

# Designing ETL Patterns for Data Warehousing

---



**Stacia Varga**

CONSULTANT, INSTRUCTOR, AUTHOR

@\_StaciaV\_ [datainspirations.com](http://datainspirations.com)



# Overview



**Package types**

**Template package**

**Staging table loads**

**Dimension table loads**

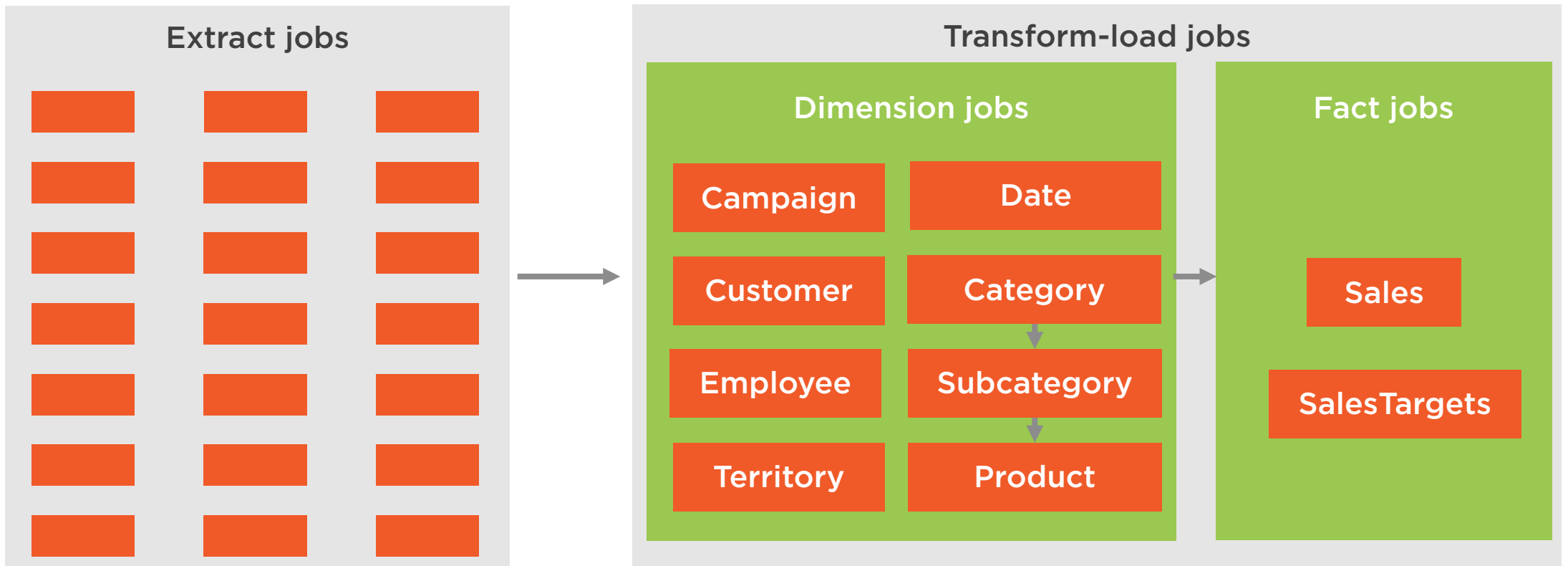
**Fact table loads**

**Parent packages**



# Package Types

## Data warehouse jobs



# Package Types

## Parent package

Auditing and logging mechanism

Execute package tasks only

Execute SQL tasks for snapshots in top-level parent package only

## Leaf-level child package

Auditing and logging mechanism

Miscellaneous tasks + Data Flow Task

Job-level details from ETL requirements



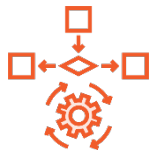
# Template Package



**Configuration components**



**Auditing and logging mechanism**



**Business logic**



**Connection managers**



**Variables and parameters**



# Configuration Components

Source database  
connection  
string(s)

Staging database  
connection string

Target database  
connection string

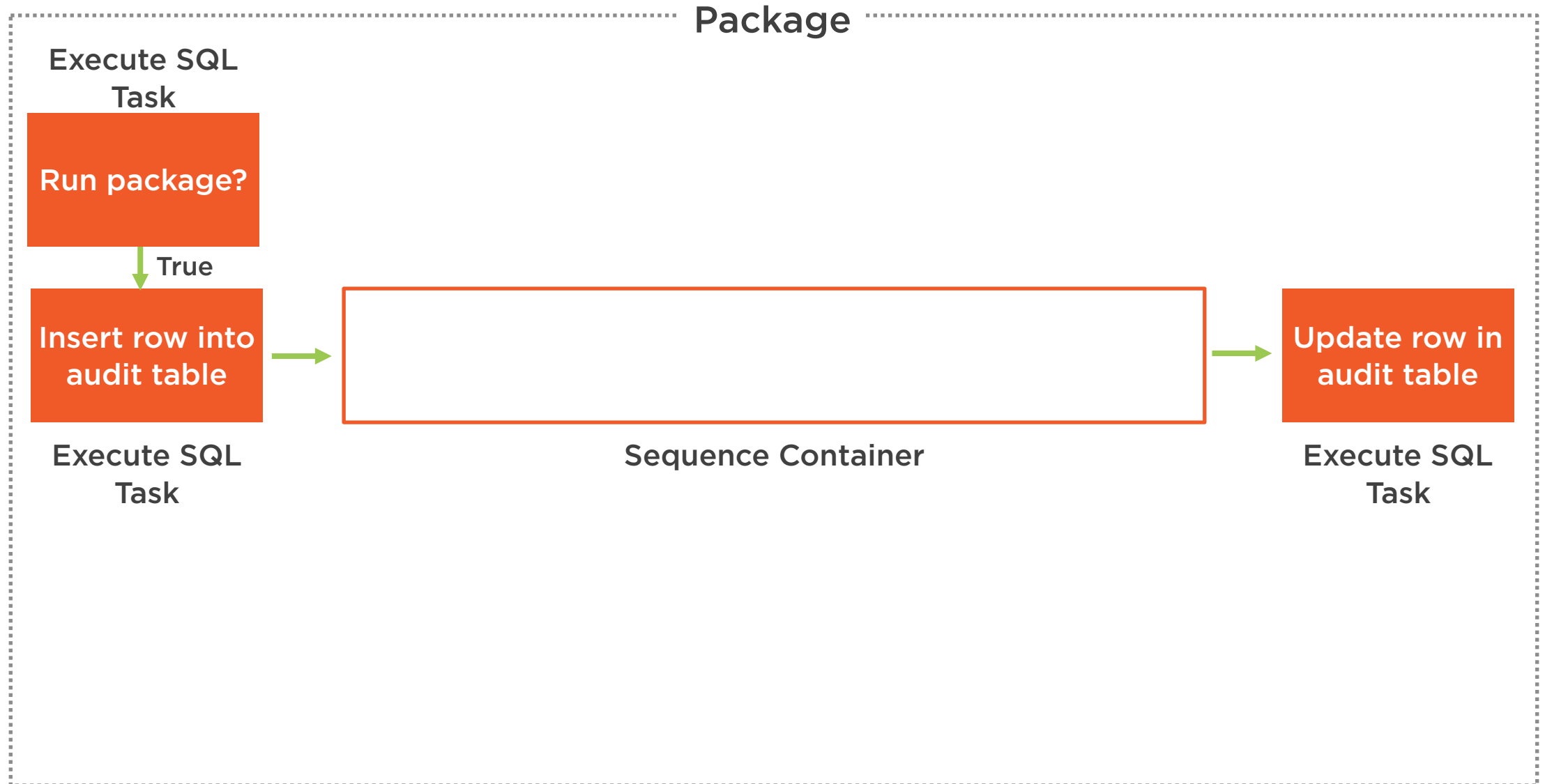
Target  
database name

SQL Server file  
path

Error file path



# Template Package



# Connection Managers



**Log database**

**Source database(s)**

**Staging database**

**Target database(s)**



# Variables



**AuditKey**

**RowCountErrors**

**RowCountFinal**

**RowCountInitial**

**RowCountInserts**

**RowCountUpdates**

**RunPackage**

**SourceQuery**

**SourceQueryWhere**

**TableMaxDateTime**

**TableName**



# Parameters



## Project

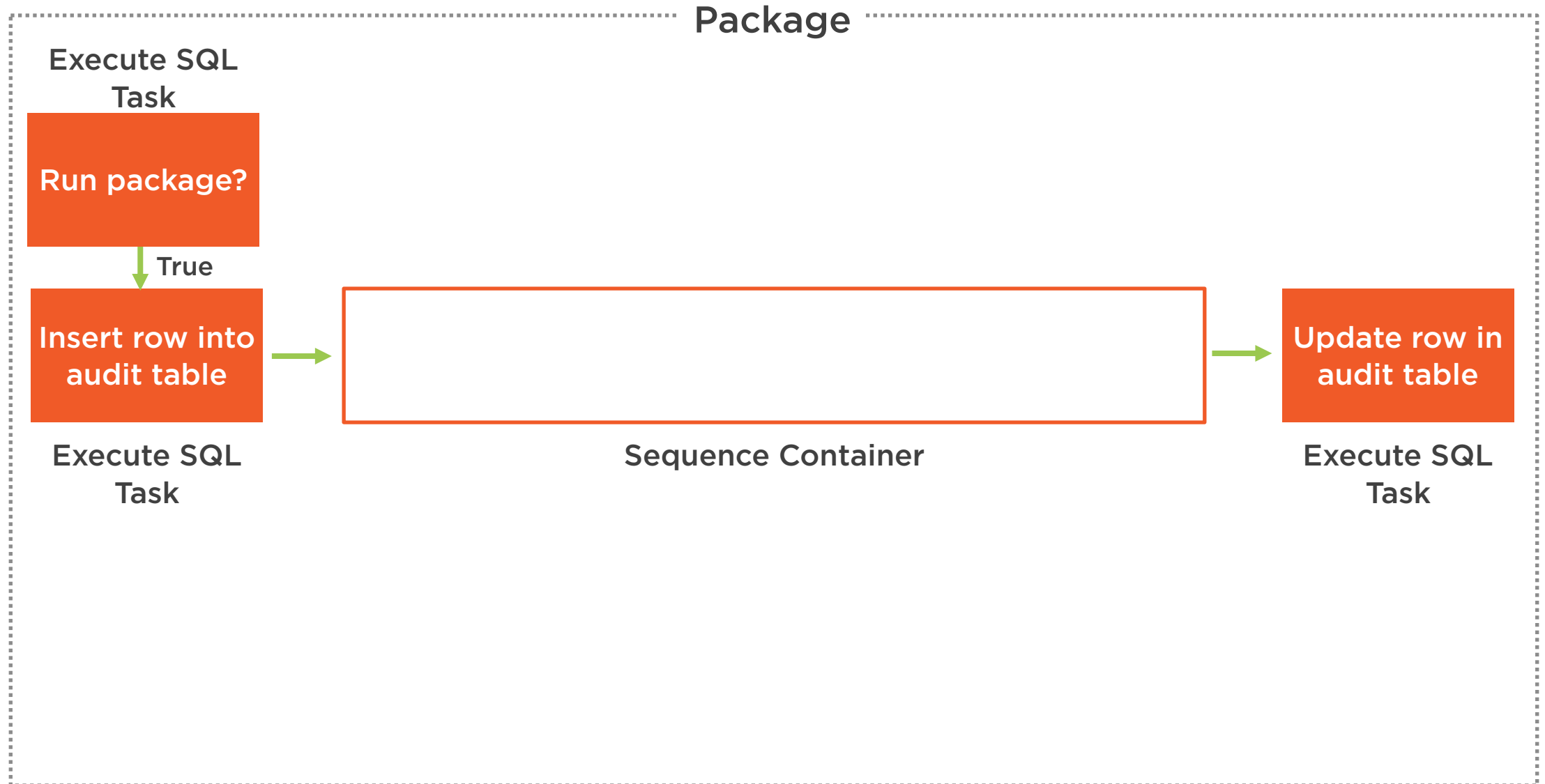
- LoadType

## Package

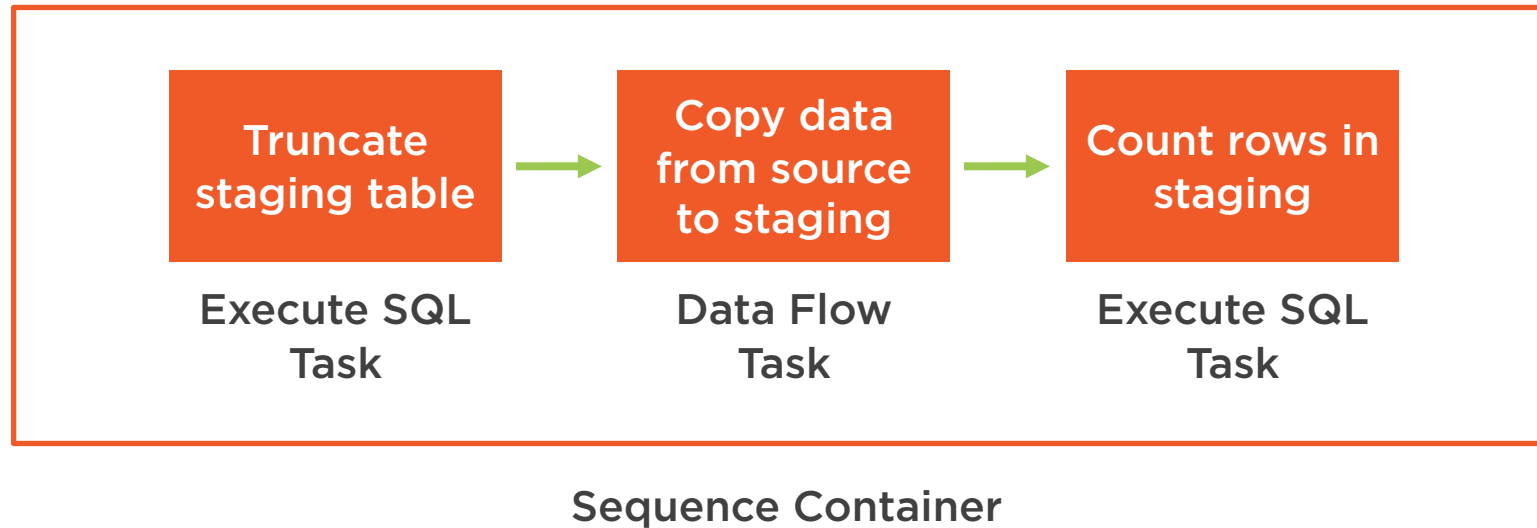
- ParentAuditKey



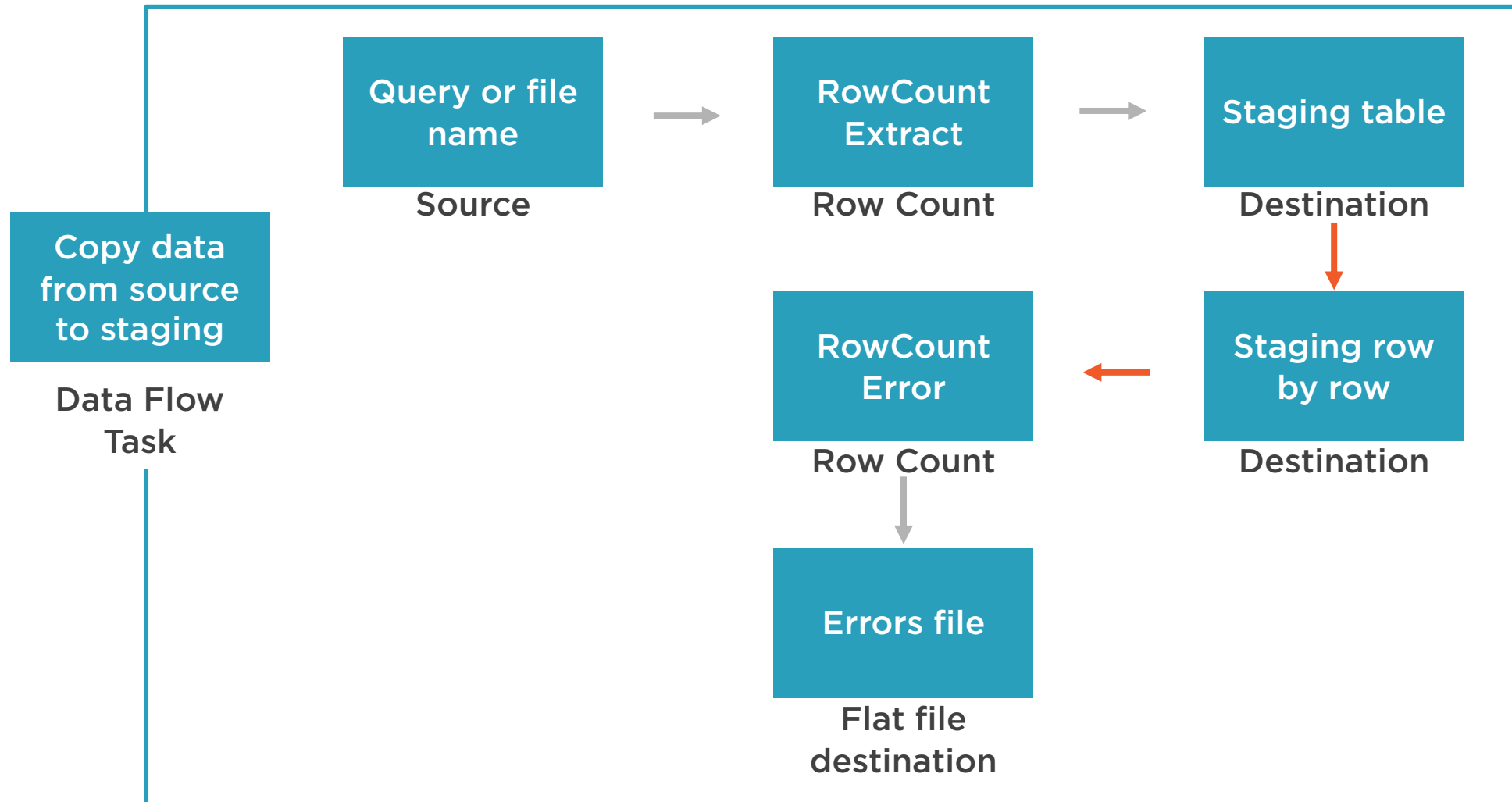
# Staging Loads



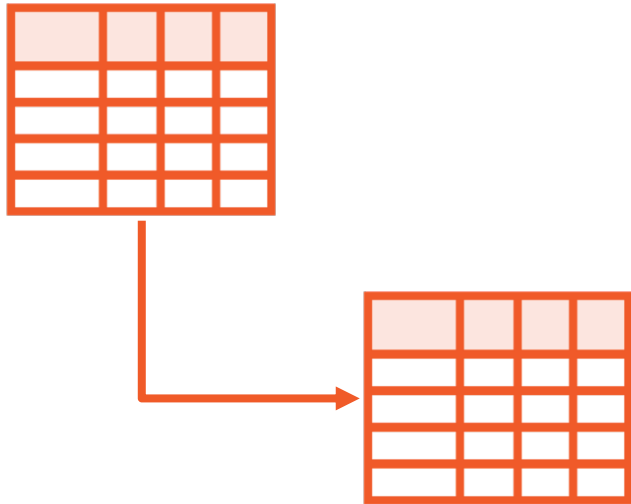
# Staging Loads



# Staging Loads

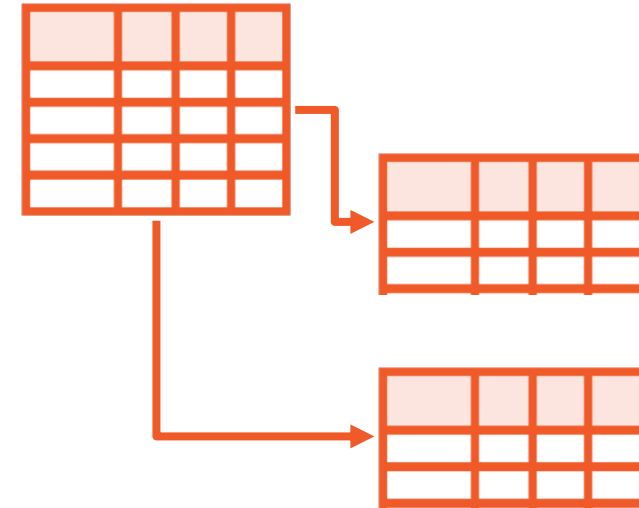


# Full Versus Incremental Extracts



## Full extract

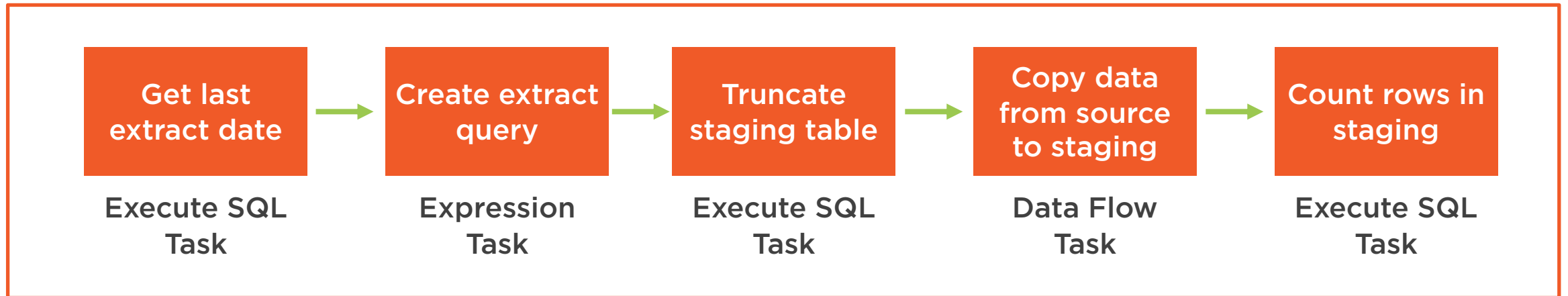
Applies to first extract of all tables or to regular extracts of dimensions



## Incremental load

Applies to sources for fact tables and large dimensions

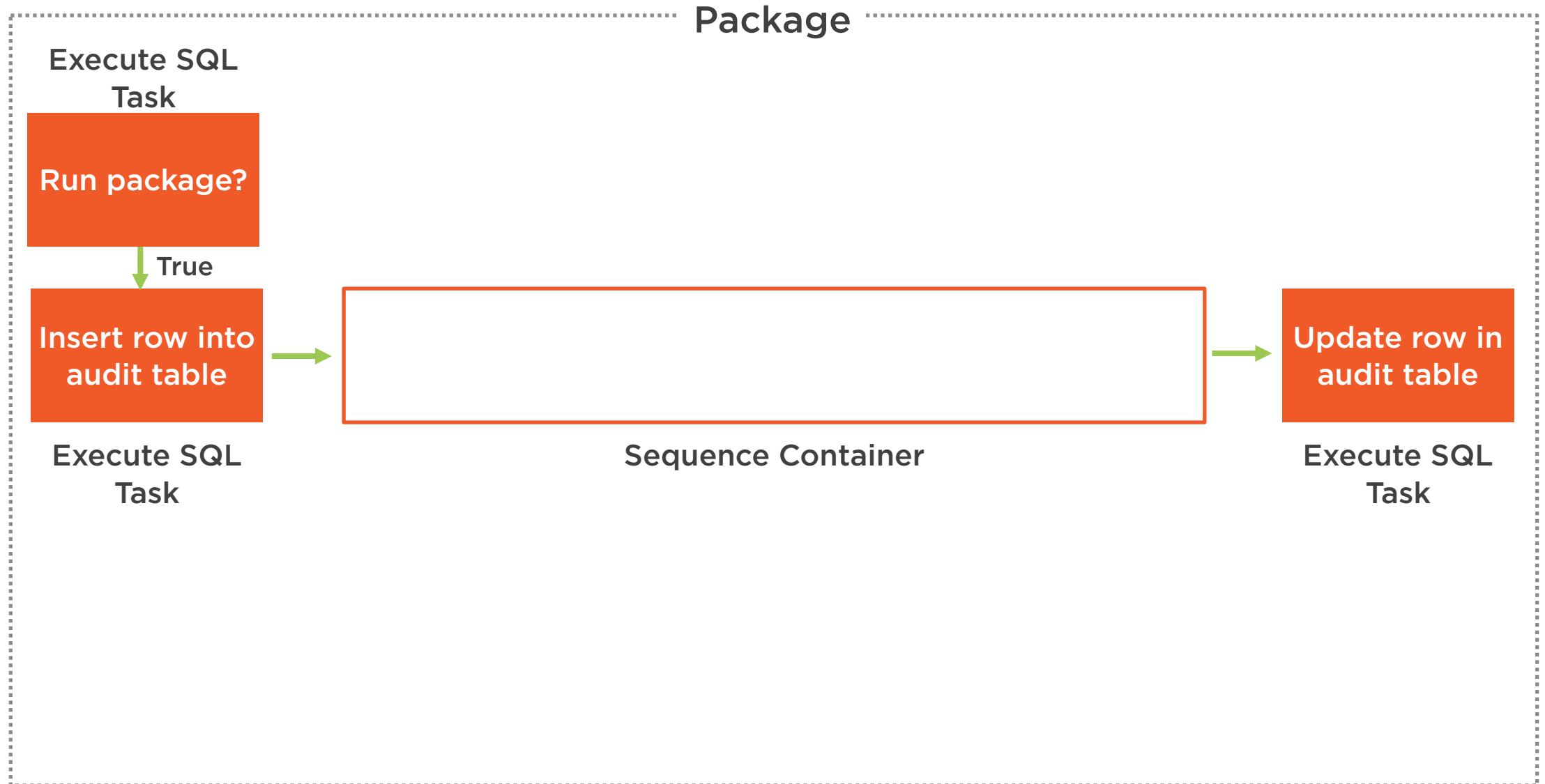
# Full Versus Incremental Extracts



Sequence Container

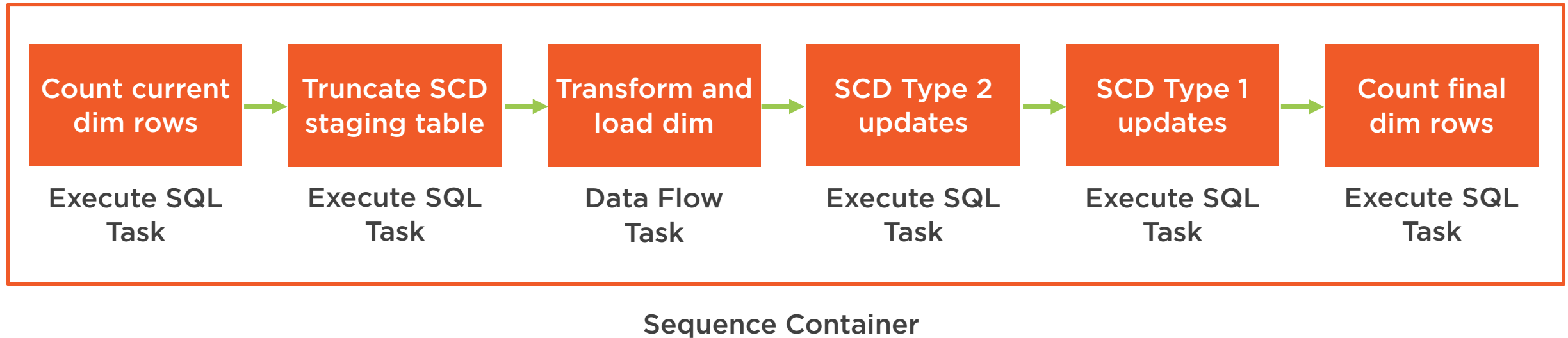


# Dimension Loads

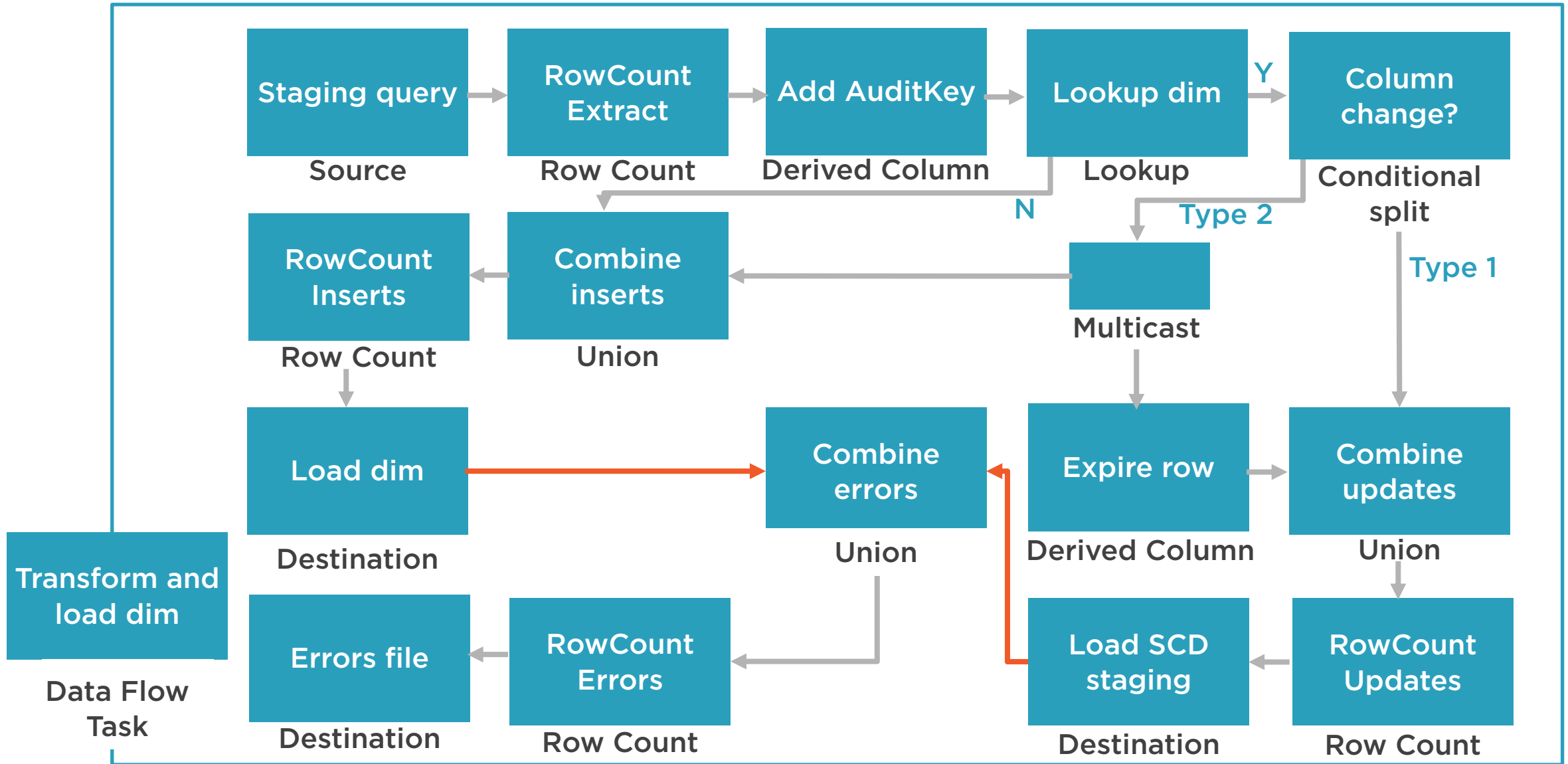




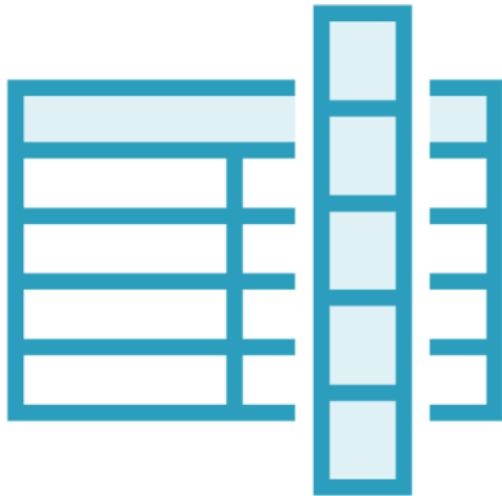
# Dimension Loads



# Dimension Loads

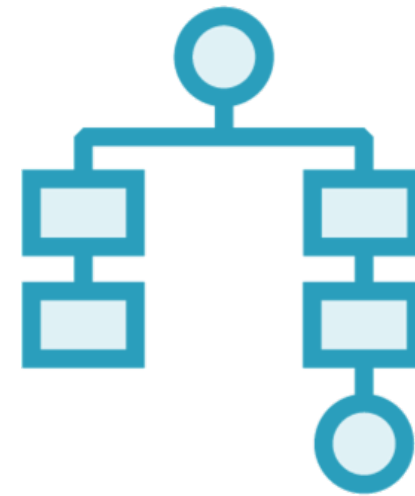


# Error Handling



## Column errors

Use derived column or data conversion transformations (SSIS) or source query expressions (SQL)



## Structure errors

Use lookup transformation to test for potential failure and add steps to correct

# SQL or SSIS Transforms

## SQL

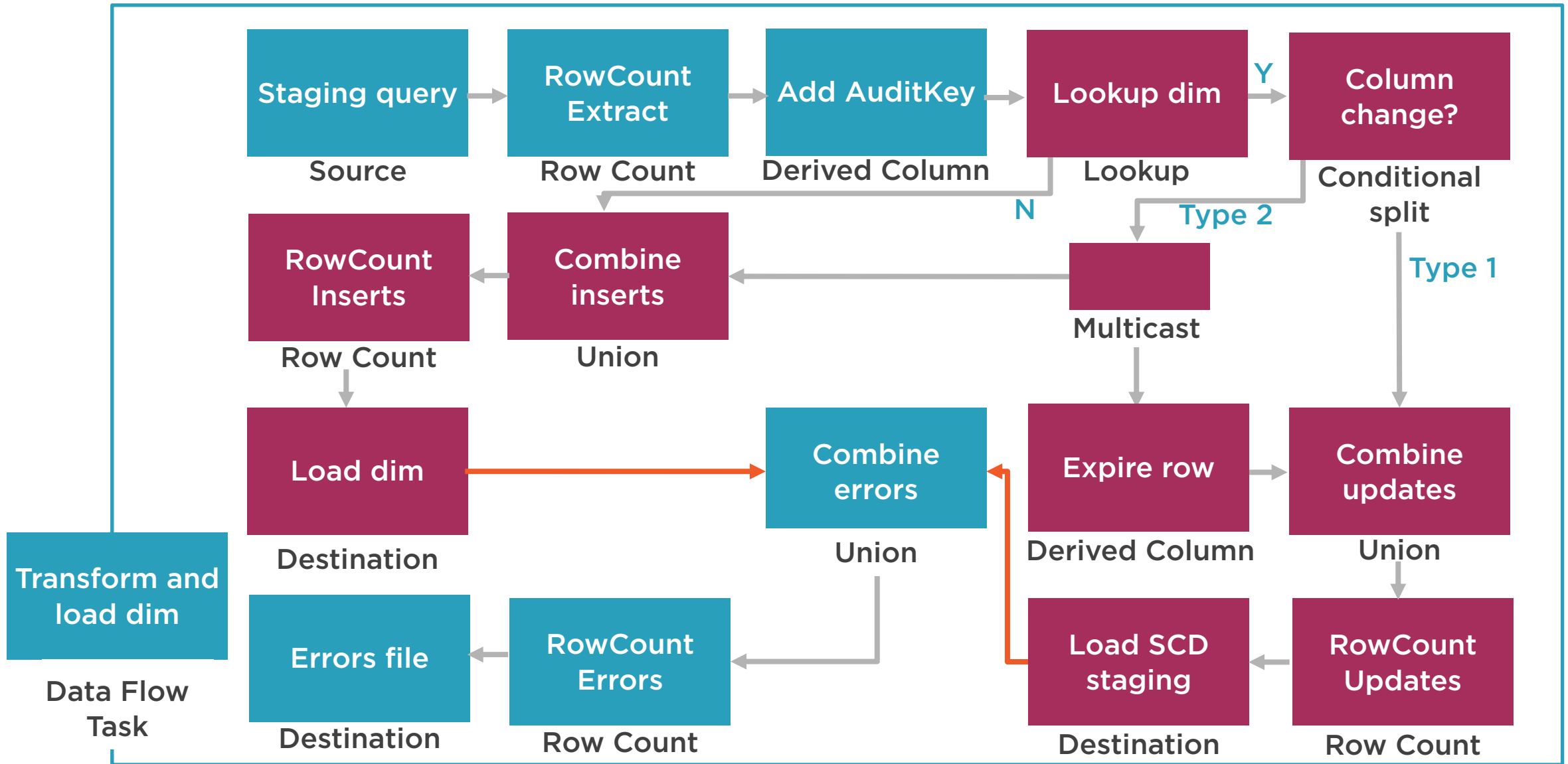
- Database-managed disk I/O
- Rollback statement for errors
- Complex queries
- All relational data
- Set operations
- Synchronous operations
- Efficient sorting

## SSIS

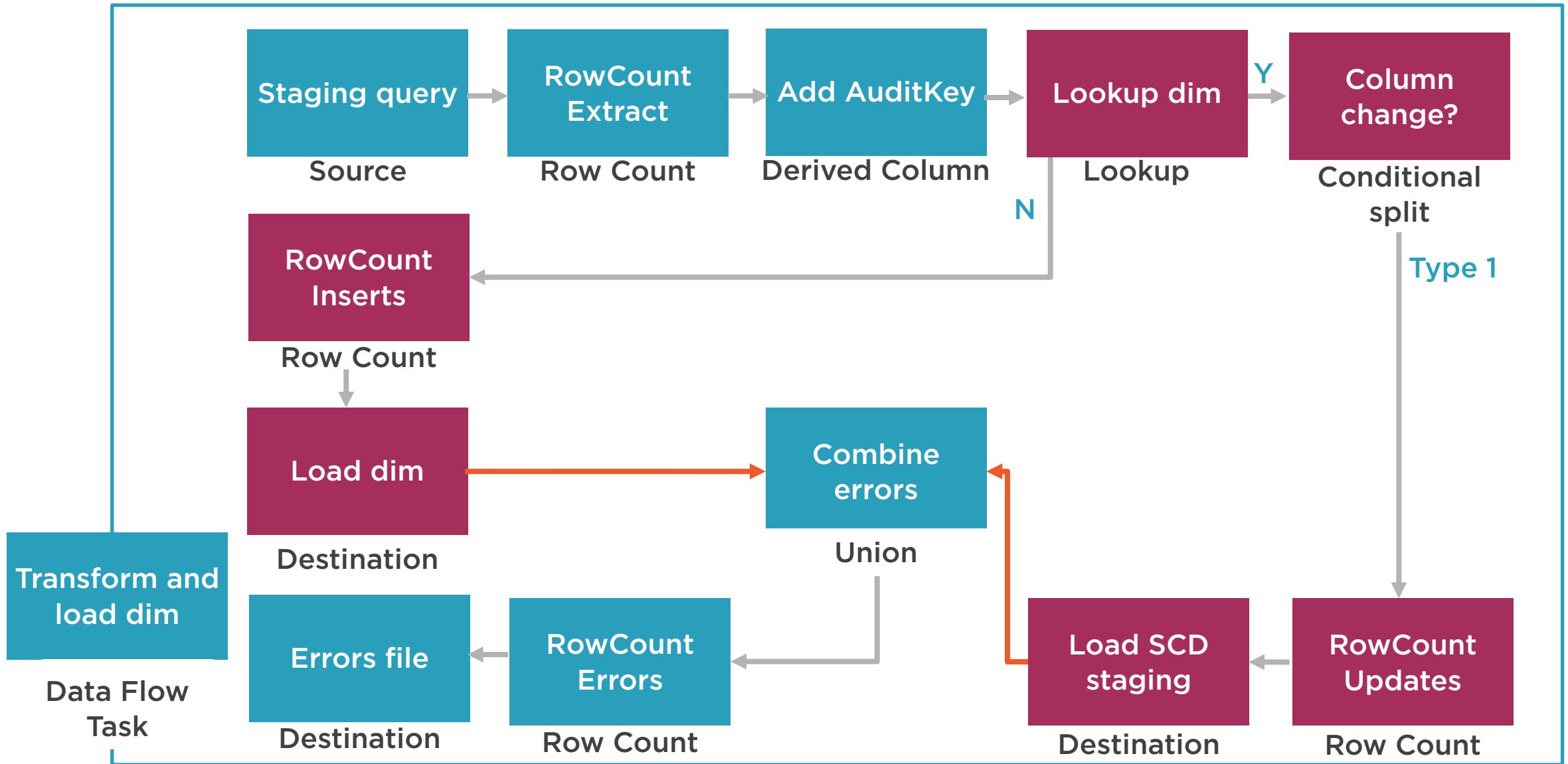
- In-memory buffers
- Redirection of errors
- Built-in transformations
- Multiple data source types
- Row-by-row operations
- Parallel operations
- Sort in memory can spill to disk



# Slowly Changing Dimension Handling



# Slowly Changing Dimension Handling



```
UPDATE target
SET  target.EndDate = source.EndDate,
     target.UpdateAuditKey = source.AuditKey
FROM DimProduct target
JOIN stage.tmp_scd_DimProduct source on
     target.ProductAlternateKey = source.ProductAlternateKey AND
     target.UpdateAuditKey <> source.AuditKey AND
     target.EndDate = '9999-12-31'
```

# Slowly Changing Dimension Handling

## Type 2



```
UPDATE target
SET target.ProductName = source.ProductName,
    target.ProductModelID = source.ProductModelID,
    <type 1 columns>...,
    target.Status = source.Status,
    target.UpdateAuditKey = source.AuditKey
FROM DimProduct target
JOIN stage.tmp_scd_DimProduct source on
    target.ProductID = source.ProductID
```

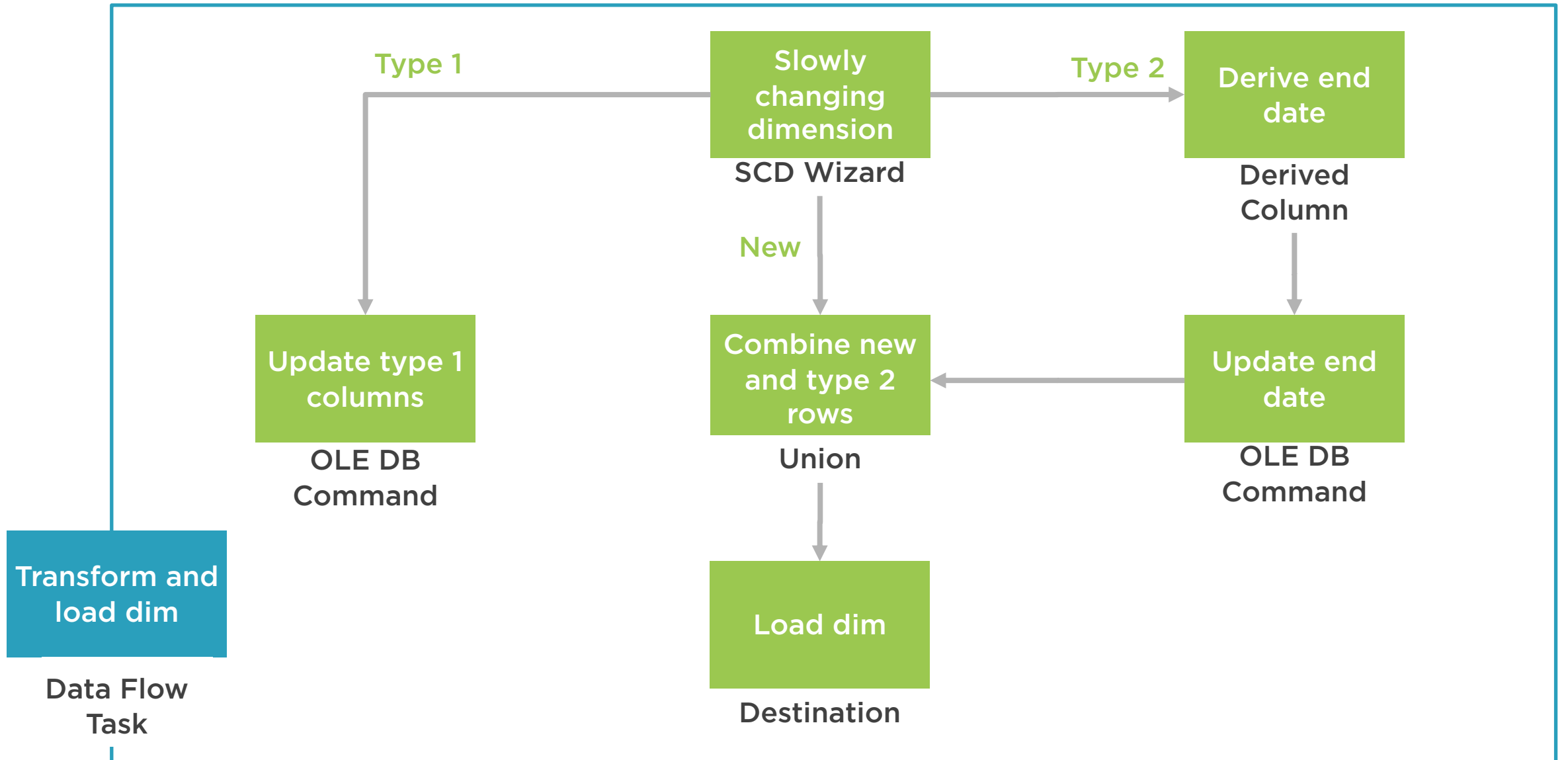
# Slowly Changing Dimension Handling

## Type 1

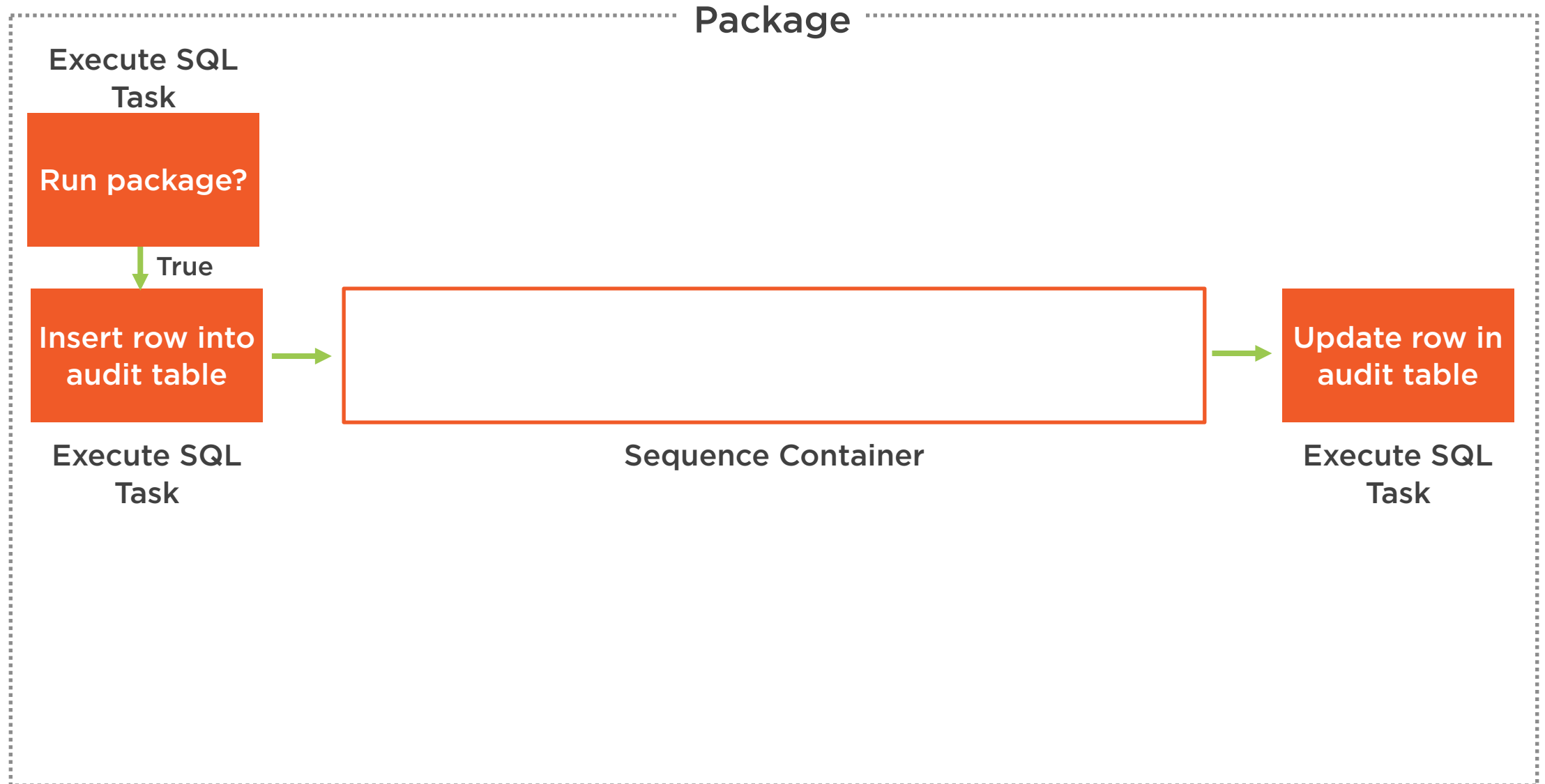




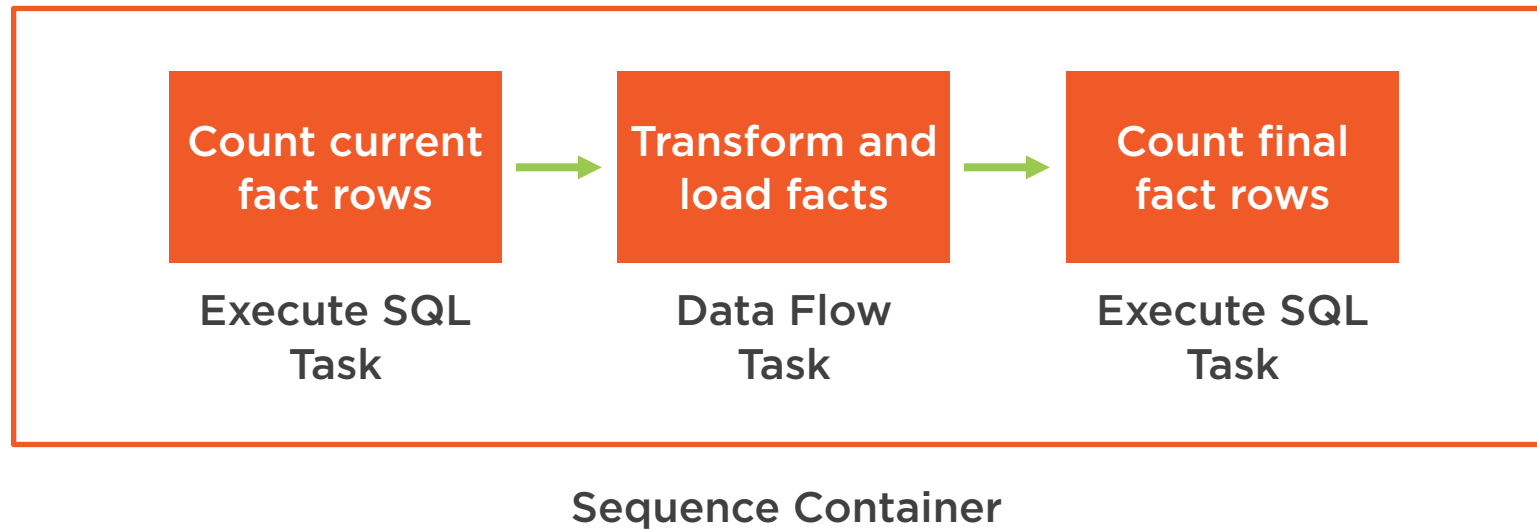
# Slowly Changing Dimension Wizard



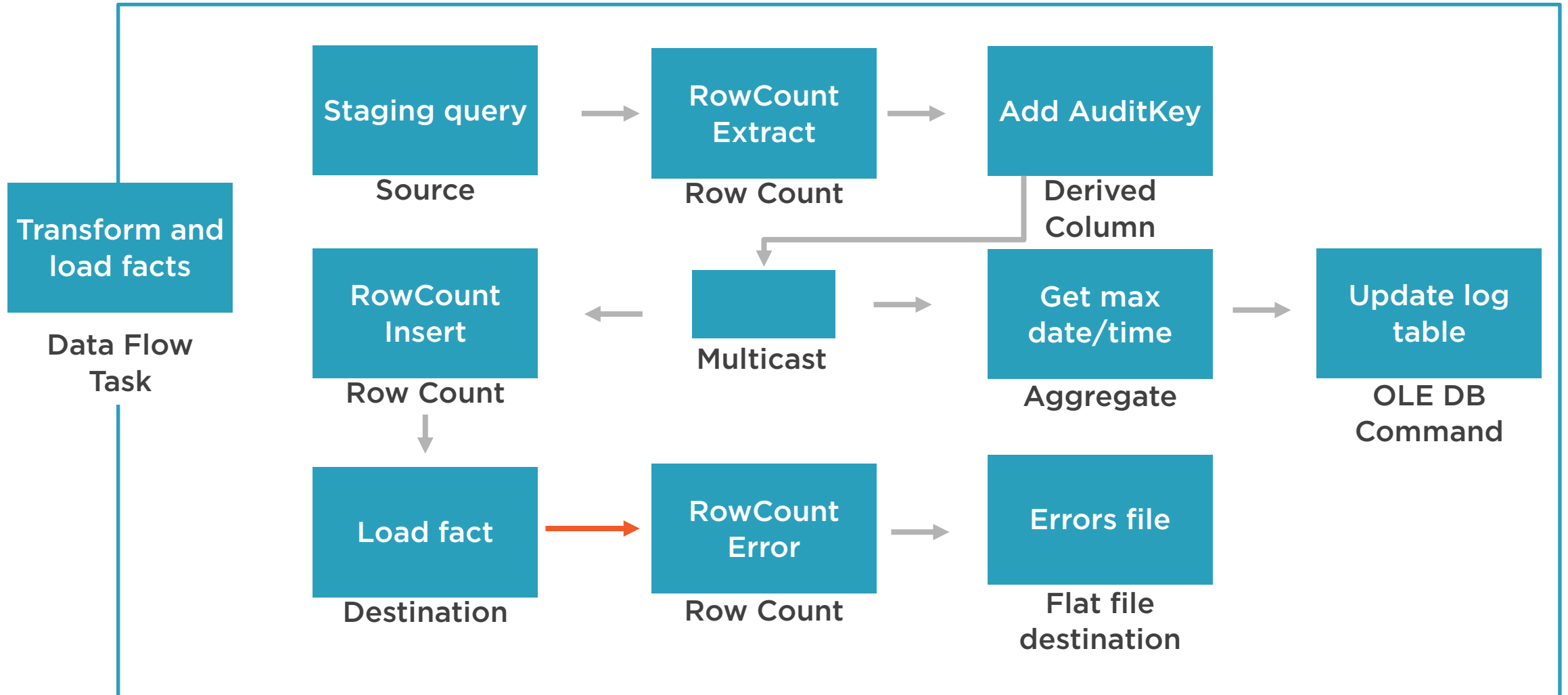
# Fact Loads



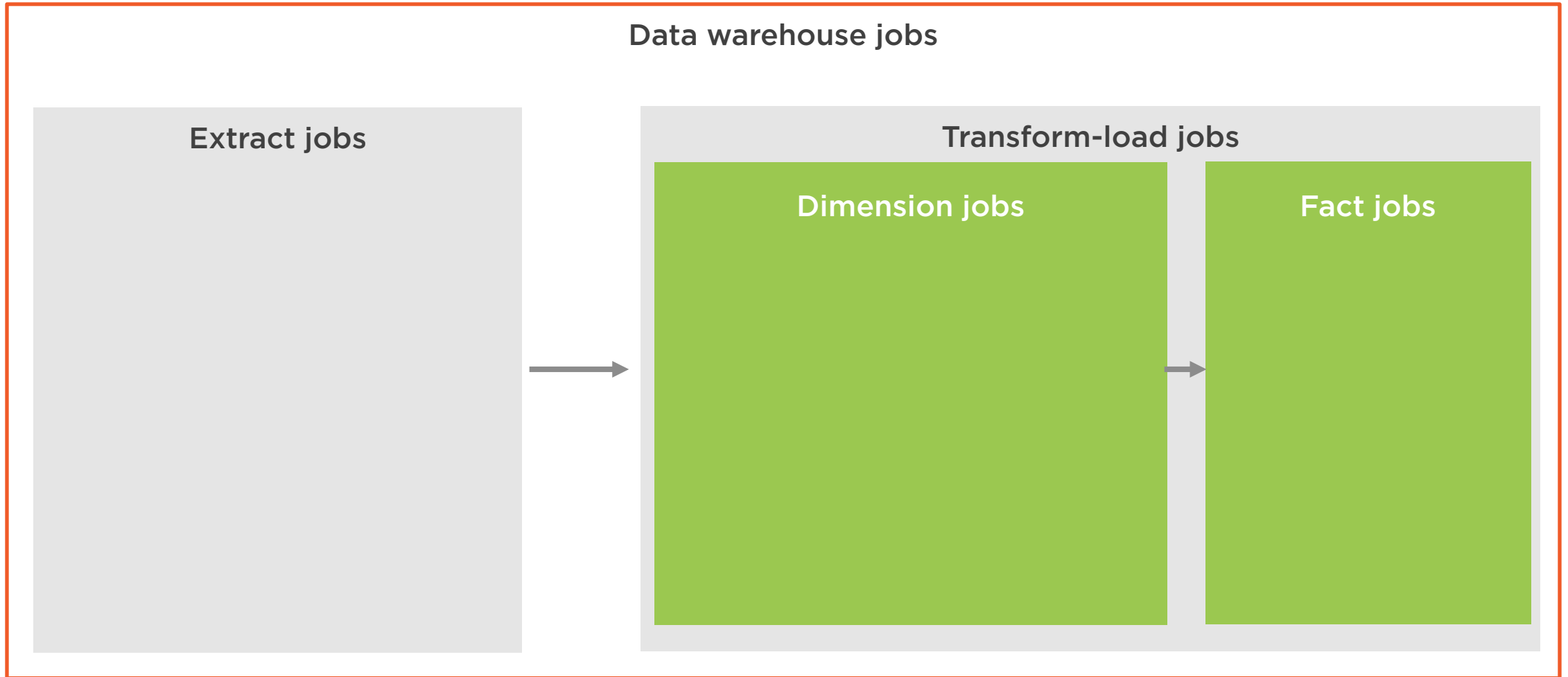
# Fact Loads



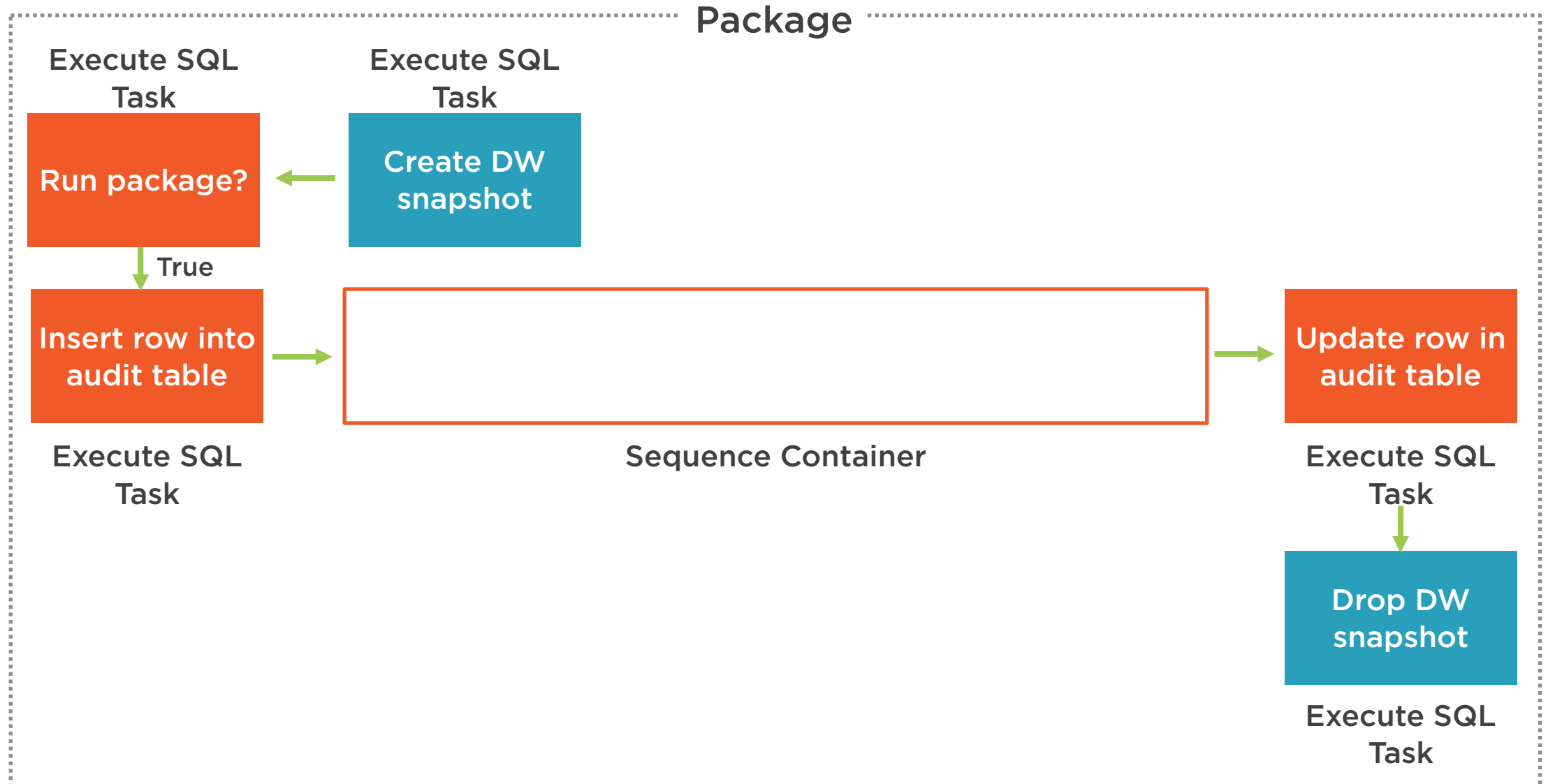
# Fact Loads



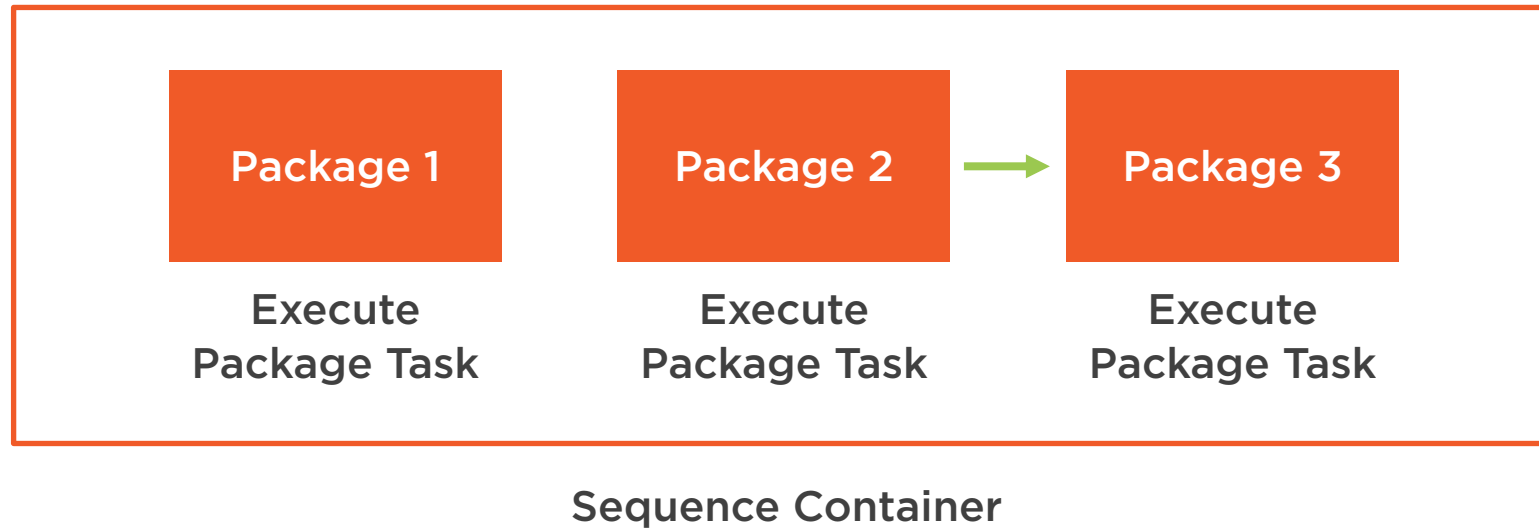
# Parent Packages



# Parent Packages



# Parent Packages



# Summary



**Package types**

**Template package**

**Staging table loads**

**Dimension table loads**

**Fact table loads**

**Parent packages**

