-- Customer raw Table

```
CREATE EXTERNAL TABLE `stg_customers_raw`(
 `customer_id` string ,
 `first` string ,
 `last` string ,
 `age` string ,
 `country` string )
ROW FORMAT SERDE
 'org.apache.hadoop.hive.serde2.OpenCSVSerde'
WITH SERDEPROPERTIES (
 'escapeChar'='\\',
 'quoteChar'='\"',
 'separatorChar'=',')
STORED AS INPUTFORMAT
 'org.apache.hadoop.mapred.TextInputFormat'
OUTPUTFORMAT
 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat'
LOCATION
 's3://analytics-development/test_data/customer'
```

-- Orders raw table

```
CREATE EXTERNAL TABLE `stg_orders_raw`(
 `order_id` string ,
 `item` string ,
 `amount` string ,
 `customer_id` string )
ROW FORMAT SERDE
 'org.apache.hadoop.hive.serde2.OpenCSVSerde'
WITH SERDEPROPERTIES (
 'escapeChar'='\\',
 'quoteChar'='\"',
 'separatorChar'=',')
STORED AS INPUTFORMAT
 'org.apache.hadoop.mapred.TextInputFormat'
OUTPUTFORMAT
 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat'
LOCATION
 's3://analytics-development/test_data/order'
```

-- Shipping raw table

```
CREATE EXTERNAL TABLE `stg_shipping_raw`(
 `shipping_id` bigint,
 `status` string ,
```

```sql
  `customer_id` bigint)
ROW FORMAT SERDE
  'org.openx.data.jsonserde.JsonSerDe'
STORED AS INPUTFORMAT
  'org.apache.hadoop.mapred.TextInputFormat'
OUTPUTFORMAT
  'org.apache.hadoop.hive.ql.io.IgnoreKeyTextOutputFormat'
LOCATION
  's3://analytics-development/test_data/shipping'


-- Customer Dimension Table
CREATE TABLE customer_dim
WITH (
  format = 'PARQUET',
  external_location = 's3://analytics-development/test_data/customer_dim/',
  parquet_compression = 'SNAPPY'
) AS
SELECT
  CAST(Customer_ID AS BIGINT)      AS customer_id,
  trim(First)              AS first_name,
  trim(Last)               AS last_name,
  TRY_CAST(Age AS INTEGER)        AS age,
  trim(Country)              AS country
FROM stg_customers_raw
WHERE COALESCE(trim(Customer_ID),'') <> '';

-- Order Item Dimension Table
CREATE TABLE order_dim
WITH (
  format = 'PARQUET',
  external_location = 's3://analytics-development/test_data/order_dim/',
  parquet_compression = 'SNAPPY'
) AS
SELECT
  CAST(order_id AS BIGINT)         AS order_id,
  trim(item)                AS item,
  TRY_CAST(amount AS DECIMAL(18,2))     AS amount,
  CAST(customer_id AS BIGINT)        AS customer_id
FROM stg_orders_raw
WHERE COALESCE(trim(order_id),'') <> '';

-- Shipping Dimension Table
CREATE TABLE shipping_dim
WITH (
  format = 'PARQUET',
  external_location = 's3://analytics-development/test_data/shipping_dim/',
  parquet_compression = 'SNAPPY'
```

```
) AS
SELECT
  CAST(Shipping_ID AS VARCHAR)      AS shipping_id,
  trim(Status)                      AS shipping_status,
  CAST(Customer_ID AS BIGINT)       AS customer_id
FROM stg_shipping_raw;
```

-- Order Fact Table
```
CREATE TABLE order_fact
WITH (
  format = 'PARQUET',
  external_location = 's3://analytics-development/test_data/order_fact/',
  parquet_compression = 'SNAPPY'
) AS
SELECT
  order_id,
  MAX(customer_id)              AS customer_id,
  COUNT(*)                      AS items_count,
  SUM(amount)                   AS order_total,
  CASE WHEN MAX(customer_id) IS NULL THEN true ELSE false END AS
is_customer_missing
FROM order_dim
GROUP BY order_id;
```