ETL Exercise: Snowflake + Talend + GitHub

Objective

You will build an ETL pipeline that:

1. Creates a database, schema, and table in Snowflake
2. Reads customer data from a CSV
3. Transforms it in Talend (adds full_name)
4. Loads the data into Snowflake
5. Pushes your ETL job files to GitHub

Step 1: Create Database, Schema, and Table in Snowflake

Run this SQL in Snowflake Web UI:

```
CREATE OR REPLACE DATABASE customer_db;
CREATE OR REPLACE SCHEMA customer_schema;
USE DATABASE customer_db;
USE SCHEMA customer_schema;

CREATE OR REPLACE TABLE customer_schema.customers (
    cust_id INTEGER,
    first_name STRING,
    last_name STRING,
    email STRING,
    city STRING,
    country STRING,
    full_name STRING
);
```

Step 2: Prepare customers.csv

Create a file named 'customers.csv' with this content:

```
cust_id,first_name,last_name,email,city,country
101,John,Doe,john.doe@example.com,New York,USA
102,Jane,Smith,jane.smith@example.com,Los Angeles,USA
```

103,Alice,Williams,alice.w@example.com,Toronto,Canada

104,Bob,Brown,bob.brown@example.com,London,UK

Step 3: Talend ETL Job - LoadCustomersToSnowflake

Use the following components:

tFileInputDelimited -> tMap -> tSnowflakeOutput

Configure tFileInputDelimited

1. Drag tFileInputDelimited to the canvas.
2. File Name: Browse to 'customers.csv'
3. Field Separator: ,
4. Header: 1
5. Edit Schema and add columns:
   - cust_id (Integer)
   - first_name (String)
   - last_name (String)
   - email (String)
   - city (String)
   - country (String)

Configure tMap

1. Drag tMap and connect it to tFileInputDelimited.
2. Double-click tMap.
3. Add a new output column: full_name (String)
4. Expression: row1.first_name + " " + row1.last_name
5. Map all other fields from input to output.
6. Click OK.

Configure tSnowflakeOutput

1. Drag tSnowflakeOutput and connect from tMap.
2. Use existing Snowflake connection.
3. Schema: customer_schema
4. Table: customers
5. Action on table: Create if not exists
6. Action on data: Insert

7. Edit Schema to include all fields including full_name.

Run and Verify

1. Click Run tab -> Run the job
2. In Snowflake, run:

```
SELECT * FROM customer_schema.customers;
```

Push to GitHub

1. Export the Talend job
2. Run these commands:

```
git init
git remote add origin https://github.com/YOUR_USERNAME/customer-etl.git
git add .
git commit -m "ETL Job: Load customer data into Snowflake"
git push -u origin main
```