

Setting up an ETL Project in SSIS



Ana Voicu

@ana_voicu



Overview



What is an ETL tool

Advantages of using an ETL tool

Overview of SSIS

Demos

- Creating a new SSIS project with Visual Studio
- Adding connection managers
- Creating control flow components
- Linking tasks with precedence constraints
- Adding project variables
- Executing SQL tasks with parameters



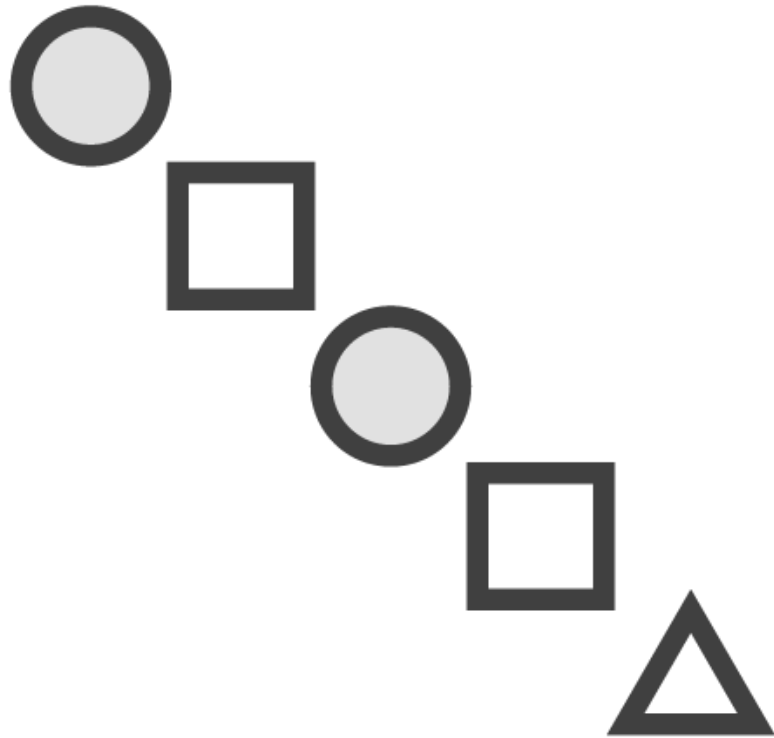
Disclaimer: This is not an
intensive course on SSIS



Advantages of Using an ETL Tool



What Are ETL Tools?

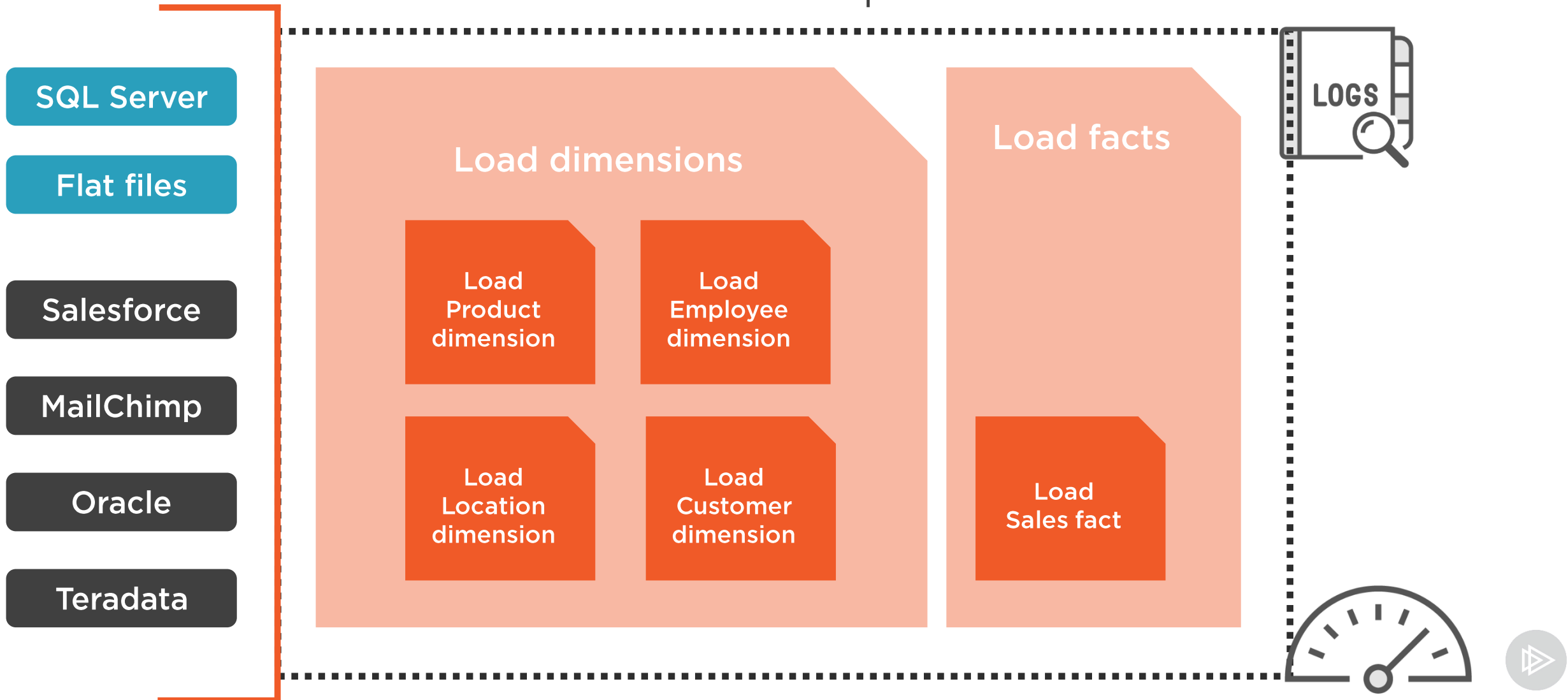


Software systems for building and automating ETL processes

Offer an alternative to creating all the code for a DW solution by hand



Loading a Data Warehouse Only Using SQL Scripts



What Is an ETL Tool?



A system comprised of all the components needed to build ETL processes

Components:

- Data flow management
- Connectors to other systems
- Data transformation components
- Logging and auditing
- Debugging
- Deployment

Advantages of Using an ETL Tool

Flexibility

Maintainability

- Easier than working with hundreds of stored procedures
- ETL tools are very graphical and easy to interpret
- Easy to learn and use



Advantages of Using an ETL Tool

Flexibility

Maintainability



Advantages of Using an ETL Tool

Flexibility

Maintainability

Performance

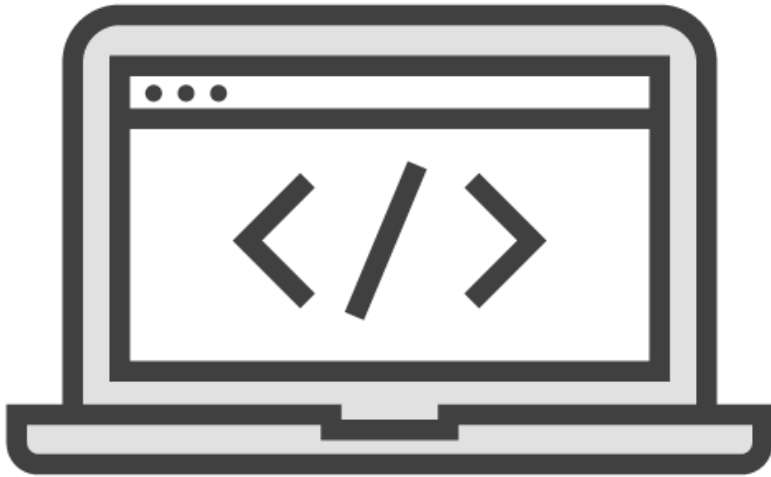
Logging and audit



Overview of SSIS



What Is SSIS?



Component of SQL Server

Used for data integration and automating processes

Provides all the components necessary for creating an ETL process

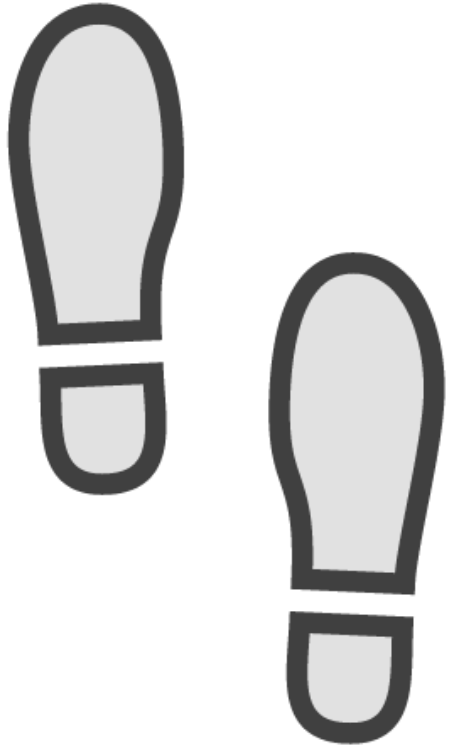
Graphical tool

- Doesn't require writing complex code
- Based on drag and drop actions

The output is a workflow, or pipeline

The unit of work in SSIS is called a package

Steps for Loading the Product Dimension



1. **Save the current load date**
2. **Insert new row in the Lineage table**
3. **Truncate the staging table**
4. **Populate the staging table**
5. **Transfer data from staging to dimension**

These steps are part of a flow or process

- Executed together
- In a specified order
- Executed multiple times

Overview of a SSIS Package



Get date of current load



Overview of a SSIS Package



Get date of current load

`@[User::LoadDate] = GETDATE()`



Overview of a SSIS Package



Get date of current load

`@[User::LoadDate] = GETDATE()`



Update Lineage table



Overview of a SSIS Package



Get date of current load

`@[User::LoadDate] = GETDATE()`



Update Lineage table

`EXEC [int].[Get_LineageKey] ?, ?;`



Overview of a SSIS Package



Get date of current load

`@[User::LoadDate] = GETDATE()`



Update Lineage table

`EXEC [int].[Get_LineageKey] ?, ?;`



Truncate staging



Overview of a SSIS Package



Get date of current load

`@[User::LoadDate] = GETDATE()`



Update Lineage table

`EXEC [int].[Get_LineageKey] ?, ?;`



Truncate staging

`TRUNCATE TABLE Staging_Product`



Overview of a SSIS Package



Get date of current load

`@[User::LoadDate] = GETDATE()`



Update Lineage table

`EXEC [int].[Get_LineageKey] ?, ?;`



Truncate staging

`TRUNCATE TABLE Staging_Product`



Populate staging



Overview of a SSIS Package



Get date of current load

`@[User::LoadDate] = GETDATE()`



Update Lineage table

`EXEC [int].[Get_LineageKey] ?, ?;`



Truncate staging

`TRUNCATE TABLE Staging_Product`



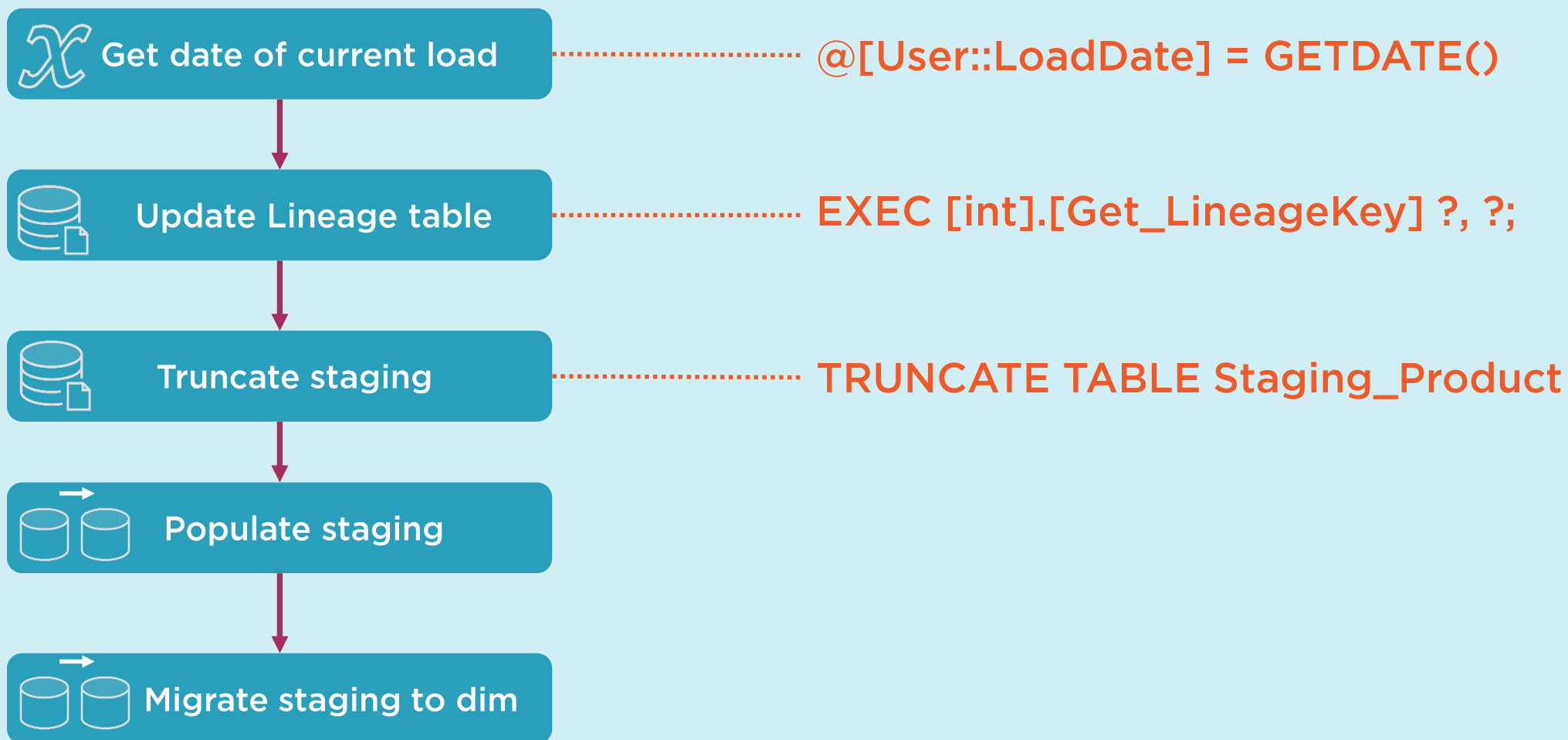
Populate staging



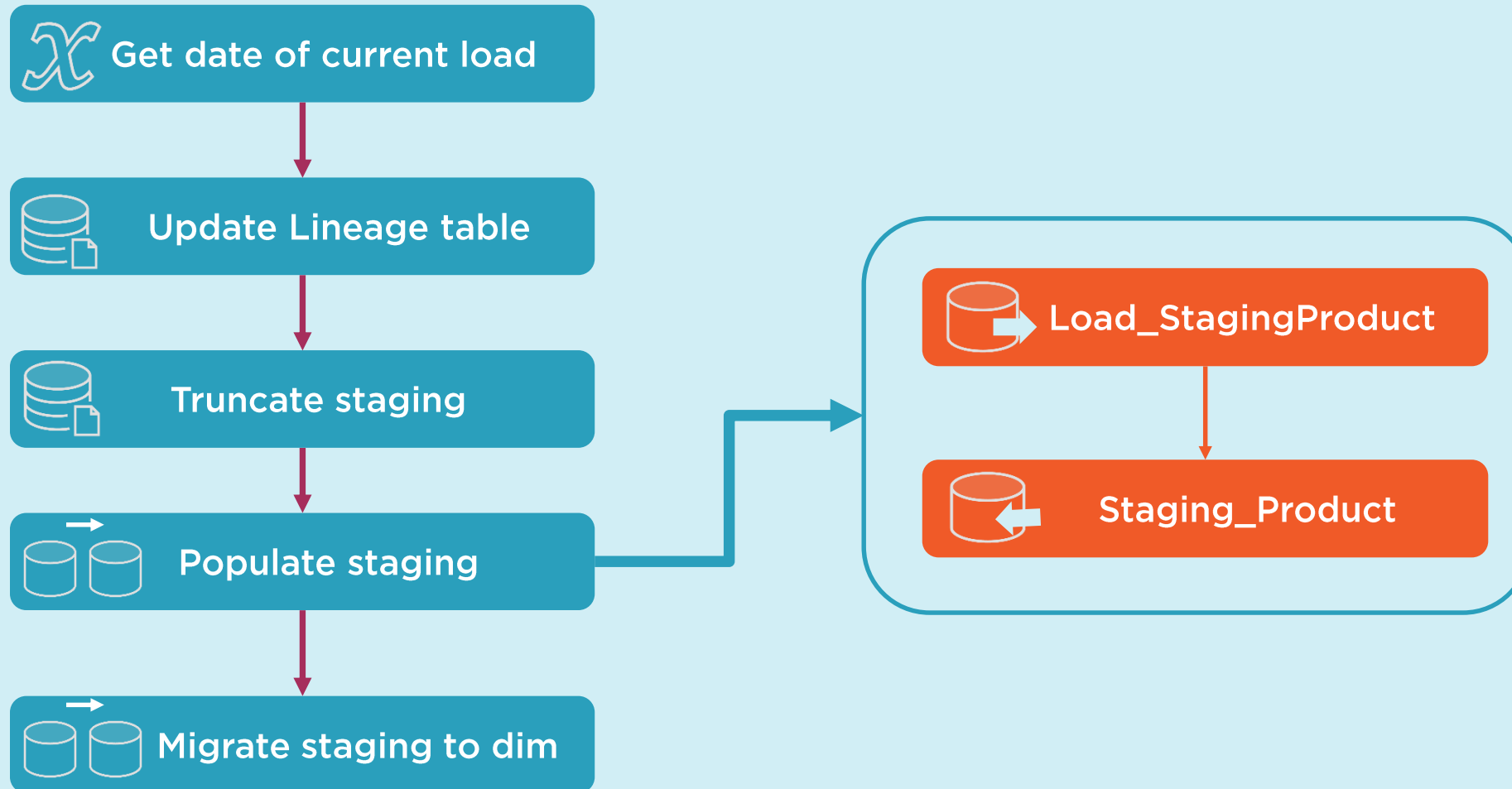
Migrate staging to dim



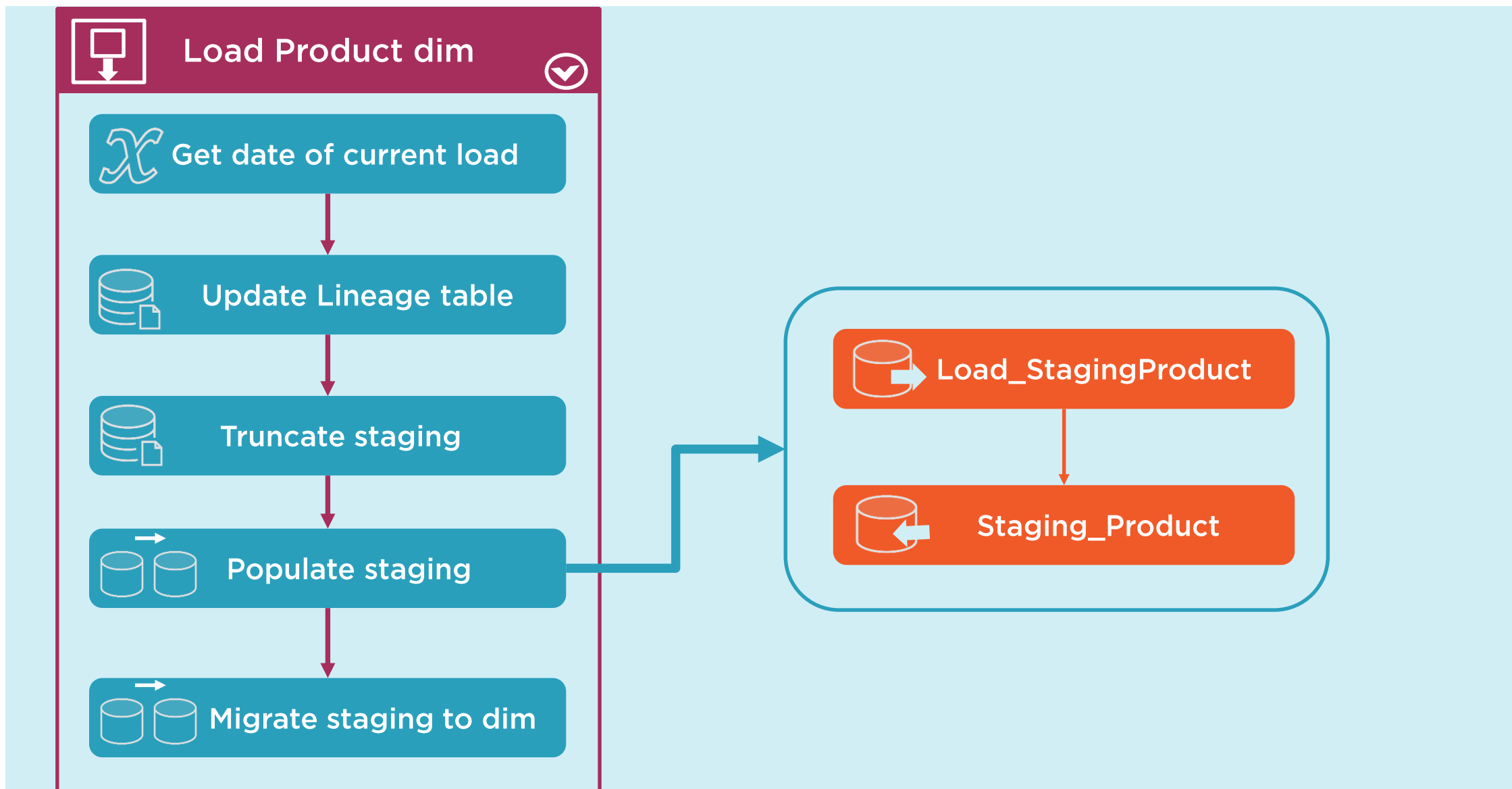
Overview of a SSIS Package



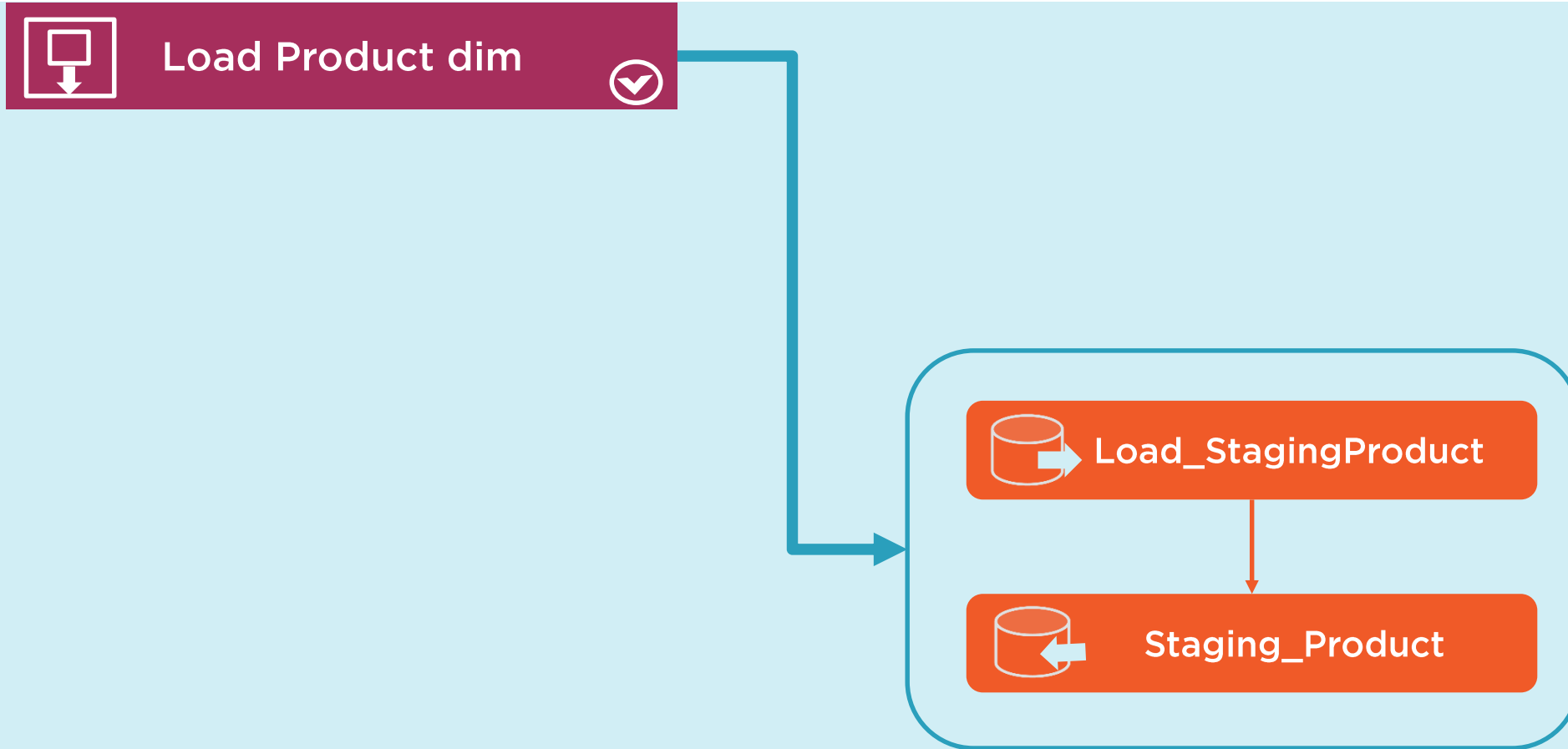
Overview of a SSIS Package



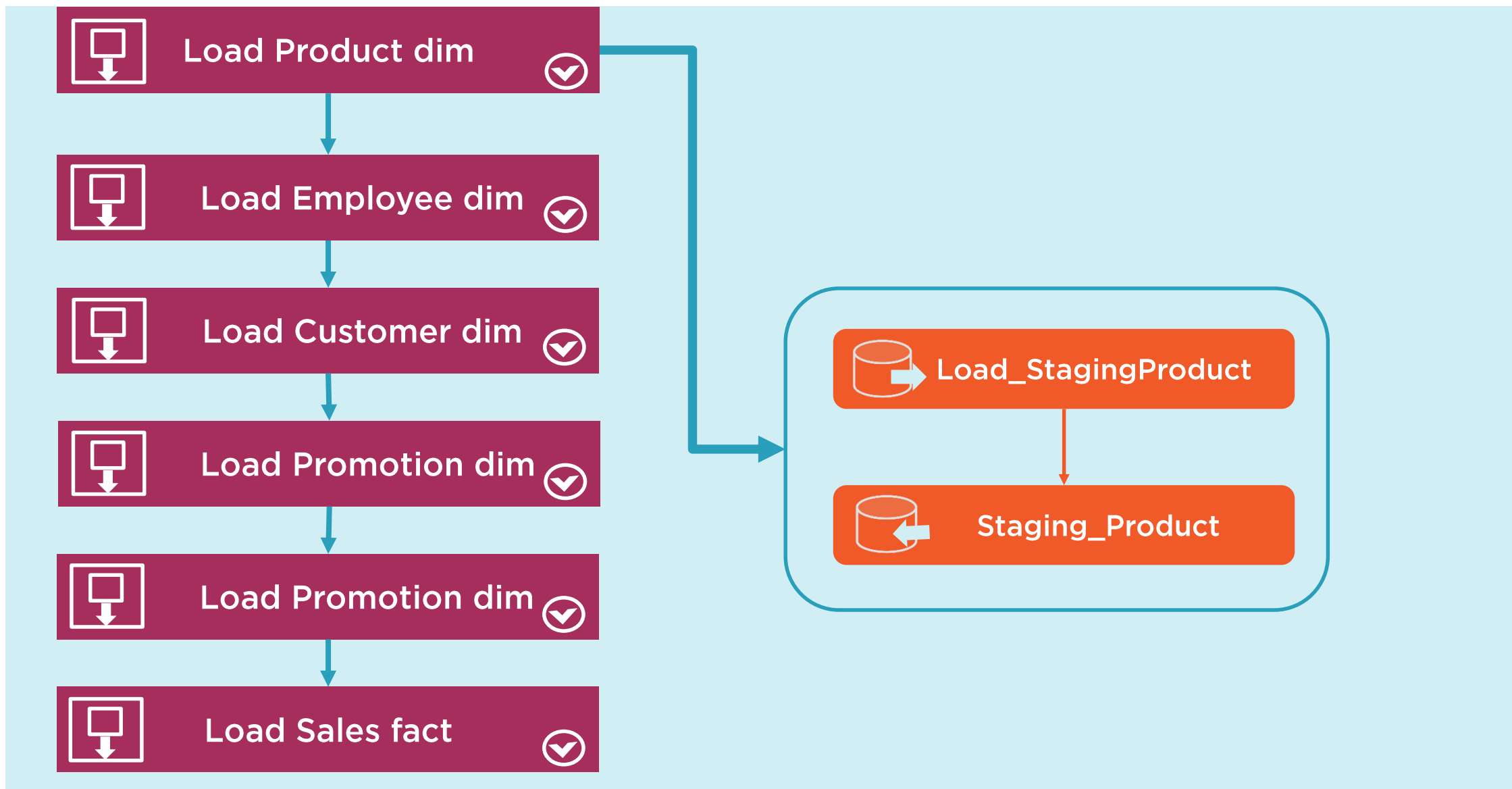
Overview of a SSIS Package



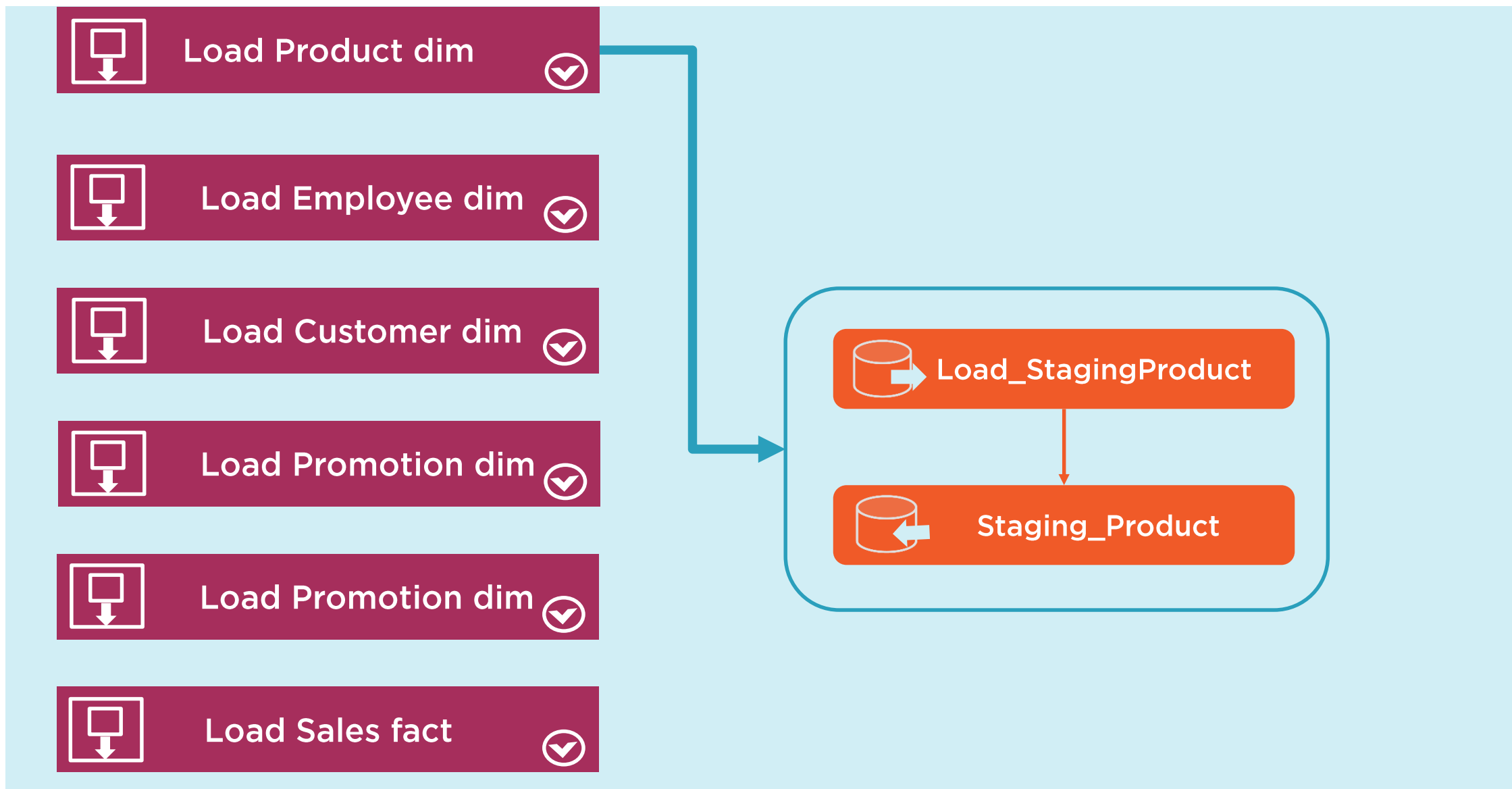
Overview of a SSIS Package



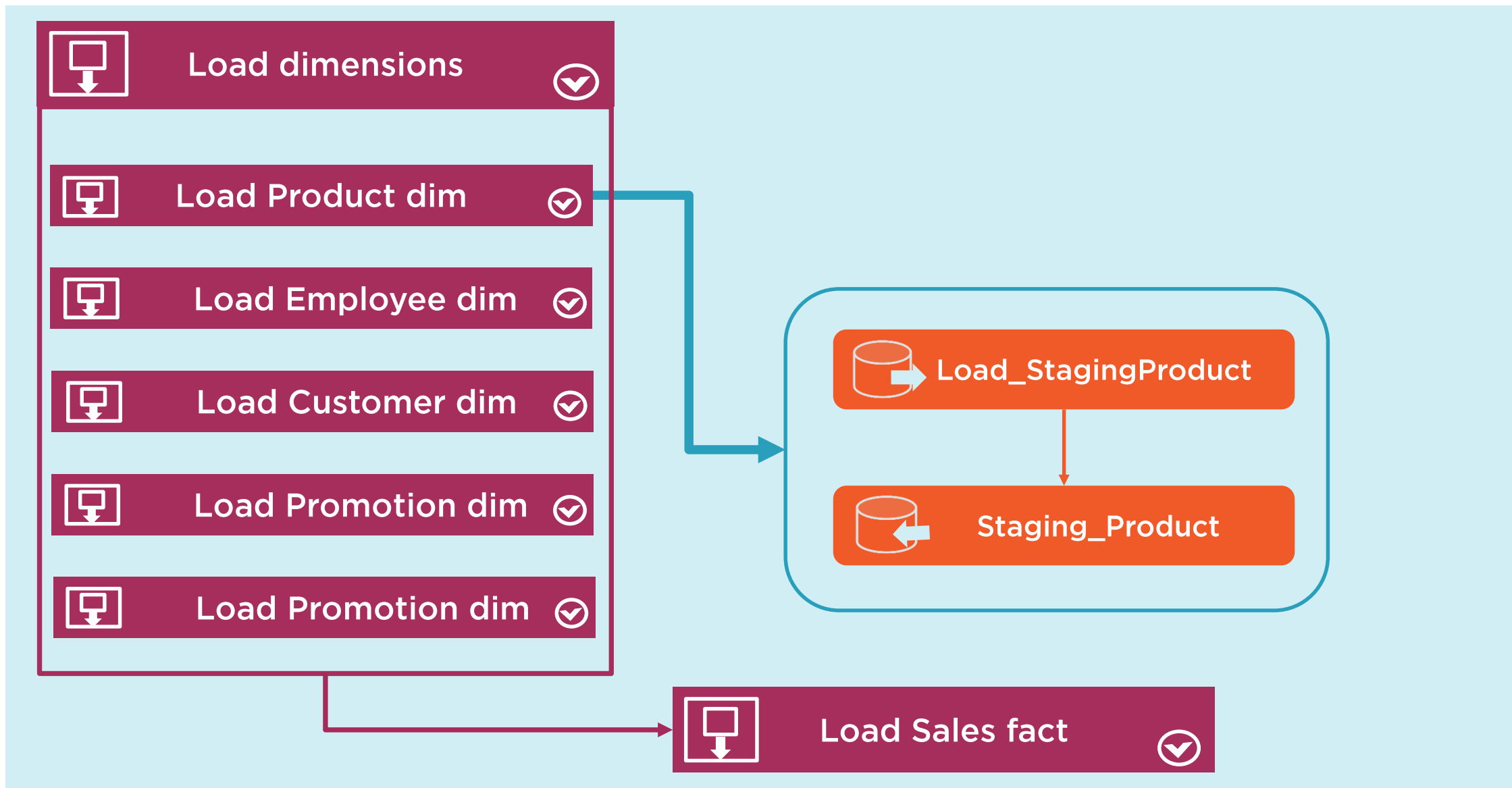
Overview of a SSIS Package



Overview of a SSIS Package



Overview of a SSIS Package



SSIS Terminology



SSIS Terminology

SSIS package	



SSIS Terminology

SSIS package	
Task	



SSIS Terminology

SSIS package	
Task Control flow task	



SSIS Terminology

SSIS package	
Task	
Control flow task	
Data flow task	



SSIS Terminology

SSIS package	
Task Control flow task Data flow task	
Control flow	



SSIS Terminology

SSIS package	
Task Control flow task Data flow task	
Control flow	
Data flow	



SSIS Terminology

SSIS package	
Task Control flow task Data flow task	
Control flow	
Data flow	
Connection manager	



SSIS Terminology

SSIS package	Container
Task Control flow task Data flow task	
Control flow	
Data flow	
Connection manager	



SSIS Terminology

SSIS package	Container
Task Control flow task Data flow task	Precedence constraint
Control flow	
Data flow	
Connection manager	



SSIS Terminology

SSIS package	Container
Task Control flow task Data flow task	Precedence constraint
Control flow	Variable
Data flow	
Connection manager	



SSIS Terminology

SSIS package	Container
Task Control flow task Data flow task	Precedence constraint
Control flow	Variable
Data flow	Parameter
Connection manager	



Demo



Creating a new SSIS project



Demo



Creating connection managers

- This is one of the first tasks to do in a new project
- Almost all components in SSIS need a connection



Demo



Creating SSIS components for loading Product dimension

- Connection managers
- Tasks
- Precedence constraints
- Containers

The package will do the following:

- Update Lineage and Incremental loads tables
- Transfer data into the staging table
- Update the dimension table



Demo



Creating and working with variables



Demo



Creating precedence constraints



Demo



Executing a SQL task with parameters



Practical Exercise

Retrieve the date of the previous load

- Create a SQL task that executes the stored procedure `[int].[Get_LastLoadedDate]`
- Save the result in the variable: `[User::PreviousLoadDate]`
- Execute the task after the staging table is truncated
- Rename the procedure “**SQL Get date of the previous load**”



Summary



Advantages of using ETL tools

High-level overview of SSIS

Creating an SSIS project

Configuring connections to the data sources

Creating control flow tasks

Setting up the order of execution with precedence constraints

Creating and working with package variables

