**Project 1: Batch Time Analysis of Transactional Data**

**Objective:** To use AWS Big Data stack for data engineering to analyze transactions, uncover patterns, and share actionable insights.

**Steps to perform:**

* + S3 > Create Bucket > Name: [batch-time-analysis-of-transaction-data](https://s3.console.aws.amazon.com/s3/buckets/batch-time-analysis-of-transaction-data?region=us-east-1):
    - Create folder: input
    - Create folder: output
    - Upload CSV data into: s3://batch-time-analysis-of-transaction-data/input/
  + AWS Glue > Crawlers > Add Crawler:
    - Name: input-crawler
    - Data Stores Path: s3://batch-time-analysis-of-transaction-data/Products\_dataset.csv
    - IAM Role: AWSGlueServiceRole-Admin
    - Schedule: Run on Demand
    - Add Database: input-database
  + Run Crawler: input-crawler
  + AWS Glue> Databases:
    - Create Database: output-database
  + Create Job > Source S3 to Target S3:
    - Job Name: csv-to-parquet
    - Data Source:
      * S3 Location URL: s3://batch-time-analysis-of-transaction-data/input/Products\_dataset.csv
      * Data Format: CSV
      * Delimiter: Tab
      * Optimize CSV: True
    - ApplyMapping:
      * Adjust data types
      * Drop unwanted columns
    - Data Target
      * S3 Location URL: s3://batch-time-analysis-of-transaction-data/output/
      * Data Format: Parquet
      * Compression Type: Snappy
    - Save Job > Run Job > Run Details
  + AWS Glue > Crawlers > Add Crawler
    - Name: parquet-crawler
    - Data Stores Path: s3://batch-time-analysis-of-transaction-data/output/Products\_dataset.parquet
    - IAM Role: AWSGlueServiceRole-Admin
    - Schedule: Run on Demand
    - Database: output-database
    - Run Crawler: parquet-crawler

SELECT \* FROM "input-database"."products\_dataset\_csv" limit 10;

SELECT country AS Country, quantity AS Quantity

FROM "input-database"."products\_dataset\_parquet"

GROUP BY quantity DESC

limit 2;

ANSWER:  
United Kingdom - PAPEL CRAFT, LITTLE BIRDIE - 80995

United Kingdom - MEDIUM CERAMIC TOP STORAGE JAR - 74215