Extract, Transform & Load Basics

- Prof. Zarna Barai
- Unit 4, Chap 7

Topics Covered

- Source structures are defined
- Target structures are defined
- Now we need to work on
  - Extracting the data from our sources
  - Perform any transformation on that data
  - Load it into our target data warehouse structure
- This can be done by
  - Designing mappings in OWB
  - Deploying mapping to database
  - Running the mappings

Manual ETL Processes

- Extract the data from the source system by some method.
- Load flat files using SQL*Loader or via direct database link. Transform that data with SQL or PL/SQL code in the database to match and fit into the target structure.
- Finally load the data into the target structure.

Staging

- Staging is a process of copying the source data temporarily into a table(s) in our target database.
- Here perform transformations that are required before loading the source data into the final target tables.
- This process involves saving data to storage at any step along the way to the final target structure.

To Stage Or Not To Stage

- No. of considerations taken into account when deciding whether to use staging area or not:
  - To keep the process flowing as fast as possible
    - Amount of source data
    - Amount of manipulations on source data
    - Reliable connection to the database and the performance of the link while pulling data across
  - Severity of the impact of failure of ETL Process
    - Restart the ETL process
    - Source data is changing while trying to load the data initially. That means we will have a different data during the restart process, which makes it difficult to debug the error that caused this failure.

Configuration Of A Staging Area

- Create staging area either
  - Outside the database – flat files
    - A flat file can be treated as another table in database
    - Access flat files with all the benefits of SQL
    - Cannot be used as target
  - Inside the database
- At ACME TOYS & Gizmos we are going to use a table in the database for initial staging

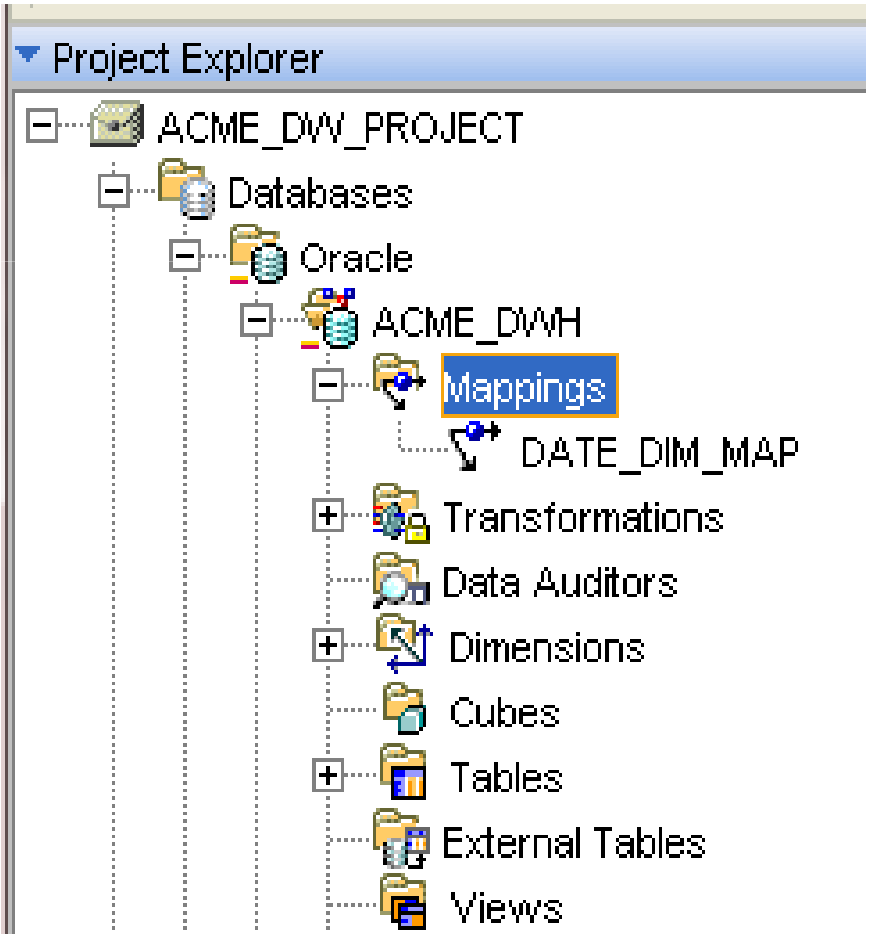
Mappings & Operators In OWB

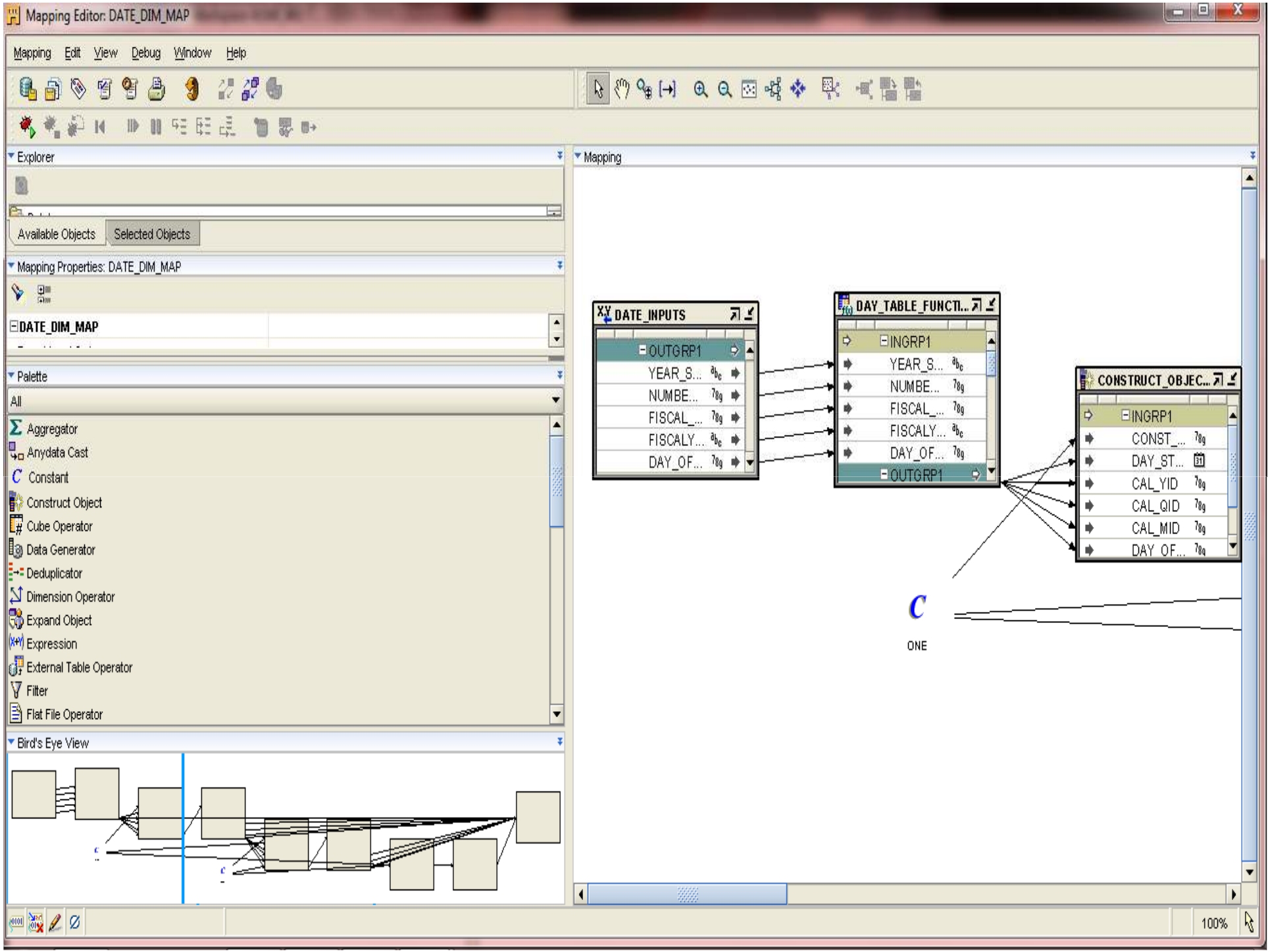
- A mapping is composed of as series of operators that describes the sources, targets and a series of operations that flow from source to target to load the data.
- Mappings are designed in graphical manner using Mapping editor.
- Mapping are created under mappings node.

Mappings & Operators In OWB

DATE_DIM_MAPPING

- To start the mapping editor, login into the design center, in the Project Explorer window expand the ACME_DW_PROJECT & then expand the MAPPINGS node underneath it.
- Double – click on DATE_DIM_MAPPING, it will launch the mapping editor.
- The mapping editor window appears




Mapping Editor: DATE_DIM_MAP
Mapping Edit View Debug Window Help
Explorer
Available Objects
Selected Objects
Mapping Properties: DATE_DIM_MAP
DATE_DIM_MAP
Palette
All
Aggregator
Anydata Cast
Constant
Construct Object
Cube Operator
Data Generator
Deduplicator
Dimension Operator
Expand Object
Expression
External Table Operator
Filter
Flat File Operator
Bird's Eye View
Mapping
DATE_INPUTS
OUTGRP1
YEAR_S...
NUMBE...
FISCAL_...
FISCALY...
DAY_OF...
DAY_TABLE_FUNCTI...
INGRP1
YEAR_S...
NUMBE...
FISCAL_...
FISCALY...
DAY_OF...
OUTGRP1
CONSTRUCT_OBJEC...
INGRP1
CONST_...
DAY_ST...
CAL_YID
CAL_QID
CAL_MID
DAY_OF...
ONE
100%

Mappings & Operators In OWB

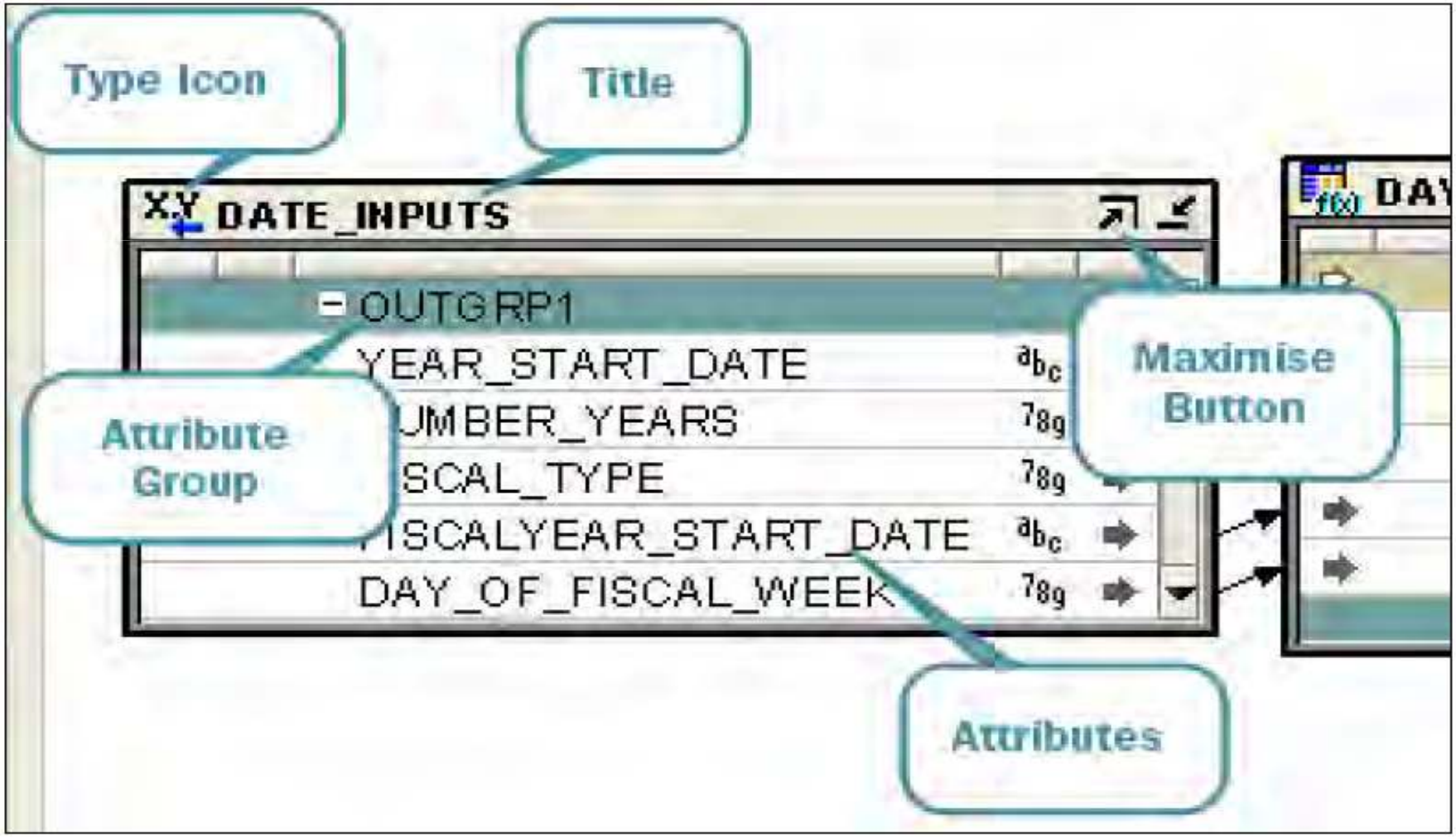
- Mapping Editor consists of:
  - Mapping
  - Explorer
    - Available Objects
    - Selected Objects
  - Mapping Properties
  - Palette
  - Bird's Eye View

The Canvas Layout

- Operators on the canvas are represented by boxes, with a title that represents the name of the operator & an icon that indicates the type.
- Operator box can be resized.
- 2 major types of attributes:
  - An Input Group - INGRP1 (generic name)
  - An Output Group – OUTGRP1 (generic name)
- An attribute group name is edited in the 'DETAILS' window for the group name in the canvas and selecting 'Open Details'

K1

The Canvas Layout

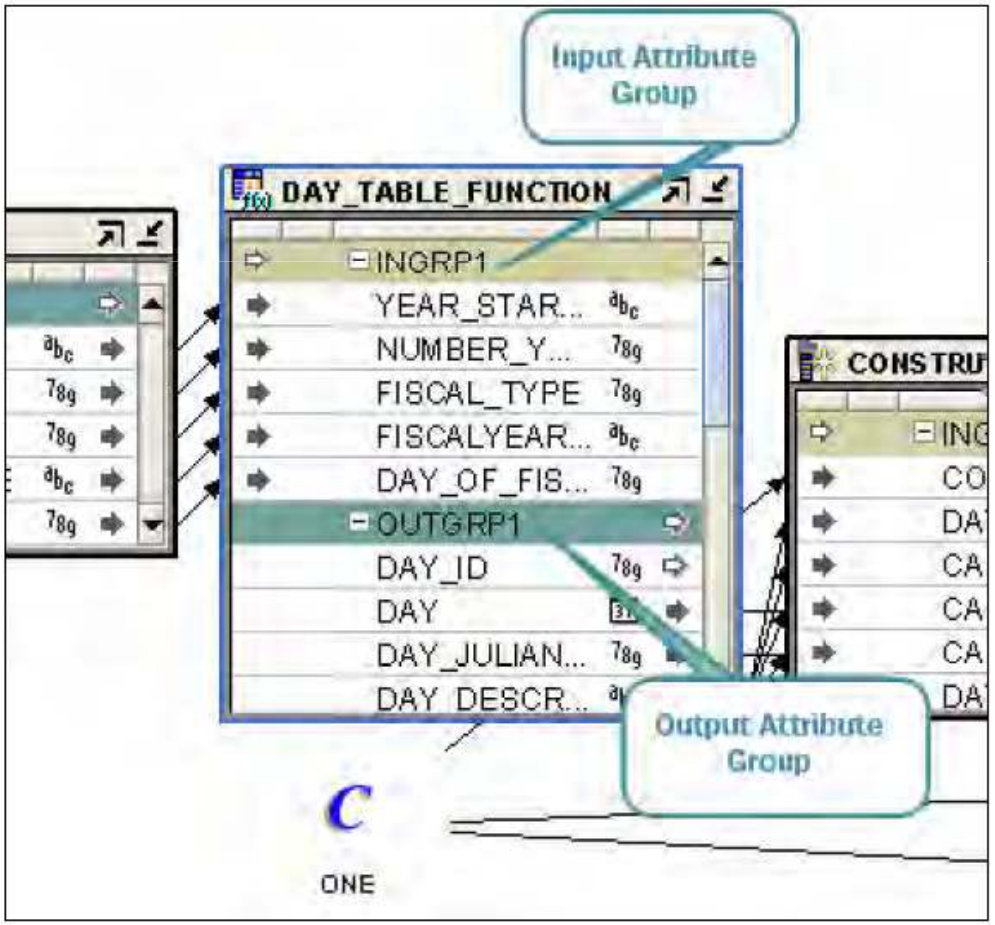


Slide 12

K1 let us take a closer look at some of the general features of the operators that we can see in canvas and then at some of the features specific to different operators.

Krishna, 1/29/2014

The Canvas Layout



OWB Operators

- Source & Target Operators
  - Cube operator
  - Dimension operator
  - External tables
  - Table operator
- Most commonly used operators
  - Constant
  - View Operator
  - Sequence Operator
  - Construct Object

OWB Operators

- Data flow operators
  - Aggregator
  - Deduplicator
  - Expression
  - Filter
  - Joiner
  - Key lookup
  - Pivot
  - Set Operator
  - Splitter operator
  - Transformation Operator

OWB Operators

- Pre/Post Processing Operators
  - Mapping Input Parameter
  - Mapping Output Parameter
  - Post Mapping Process
  - Pre Mapping Process

The END