire Date": {"type": "date"},
            "Annual Salary": {"type": "integer"},
            "Bonus %": {"type": "text"},
            "Country": {"type": "text"},
            "City": {"type": "text"},
            "Exit Date": {"type": "date"}
        }
    }
}
if not es.indices.exists(index=index_name):
    es.indices.create(index=index_name, body=index_settings)
    print(f"Index '{index_name}' created")

csv_file_path = 'C:\\Users\\nilof\\Downloads\\archive\\employee.csv'
```

```
if os.path.exists(csv_file_path):
    print("File found, attempting to load...")
    try:
        df = pd.read_csv(csv_file_path, encoding='latin1')
        print("CSV file loaded successfully.")
    except UnicodeDecodeError as e:
        print(f"Error reading CSV file due to encoding issue: {e}")
else:
    print(f"File not found at {csv_file_path}")
    exit()

def convert_to_datetime(date_str):
    if isinstance(date_str, str):
        try:
            return datetime.strptime(date_str, '%m/%d/%Y').isoformat()
        except ValueError:
            print(f"Error processing date: {date_str}")
            return None
    else:
        return None

if 'Hire Date' in df.columns:
    df['Hire Date'] = df['Hire Date'].apply(convert_to_datetime)

if 'Exit Date' in df.columns:
    df['Exit Date'] = df['Exit Date'].apply(convert_to_datetime)

def clean_salary(salary_str):
    if isinstance(salary_str, str):
        try:
            return int(salary_str.replace('$', '').replace(',', '').strip())
        except ValueError:
            print(f"Error processing salary: {salary_str}")
            return None
    return None

if 'Annual Salary' in df.columns:
    df['Annual Salary'] = df['Annual Salary'].apply(clean_salary)
```

```
def generate_data(df):
    for index, row in df.iterrows():
        try:
            yield {
                "_index": index_name,
                "_source": {
                    "Employee ID": str(row["Employee ID"]) if pd.notnull(row["Employee ID"]) else None,
                    "Full Name": str(row["Full Name"]),
                    "Job Title": str(row["Job Title"]),
                    "Department": str(row["Department"]),
                    "Business Unit": str(row["Business Unit"]),
                    "Gender": str(row["Gender"]),
                    "Ethnicity": str(row["Ethnicity"]),
                    "Age": int(row["Age"]) if pd.notnull(row["Age"]) else None,
                    "Hire Date": row["Hire Date"],
                    "Annual Salary": row["Annual Salary"],
                    "Bonus %": str(row["Bonus %"]) if pd.notnull(row["Bonus %"]) else None,
                    "Country": str(row["Country"]),
                    "City": str(row["City"]),
                    "Exit Date": row["Exit Date"]
                }
            }
        except Exception as e:
            print(f"Error processing row {index}: {e}")

try:
    response = helpers.bulk(es, generate_data(df), raise_on_error=False, stats_only=False)

    if response[0] > 0:
        print(f"Successfully indexed {response[0]} documents.")
    if response[1] > 0:
        print(f"{response[1]} document(s) failed to index.")
except Exception as e:
    print(f"Error indexing data: {str(e)}")
```

EXECUTION

```
(c) Microsoft Corporation. All rights reserved.

C:\Users\nilof\Downloads>n.py
File found, attempting to load...
CSV file loaded successfully.
```

OUTPUT

```
C:\Users\nilof\Downloads>n.py
File found, attempting to load...
CSV file loaded successfully.
Error indexing data: 500 document(s) failed to index.

C:\Users\nilof\Downloads>curl -u elastic:elastic -X GET "http://localhost:9200/my_csv_index/_search?pretty"
{
  "took" : 690,
  "timed_out" : false,
  "_shards" : {
    "total" : 1,
    "successful" : 1,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : {
      "value" : 13,
      "relation" : "eq"
    },
    "max_score" : 1.0,
    "hits" : [
      {
        "_index" : "my_csv_index",
        "_id" : "Qj4MTpIBVLfX5wZA6ilJ",
        "_score" : 1.0,
        "_source" : {
          "Employee ID" : null,
          "Full Name" : "nan",
          "Job Title" : "Manager",
          "Department" : "IT",
          "Business Unit" : "Specialty Products",
          "Gender" : "Female",
          "Ethnicity" : "Latino",
          "Age" : 64,
          "Hire Date" : null,
          "Annual Salary" : null,
          "Bonus %" : null,
          "Country" : "United States",
          "City" : "Miami",
          "Exit Date" : null
        }
      },
      {
        "_index" : "my_csv_index",
        "_id" : "Qz4MTpIBVLfX5wZA6ilJ",
```

```
    "_score" : 1.0,
    "_source" : {
      "Employee ID" : null,
      "Full Name" : "nan",
      "Job Title" : "Director",
      "Department" : "Human Resources",
      "Business Unit" : "Specialty Products",
      "Gender" : "Female",
      "Ethnicity" : "Caucasian",
      "Age" : 53,
      "Hire Date" : null,
      "Annual Salary" : null,
      "Bonus %" : null,
      "Country" : "United States",
      "City" : "Miami",
      "Exit Date" : null
    }
  },
  {
    "_index" : "my_csv_index",
    "_id" : "RD4MTpIBVLfX5wZA6ilJ",
    "_score" : 1.0,
    "_source" : {
      "Employee ID" : null,
      "Full Name" : "Logan Rahman",
      "Job Title" : "Manager",
      "Department" : "nan",
      "Business Unit" : "Specialty Products",
      "Gender" : "Male",
      "Ethnicity" : "Asian",
      "Age" : 39,
      "Hire Date" : null,
      "Annual Salary" : null,
      "Bonus %" : null,
      "Country" : "United States",
      "City" : "Austin",
      "Exit Date" : null
    }
  },
  {
    "_index" : "my_csv_index",
    "_id" : "Rj4RTpIBVLfX5wZAZymT",
    "_score" : 1.0,
    "_source" : {
      "Employee ID" : null,
      "Full Name" : "nan",
      "Job Title" : "Manager",
      "Department" : "IT",
      "Business Unit" : "Specialty Products",
```

```
    },
    {
      "_index" : "my_csv_index",
      "_id" : "Rz4RTpIBVLfX5wZAZymT",
      "_score" : 1.0,
      "_source" : {
        "Employee ID" : null,
        "Full Name" : "nan",
        "Job Title" : "Director",
        "Department" : "Human Resources",
        "Business Unit" : "Specialty Products",
        "Gender" : "Female",
        "Ethnicity" : "Caucasian",
        "Age" : 53,
        "Hire Date" : null,
        "Annual Salary" : null,
        "Bonus %" : null,
        "Country" : "United States",
        "City" : "Miami",
        "Exit Date" : null
      }
    },
    {
      "_index" : "my_csv_index",
      "_id" : "SD4RTpIBVLfX5wZAZymT",
      "_score" : 1.0,
      "_source" : {
        "Employee ID" : null,
        "Full Name" : "Logan Rahman",
        "Job Title" : "Manager",
        "Department" : "nan",
        "Business Unit" : "Specialty Products",
        "Gender" : "Male",
        "Ethnicity" : "Asian",
        "Age" : 39,
        "Hire Date" : null,
        "Annual Salary" : null,
        "Bonus %" : null,
        "Country" : "United States",
        "City" : "Austin",
        "Exit Date" : null
      }
    },
    {
      "_index" : "my_csv_index",
      "_id" : "Tz4TTpIBVLfX5wZA7ymD",
      "_score" : 1.0,
      "_source" : {
```

```
},
{
  "_index" : "my_csv_index",
  "_id" : "UD4TTpIBVLfX5wZA7ymD",
  "_score" : 1.0,
  "_source" : {
    "Employee ID" : null,
    "Full Name" : "nan",
    "Job Title" : "Director",
    "Department" : "Human Resources",
    "Business Unit" : "Specialty Products",
    "Gender" : "Female",
    "Ethnicity" : "Caucasian",
    "Age" : 53,
    "Hire Date" : null,
    "Annual Salary" : null,
    "Bonus %" : null,
    "Country" : "United States",
    "City" : "Miami",
    "Exit Date" : null
  }
},
{
  "_index" : "my_csv_index",
  "_id" : "Uj4TTpIBVLfX5wZA7ymD",
  "_score" : 1.0,
  "_source" : {
    "Employee ID" : null,
    "Full Name" : "Logan Rahman",
    "Job Title" : "Manager",
    "Department" : "nan",
    "Business Unit" : "Specialty Products",
    "Gender" : "Male",
    "Ethnicity" : "Asian",
    "Age" : 39,
    "Hire Date" : null,
    "Annual Salary" : null,
    "Bonus %" : null,
    "Country" : "United States",
    "City" : "Austin",
    "Exit Date" : null
  }
},
{
  "_index" : "my_csv_index",
  "_id" : "dD4XTpIBVLfX5wZA_Snb",
  "_score" : 1.0,
  "_source" : {
    "Employee ID" : null,
```

```
        "Gender" : "Male",
        "Ethnicity" : "Asian",
        "Age" : 39,
        "Hire Date" : null,
        "Annual Salary" : null,
        "Bonus %" : null,
        "Country" : "United States",
        "City" : "Austin",
        "Exit Date" : null
      }
    },
    {
      "_index" : "my_csv_index",
      "_id" : "dD4XTpIBVLfX5wZA_Snb",
      "_score" : 1.0,
      "_source" : {
        "Employee ID" : null,
        "Full Name" : "nan",
        "Job Title" : "Manager",
        "Department" : "IT",
        "Business Unit" : "Specialty Products",
        "Gender" : "Female",
        "Ethnicity" : "Latino",
        "Age" : 64,
        "Hire Date" : null,
        "Annual Salary" : null,
        "Bonus %" : null,
        "Country" : "United States",
        "City" : "Miami",
        "Exit Date" : null
      }
    }
  ]
 }
}
```