

Assignment 3

Screenshots:

The screenshot displays the Apache Spoon IDE interface. The main window shows a transformation job design with three components: 'customer source', 'Sort rows', and 'dim_customer', connected in a linear sequence. The 'dim_customer' component is highlighted, showing its DDL script in a yellow background. The script includes a DROP statement and a CREATE TABLE statement for 'dim_customer' with various fields and an index. Below the design, the 'Execution Results' tab is active, showing a log of the transformation execution. The log indicates that the transformation started at 2021/10/25 05:56:24, finished at 2021/10/25 05:56:24, and ended at 2021/10/25 05:59:14.

```
/*
 * DDL Script
 */

DROP TABLE IF EXISTS dim_customer;

/* DDL script for dim_customer */
CREATE TABLE dim_customer
(
  customer_skey BIGINT AUTO_INCREMENT NOT NULL PRIMARY KEY
, customer_id VARCHAR(50)
, first_name VARCHAR(50)
, last_name VARCHAR(50)
, date_of_birth DATE
, city VARCHAR(50)
, state VARCHAR(50)
, version INT
, date_from DATE
, date_to DATET
);

CREATE INDEX idx_dim_customer_lookup ON dim_customer(customer_id);
```

Execution Results

Logging | Execution History | Step Metrics | Performance Graph | Metrics | Preview data

2021/10/25 05:56:24 - Spoon - Started the transformation execution.

2021/10/25 05:56:24 - slowly-changing-dimension - Dispatching started for transformation [slowly-changing-dimension]

2021/10/25 05:56:24 - Spoon - The transformation has finished!!

2021/10/25 05:59:14 - Spoon - Starting job...

2021/10/25 05:59:14 - Spoon - Job has ended.

SCD Type 1						
Before						
customer_id	first_name	last_name	date_of_birth	city	state	effective_Date
C1197	Panpan	Bressler	30-01-1992	PHILADELPHIA	Pennsylvania	01-01-2015
After						
customer_id	first_name	last_name	date_of_birth	city	state	effective_Date
C1050	BethAnn	Cox	03-01-1993	New York	New York	01-01-2015

Application of SCD type1 is applied to the first name, last name, city, state and effective date as the type1 is direct replacing of the old value, the last value in the csv is updated for each record of the customer C1050.

Spoon - scd-main

File Edit View Action Tools Help

Connect

View Design

Search

Jobs

- scd-main
 - Run configurations
 - Database connections
 - Job entries
 - Slave server
 - Hadoop clusters
 - VFS Connections

START → truncate table → scd

Execution Results

Logging History Job metrics Metrics

2021/10/25 05:59:14 - Spoon - Starting job...

2021/10/25 05:59:14 - scd-main - Start of job execution

2021/10/25 05:59:14 - scd-main - Starting entry [truncate table]

2021/10/25 05:59:14 - scd-main - Starting entry [scd]

2021/10/25 05:59:14 - scd - Using run configuration [Pentaho local]

2021/10/25 05:59:14 - scd - Using legacy execution engine

2021/10/25 05:59:14 - slowly-changing-dimension - Dispatching started for transformation [slowly-changing-dimension]

2021/10/25 05:59:14 - scd-main - Finished job entry [scd] (result=[true])

2021/10/25 05:59:14 - scd-main - Finished job entry [truncate table] (result=[true])

2021/10/25 05:59:14 - scd-main - Job execution finished

2021/10/25 05:59:14 - Spoon - Job has ended.