# Data Warehousing Fundamentals

WISE LTD.

**COURSE CODE:** DWHFUN-G

**COURSE TITLE:** Data Warehousing Fundamentals

**FEATURES:** Dimensional modelling per Ralph Kimball.

**PLATFORM:** **Multi-platform:** Oracle, SQL Server, DB2 for z/OS, DB2 for Linux/UNIX/Windows, Sybase, etc.

**AUDIENCE:** Would-be data warehouse architects, IT developers, database administrators, DW project managers or anyone responsible for a data warehouse or related discipline.

**PREREQUISITES:** At least six (6) months in a relational database environment.

**DURATION:** 3 days

**SUMMARY:** This dimensional modelling techniques course is designed to answer questions, such as the following:

- What is data warehousing?
- What is a data mart?
- What are the data modelling options?
- What is Extract, Transform and Load (ETL)?
- What are the terms and concepts specific to data warehousing and OLAP design?
- How to plan and implement a data warehouse with high availability, simplified manageability and optimal performance
- What are common statistics, analytic and OLAP SQL queries?

**OBJECTIVES:** Upon completion of this course, the participant should be able to design a data warehouse using both star and snowflake schemas. And the delegate should understand the implication of such terms as cubes, dimensions, attributes, joins, hierarchies, measures, etc.

# Data Warehousing Fundamentals
# Core Topics

WISE LTD.

## 1. DATA WAREHOUSING OVERVIEW
- Overview
- Typical uses
- Architecture

## 2. DEFINITION, ARCHITECTURE AND CONCEPTS
- Enterprise Data Model
- Operational vs. historical data
- Extract Transform Load (ETL)
- Metadata
- Data warehouse vs. data mart
- Data mining
- OLAP vs. OLTP
- Logical design vs. physical design
- Normalization vs. denormalization
- Referential constraints

## 3. DATA MODELLING OPTIONS
- Entity model
- Star schema
- Snowflake schema

## 4. DIMENSIONAL MODELLING DEVELOPMENT LIFE CYCLE
- Four steps of dimensional modelling
- Requirements analysis
- Requirements gathering
- Requirements validation
- Requirements modelling
- Schema design
- Project definition
- Warehouse design
- Implementation
- Follow-up and review

## 5. DIMENSIONAL MODELLING DESIGN
- Overview
- Metadata properties
- Star schema
- Snowflake schema
- Cubes
- Measures and facts
- Attributes and relationships
- Dimensions
- Hierarchies
- Joins
- Summary tables and aggregation (i.e., materialized views/indexed views)

## 6. IMPLEMENTATION OPTIONS

- Overview
- Top down
- Bottom up
- Sizing
- Cleaning
- Populating the data warehouse

## 7. EXTRACT, TRANSFORM, LOAD (ETL)

- ETL vs. ELT: pros and cons
- ETL planning and monitoring
- Transformation options
- Loading options
- Change Data Capture and publishing
- Data Staging
- Restart recovery

## 8. EXTRACTING

- Logical-to-physical data mapping
- Disparate (heterogeneous) data sources
- Extracting changes data – delta or other
- Data profiling

## 9. DATA CLEANING & CONFORMING

- Data quality criteria
- Design methods and alternatives
- Cleaning deliverables
- Conforming dimension tables
- Conforming fact tables

## 10. DIMENSION TABLE DELIVERY

- Dimension table structure
- Surrogate key generation
- Dimension table grain
- Flat (denormalized) or snowflake?
- Date and time dimensions
- ‘Big’ vs. ‘small’ dimensions
- Dimensional roles
- Dimensions as subdimensions
- Degenerate dimensions

 

## 11. SLOWLY CHANGING DIMENSIONS

- Type 1
- Type 2
- Type 3
- Hybrid
- Inferred members/late arriving dimensions
- Early arriving facts

## 12. MULTIVALUED DIMENSIONS

- Definition
- Bridge tables

## 13. FACT TABLE DELIVERY

- Fact table structure
- Referential integrity (RI)
- Surrogate key derivation and flow
- Fundamental grain
- Transaction fact tables
- Factless fact tables
- Periodic snapshots
- Accumulating snapshots

## 14. FACT TABLE LOAD CONSIDERATIONS

- Index management
- Partition management
- Updates, deletes and inserts
- Recovery
- Summary tables
- Parallelism

## 15. DATA WAREHOUSE PERFORMANCE DESIGN

- Materialized views
- Large concurrent reports
- Short running queries
- Long running queries
- Random queries
- Occasional updates
- On-line utilities
- Index options
- Partitioning and parallelism (e.g., LOADs)

 

## 16. PHYSICAL DESIGN CONSIDERATIONS

- Denormalization
- Index choices
- Data placement
- Free space
- Summary tables
- Data compression

## 17. DATA ANALYSIS & REPORTING

- BI User Types
- Query and reporting concepts
- OLAP and cubes
- OLAP architecture – MOLAP, ROLAP, HOLAP
- Multidimensional analysis, e.g., slicing, dicing, drill-down, etc.
- GROUPING, ROLLUP & CUBE
- OLAP tools

## 18. CASE STUDY

- Project definition and scoping
- Specify the requirements
- Specify the grain (e.g., fact table types)
- Specify the dimensions (e.g., handling slowly changing dimensions)
- Specify the facts (e.g., conformed facts)

 