

ETL Pipeline : Extract : Transform : Load. —————> The Architecture

DependencyCreator

Google Review

Web Scrapper 1

Web Scrapper 2

Web Scrapper 3

CSV.

CSV.

CSV.

S3
Cloud
Object
Storage

ID	Status	Filename	Folder
1	NEW	Rest 1 .01	Rest 1
2	NEW	Rest 2 .02	Rest 2
3	NEW	Rest 1 .02	Rest 1

Data Ingestor 1

Data Ingestor 2

Data Ingestor 3

MySQL

Extract

Transform

Create New Work Item

Add CSV Files

Ingest

- ① Look for the latest job
- ② Go to Folder → filename
- ③ Ingest data to MySQL DB
- ④ Delete CSV from S3
- ⑤ Mark Status as Done

Data Analyst: SQL in Python

```
def get_customers(rest):
```

```
    db_handle = "SELECT u.FIRST-NAME,  
                  u.LAST-NAME
```

```
                FROM USER u
```

```
                RIGHT JOIN j ON user_id"
```

```
    return db_handle.
```

```
Customer = get_customers(rest-01)
```

```
Customer-df = pd.DataFrame(customer) SQL → DF
```

① Cleaning

② Analysis

③ Clustered Review Results.

Cluster Review By
Sentiment API

Export to CSV

To Google Cloud?

Matplotlib
Visualization

Tableau / Power BI

Load

1. Dependency creator = Jessie
2. Web scraper = Kevin
3. Data ingester = Joseph
4. Data analyser