Software Development
Home Assignment

Please solve the test using one of the following programming languages: Python, C++, or Java, and provide the files as a ZIP.

Objective

Implement an ETL (Extract, Transform, Load) pipeline in Python that processes JSON data from three different sources: req, application, and user. The ETL process should read the data, perform necessary transformations, and load the processed data into a MongoDB collection.

Requirements

1. Extract data from JSON files.
2. Transform the data to meet the specified schema requirements.
3. Load the transformed data into a MongoDB collection.

Deliverables

1. Python scripts for the ETL process.
2. Sample JSON files (req.json, application.json, user.json).
3. Instructions on how to run the ETL process.

```
[
    {
        "req_id": "REQ001",
        "title": "Software Engineer",
        "description": "Develop and maintain software applications.",
        "location": "New York, NY",
        "department": "Engineering",
        "posted_date": "2024-05-01"
    },
    {
        "req_id": "REQ002",
        "title": "Data Analyst",
        "description": "Analyze data to support business decisions.",
```


```
        "location": "San Francisco, CA",
        "department": "Data Science",
        "posted_date": "2024-05-15"
    }
]
```

```
[
    {
        "application_id": "APP001",
        "req_id": "REQ001",
        "candidate_name": "John Doe",
        "status": "Applied",
        "applied_date": "2024-05-02",
        "resume": "URL_TO_RESUME"
    },
    {
        "application_id": "APP002",
        "req_id": "REQ002",
        "candidate_name": "Jane Smith",
        "status": "Interviewing",
        "applied_date": "2024-05-16",
        "resume": "URL_TO_RESUME"
    }
]
```

```
[
    {
        "user_id": "USR001",
        "name": "John Doe",
        "email": "john.doe@example.com",
        "role": "Candidate",
        "created_at": "2024-04-01"
    },
    {
        "user_id": "USR002",
        "name": "Jane Smith",
        "email": "jane.smith@example.com",
        "role": "Candidate",
        "created_at": "2024-04-15"
    }
```

]

Deliverables

1. Python scripts for the ETL process.
2. Sample JSON files (req.json, application.json, user.json).
3. Instructions on how to run the ETL process.

ETL Process

1. Extract: Read the JSON data from the files.
2. Transform:
    - Join the req and application data based on req_id .
    - Merge user data based on candidate_name and update email and role.
    - Ensure all dates are in ISO format.
3. Load: Insert the transformed data into a MongoDB collection named etl_output

Tasks

1. Setup MongoDB:
    - Ensure MongoDB is running locally or provide a connection string to a MongoDB instance.
2. Create the ETL Script:
    - Create a Python script named etl.py .
    - The script should perform the following steps:
        A. Read JSON data from req.json , application.json , and user.json .
        B. Transform the data as described above.
        C. Insert the transformed data into a MongoDB collection named etl_output .
3. Submission:
    - Provide the etl.py script.
    - Provide the JSON files: req.json , application.json , and user.json .

Provide a README.md with instructions on how to run the script.