3. Data Modelling

3.0 Chapter Goals

- Fundamental Concepts
- What is Data Modelling
- Traditional Data Modelling
  - Star Schema
  - Snowflake Schema
- QlikView Data Modelling
  - Linked Tables
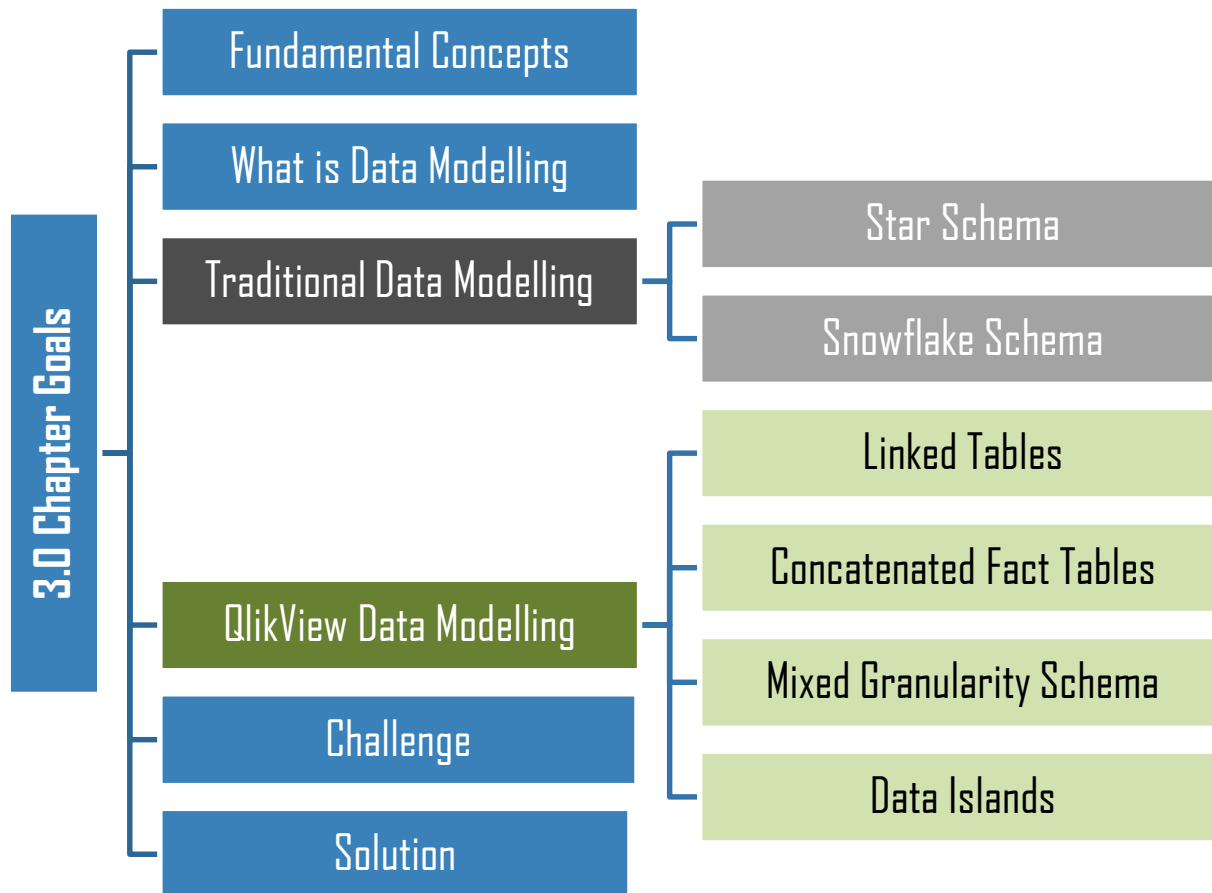 - Concatenated Fact Tables
  - Mixed Granularity Schema
  - Data Islands
- Challenge
- Solution

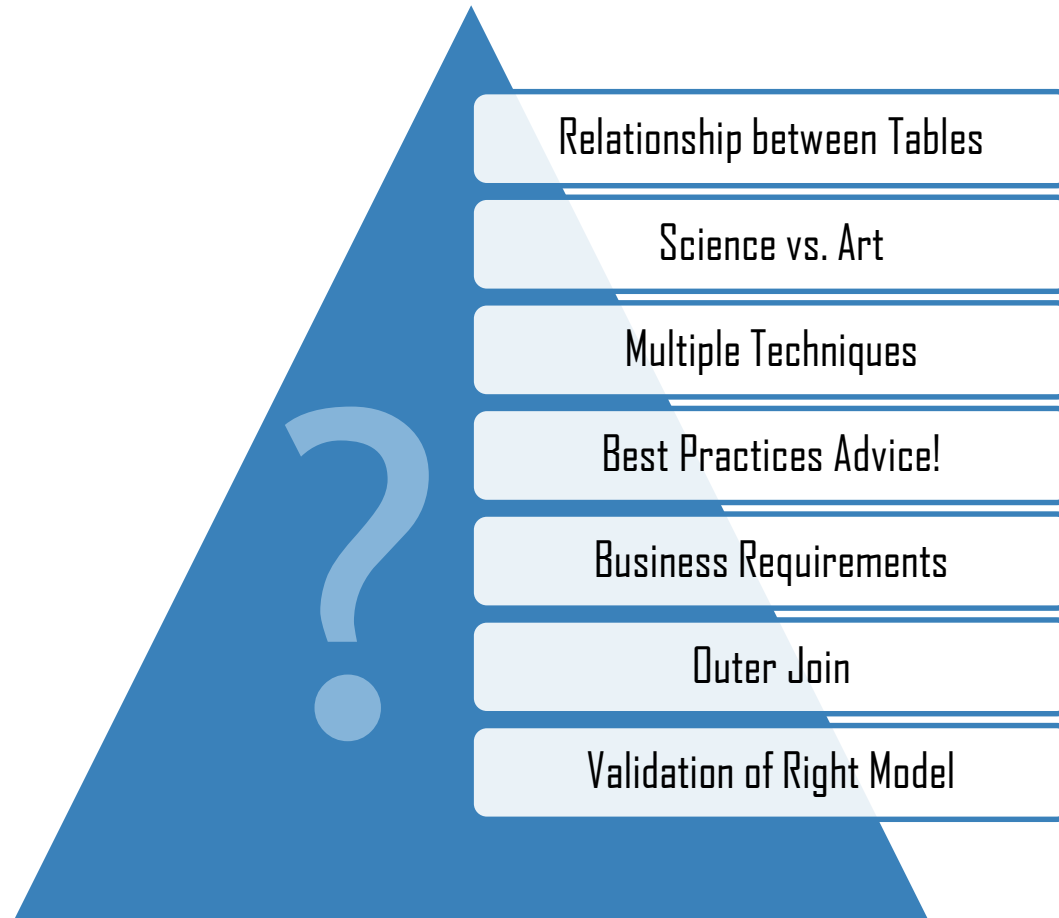## Conceptual Asides:

- OLAP vs Associative Model

- Dimensional vs Relational Modelling

- Schema

- Cardinality

- Normalization

- Metadata

- Primary Key

- Subset Ratio, Information Density and Perfect Key

- Synthetic Keys / Tables

- Circular Reference

- Importance of Outer Joins

What is Data Modelling?

Relationship between Tables

Science vs. Art

Multiple Techniques

Best Practices Advice!

Business Requirements

Outer Join

Validation of Right Model
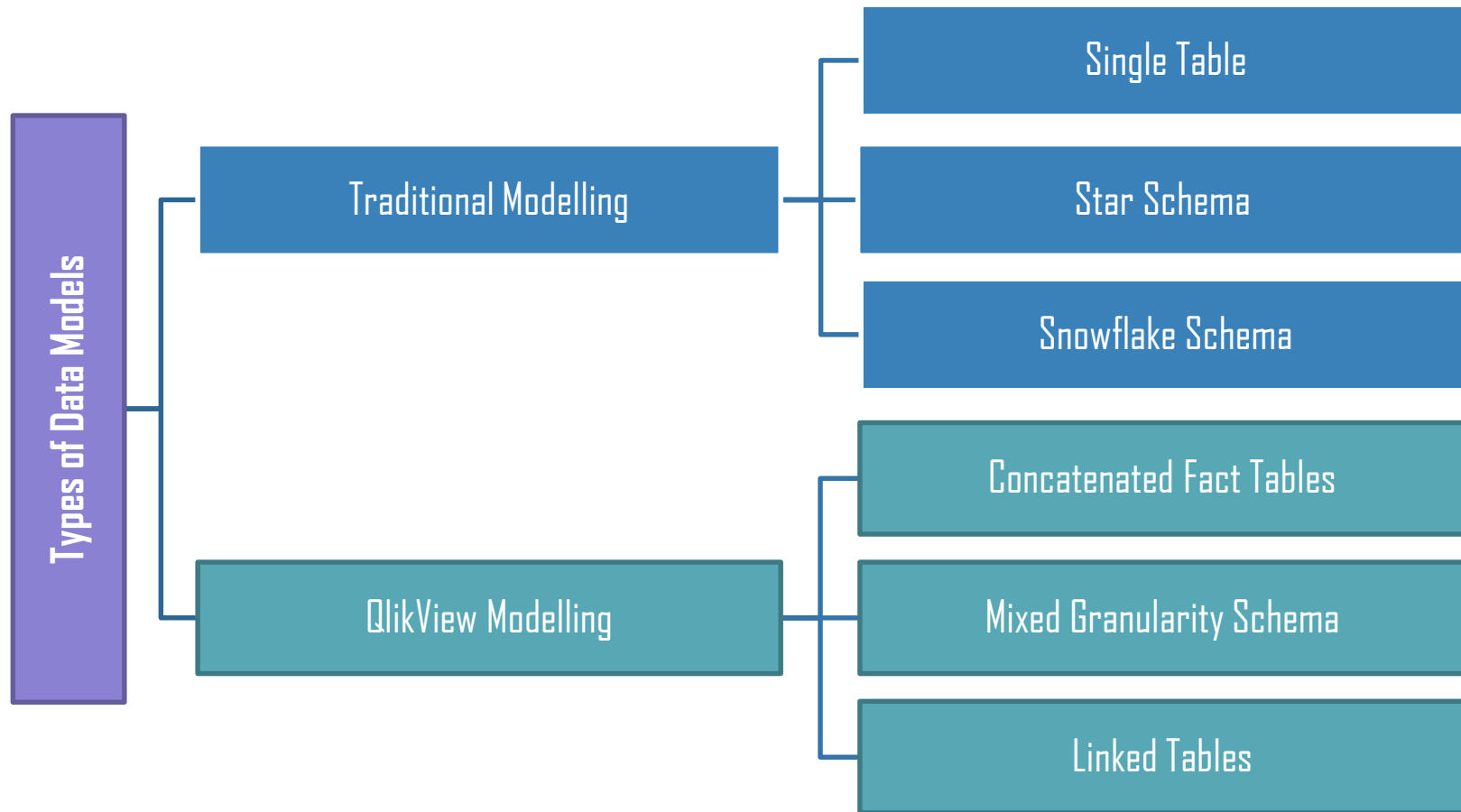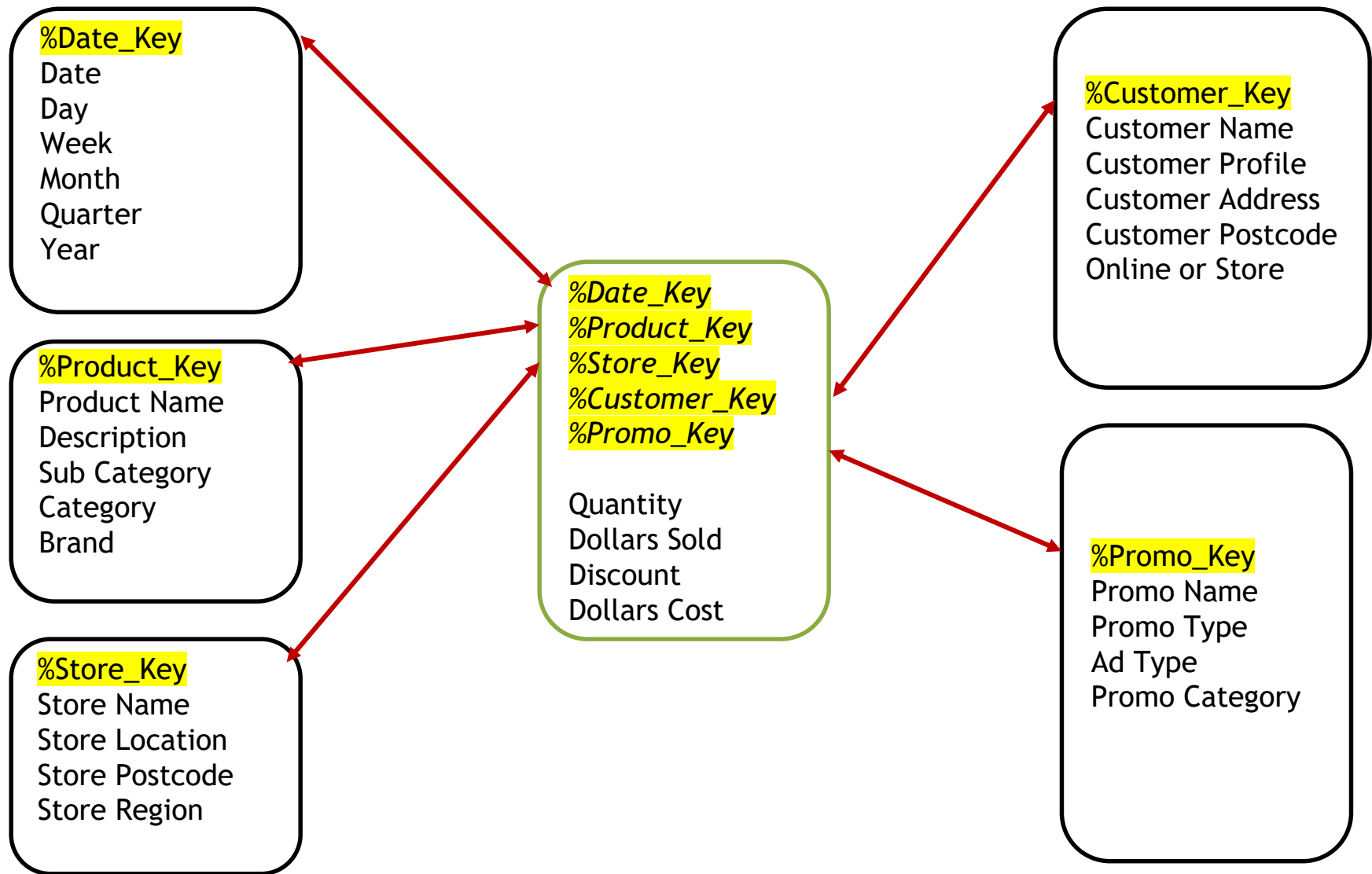
Before you start Data Modeling:

- Understand and define the granularity of the data

- Identify and separate Dimension(s) from Fact(s)

- Check if multiple Fact tables exist and wherever possible – Join or Concatenate

- Identify the potential Measures and map with the user requirements

- Identify Join Keys between Fact tables and Dimension tables

- Check if there are any Slowly Changing Dimensions (SCD)

- User requirements and user experience (UX) will determine the data model

3.3: Types of Data Model

Types of Data Models

- **Traditional Modelling**
  - Single Table
  - Star Schema
  - Snowflake Schema
- **QlikView Modelling**
  - Concatenated Fact Tables
  - Mixed Granularity Schema
  - Linked Tables

%Date_Key
Date
Day
Week
Month
Quarter
Year

%Product_Key
Product Name
Description
Sub Category
Category
Brand

%Store_Key
Store Name
Store Location
Store Postcode
Store Region

%Date_Key
%Product_Key
%Store_Key
%Customer_Key
%Promo_Key

Quantity
Dollars Sold
Discount
Dollars Cost

%Customer_Key
Customer Name
Customer Profile
Customer Address
Customer Postcode
Online or Store

%Promo_Key
Promo Name
Promo Type
Ad Type
Promo Category

Rules:

- Only one Fact Table

- Only one Field links between Dimensions and Fact Table

- Relationship can be One-to-One or One-to-Many between Dimension and Fact

- Every Dimension Table must contain a Primary Key

- Dimension Table will not have a Parent or Child Table

Advantages:

- Query performance

- Load performance and administration

- Easily understood

What is Snowflake Schema:

- Extension to Star Schema

- Normalized Dimension Tables

- Multiple Nodes

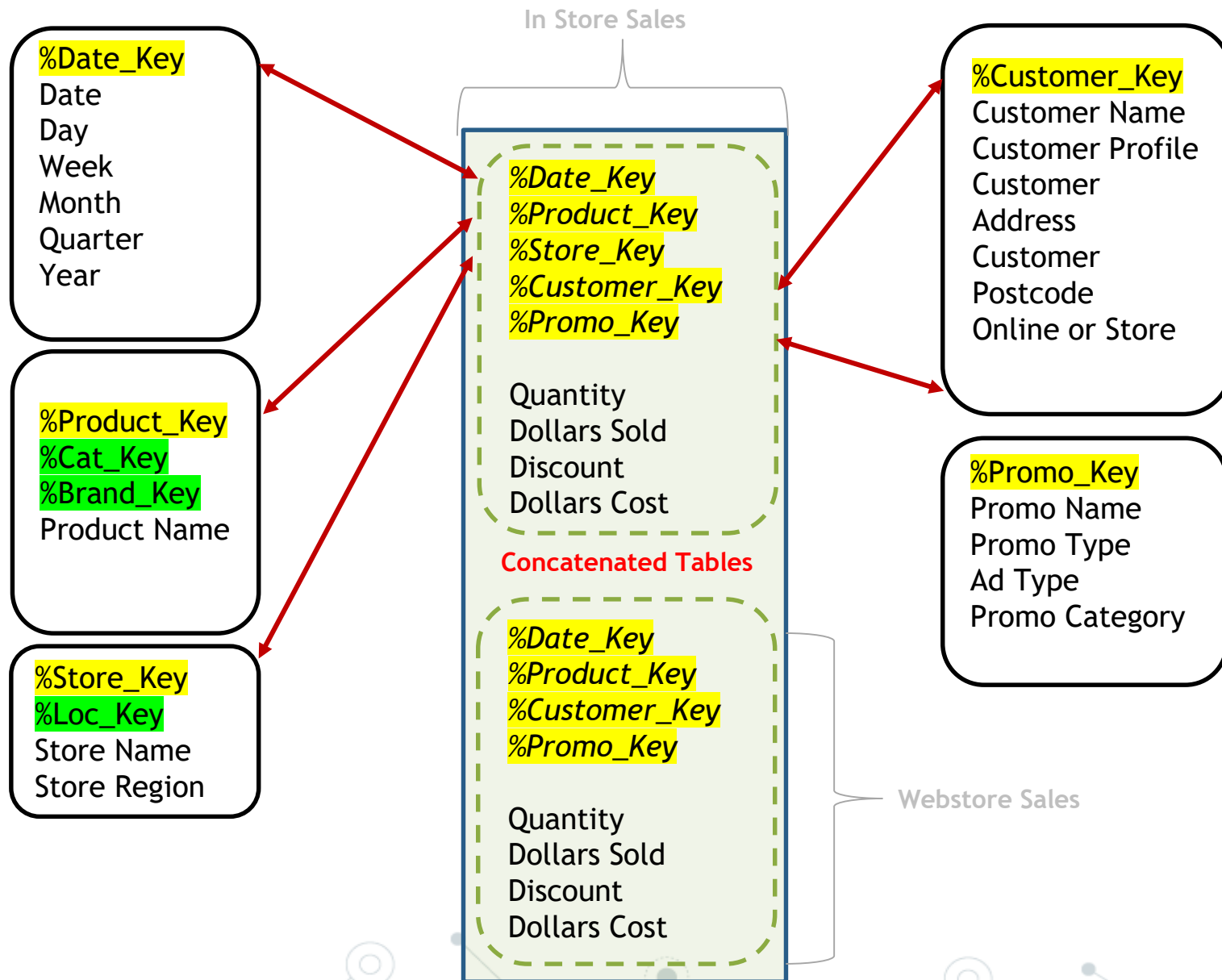- Dimension Table will have Parent or Child Table

Date dimension table:
- %Date_Key
- Date
- Day
- Week
- Month
- Quarter
- Year

Customer dimension table:
- %Customer_Key
- Customer Name
- Customer Profile
- Customer Address
- Customer Postcode
- Online or Store

Category dimension table:
- %Cat_Key
- Category
- Sub Category

Brand dimension table:
- %Brand_Key
- Brand Name

Product dimension table:
- %Product_Key
- %Cat_Key
- %Brand_Key
- Product Name

Fact table:
- %Date_Key
- %Product_Key
- %Store_Key
- %Customer_Key
- %Promo_Key
- Quantity
- Dollars Sold
- Discount
- Dollars Cost

Promo dimension table:
- %Promo_Key
- Promo Name
- Promo Type
- Ad Type
- Promo Category

Location dimension table:
- %Loc_Key
- Lat
- Long
- Location
- Postcode
- Address

Store dimension table:
- %Store_Key
- %Loc_Key
- Store Name
- Store Region

3.8: Concatenated Fact Tables

In Store Sales

%Date_Key
Date
Day
Week
Month
Quarter
Year

%Product_Key
%Cat_Key
%Brand_Key
Product Name

%Store_Key
%Loc_Key
Store Name
Store Region

%Customer_Key
Customer Name
Customer Profile
Customer Address
Customer Postcode
Online or Store

%Promo_Key
Promo Name
Promo Type
Ad Type
Promo Category

%Date_Key
%Product_Key
%Store_Key
%Customer_Key
%Promo_Key

Quantity
Dollars Sold
Discount
Dollars Cost

Concatenated Tables

%Date_Key
%Product_Key
%Customer_Key
%Promo_Key

Quantity
Dollars Sold
Discount
Dollars Cost

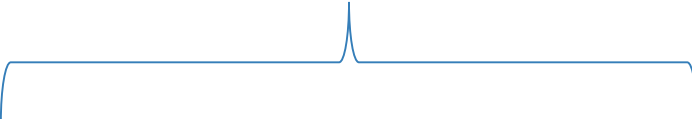**Webstore Sales**

Rules:

- Not more than one Fact Table (Multiple Flat Files and Fact data from multiple sources)

- Implicit vs Explicit Concatenation

- If Field Names are different – rename the Fields and then Concatenate the Tables

- Always identify the tables with flags/tags – 'Web Sales' and 'In Store' etc.

- Perform data transformations after Concatenation of Tables
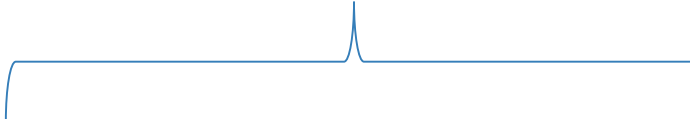
3.9: Mixed Granularity Schema

- Multiple FACT tables with different granularity (Aggregated and Transactional)
- Budget vs. Actual, Web Sales vs Retail Sales and Previous Year Aggregated vs Current Year Transactional data etc.

Granularity: Transactional Level

CustomerID	ProductID	OrderDate	Amount
32	12	18-Apr-17	28890
32	2	20-May-17	16160
40	7	21-May-17	75024
89	18	25-May-17	47660
32	1	02-Jun-17	17552
40	19	07-Jun-17	2018
32	1	30-Jun-17	86000
32	5	02-Jun-17	5000
40	5	07-Jun-17	25000
32	3	30-Jun-17	42586
89	15	25-May-17	2575
32	15	02-Jun-17	5420

Granularity: Aggregated to Month

Country	CategoryID	OrderMonth	Amount
UK	Cat_1	2017-05	6696
UK	Cat_5	2017-05	6910
UK	Cat_7	2017-05	57450
Germany	Cat_2	2017-06	208530
France	Cat_5	2017-06	70956
Germany	Cat_5	2017-06	235560
France	Cat_2	2017-04	274280

2 Cents:

- User requirements must drive the data model and app design approach
- Best practices are great but if your users reject the application, then it is failure regardless of technical execution
- Consistency beats Coolness
- Most of the traditional data warehousing techniques are applicable in the QlikView world
- Always test your model before publishing the app(s)