BUILDING THE DATA WAREHOUSE – PART 2

MODULE 11

1

DIMENSIONS

Building the Master Dimensions – SCD Type 1

2

SLOWLY CHANGING DIMENSIONS

Building the Product Dimension – SCD Type 2

3

FACTS

Building the Fact Table

4

REVIEW

Review of the Data in the Data Warehouse

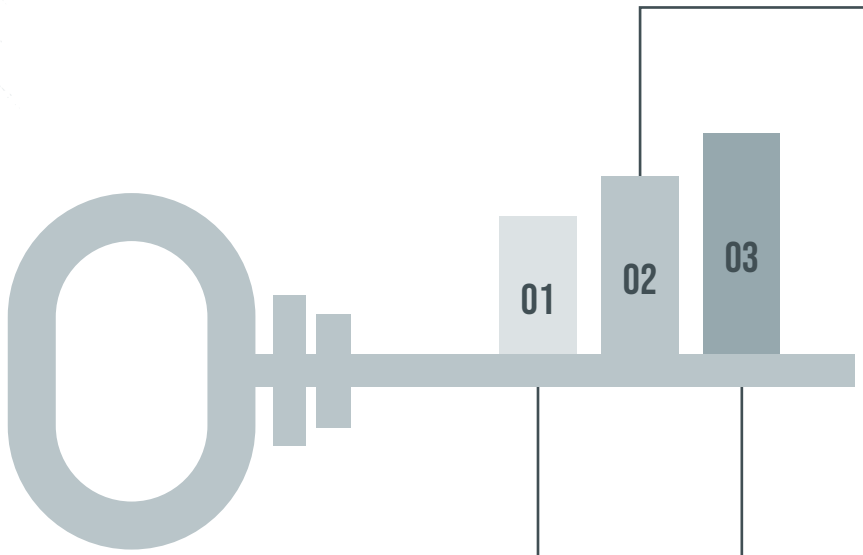
SECTION 1

DIMENSIONS

DIMENSIONS

Overview of Dimensions

DIMENSIONS PROVIDE DESCRIPTIVE CONTEXT FOR THE QUANTITATIVE DATA IN THE FACT TABLE IN A DATA WAREHOUSE.



DENORMALIZED

- *Dimensions are denormalized tables*
- *Hierarchical attributes in a dimension are flattened*
- *Prevents snow-flaking and reducing expensive joins*

DESCRIPTIVE

- *Provide descriptive information about the business*
- *Use surrogate keys to join and describe data in the fact table*
- *Allows for slicing the quantitative data over different dimensions*
- *Time dimension provides historical context for the data in the fact table*

HIERARCHICAL

- *Attributes in a dimension organized hierarchically for analytics*
- *Use hierarchies to aggregate or drill-down quantitative data in the fact table*

DIMENSIONS

Overview of Slowly Changing Dimensions

DIMENSIONS CONNECTED TO A FACT TABLE ARE ALSO AFFECTED BY THE PASSAGE OF TIME

TYPE 0

Changes Ignored

- *Data cannot be changed*
- *No history preservation*

OLD

Key	ID	Name	City
S123	123	Mike	Rome

NEW

Key	ID	Name	City
S123	123	Mike	Rome

TYPE 1

Changes Updated (No History)

- *Data can be changed*
- *Overwrite old data with new data*
- *No history preservation*

OLD

Key	ID	Name	City
S123	123	Mike	Rome

NEW

Key	ID	Name	City
S123	123	Mike	Milan

TYPE 2

Changes Inserted (History Preserved)

- *Create new record by tuple versioning*
- *Historical record is made inactive*
- *New record is made active*

OLD

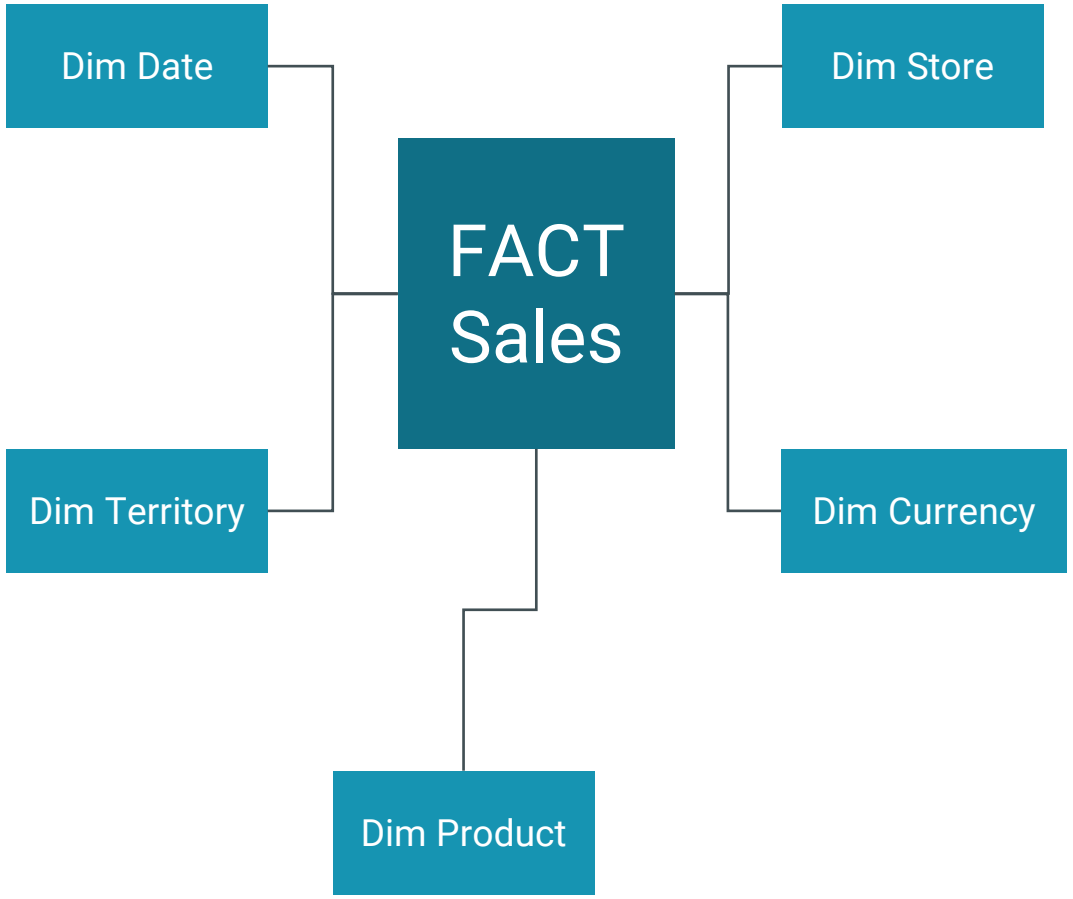
Key	ID	Name	City	From	To	Active
S123	123	Mike	Rome	Jan 1 2022	-	Yes

NEW

Key	ID	Name	City	From	To	Active
S123	123	Mike	Rome	Jan 1 2022	Dec 31 2022	No
S124	123	Mike	Milan	Jan 1 2023	-	Yes

DIMENSIONS

Master Dimensions and their type



TYPE 0

- *None*

TYPE 1

- *Dim Date*
- *Dim Currency*
- *Dim Store*
- *Dim Territory*

- *Changes to these dimensions are rare but updates are possible*
- *History preservation is not required*

TYPE 2

- *Dim Product*

- *Changes and history are required*

Important to define the type of dimension as part of the data warehouse design

DIMENSIONS

Building the SCD Type 1 Dimensions

MAIN STEPS IN BUILDING THE SCD TYPE 1 DIMENSIONS

READ SOURCE

- *Read Stage Table*

COMPARE TARGET

- *Use a conditional split to compare Source and Target*
- *Compare on natural key of the source and target*

LOAD TARGET

- *Update existing record in Target if it already exists*
- *Insert new record in Target if it doesn't exist*

Building a Type 1 dimension

- *Using Mapping Data Flows*
- *Using Stored Procedures*

SECTION 2

SLOWLY CHANGING DIMENSIONS

SLOWLY CHANGING DIMENSIONS

Product Dimension – SCD Type 2

PRODUCT DIMENSION HAS ATTRIBUTES THAT CAN CHANGE AND WHERE HISTORY PRESERVATION IS NECESSARY

TYPE 2

Changes Inserted (History Preserved)

OLD

ID	No	Title	Vintage	Score	From	To	Active
1	123	Nebbiolo	2015	95	1/1/2022	-	Yes

NEW

ID	No	Title	Vintage	Score	From	To	Active
1	123	Nebbiolo	2015	95	1/1/2022	31/12/2022	No
2	123	Nebbiolo	2015	93	1/1/2023	-	Yes

Title

- *The title of the wine, typically doesn't change*

Vintage

- *Vintage of the wine doesn't change, since it is the year of the wine*

Score

- *The score of the wine can change since it depends on reviewers*
- *Essential to preserve history*

DIMENSIONS

Building the SCD Type 2 Dimension

MAIN STEPS IN BUILDING SCD TYPE 2 DIMENSIONS

READ SOURCE

- *Read Stage Table*

COMPARE TARGET

- *Use a Lookup to compare Source and Target*
- *Compare source with matched records from lookup*

LOAD TARGET

- *Add new record with new surrogate key if records differ on Type 2 attribute*
- *Add new record if record doesn't exist in target with new surrogate key*
- *Use derived transformation to add effective start and end dates and active flag*

Building a Type 2 dimension

- *Using Mapping Data Flows*
- *Using Stored Procedures*

Assignment

- *Build all Type 1 dimensions using Stored Procedures and invoke them from ADF*
- *Build one of the Type 1 dimensions using Mapping Data Flows*
- *Test the implementations and review the dimension data*

Building remaining dimensions

- *Invoke Stored Procedures*

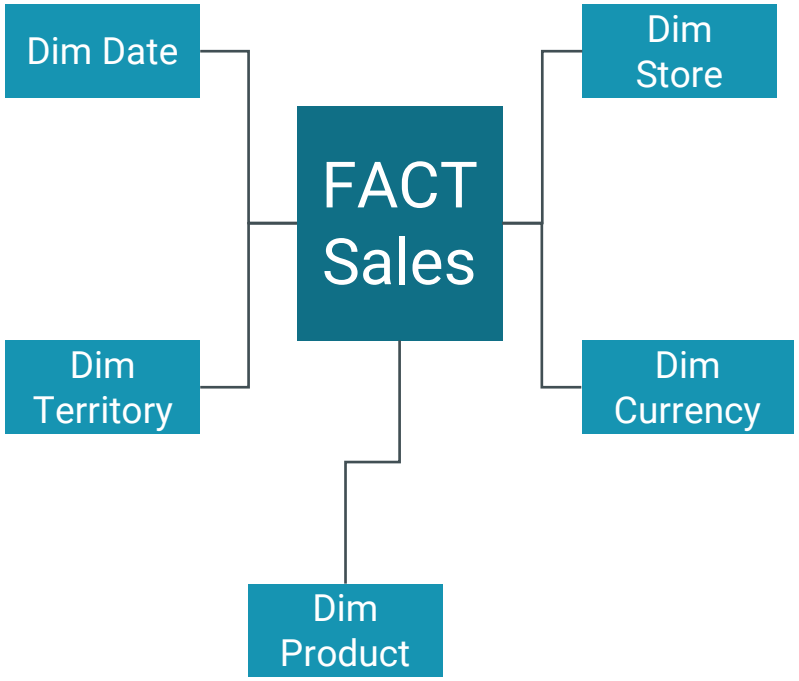
SECTION 3

FACTS

FACTS

Building the Fact Table

FACT TABLES ARE GENERALLY BUILT AFTER DIMENSIONS. THIS ENABLES THE ASSIGNMENT OF THE DIMENSION SURROGATE KEYS



FACT TABLE

	Column Name
key	StoreId
key	TerritoryId
key	DateId
key	CurrencyId
key	ProductId
	SalesQty
	SalesAmount
	CostAmount
	MarginAmount

Derive dimension surrogate keys by joining the stage sales data with the dimensions

Derive fact measures by applying the appropriate calculations

FACTS

Building the Fact Table

MAIN STEPS IN BUILDING FACT TABLES

READ SOURCE

- *Read Stage Sales Table*

COMPARE TARGET

- *Lookup relevant dimension tables*
- *Retrieve the dimension surrogate keys*
- *Calculate or derive required measures*
- *Use derived transformation to add dimension surrogate keys*

LOAD TARGET

- *Use merge transformation to merge fact table with dimension table using the surrogate keys*
- *Load merged data into the fact table*

Building the Fact Table

- *Loading Fact data from Stage Sales Transactions*
- *Deriving Dimension Keys*

SECTION 5

REVIEW

REVIEW

Review of the Data Warehouse

What did we build?

- *Dimensions – Type 1 and Type 2*
- *Facts – Dimension Keys and Measures*

Review the Data Warehouse

- *Review the Data Warehouse with canned queries*

MODULE SUMMARY

In this module we learnt

OVERVIEW

We got an overview of dimensions and their benefits.

We learnt about different types of slowly changing dimensions

INTEGRATION

We learnt the concepts of loading Type 1 and Type 2 dimensions

We learnt about the concept of loading a Fact table and the different ways to handle delta loads

HANDS-ON

We learnt how to build the Type 1 and Type 2-dimension patterns

We then built the Fact table

We then reviewed our data warehouse with various queries to analyze the data

REFERENCES

Surrogate Keys

Surrogate Keys | James Serra's Blog

Populating a Data Warehouse

Methods for populating a data warehouse | James Serra's Blog

Alter Row Transform

https://learn.microsoft.com/en-us/azure/data-factory/data-flow-alter-row