# DB2 Data Warehouse Fundamentals

WISE LTD.

**COURSE CODE:** DWHFUN

**COURSE TITLE:** DB2 Data Warehouse Fundamentals

**AUDIENCE:** IT developers, database administrators or anyone responsible for a data warehouse or related discipline.

**PREREQUISITES:** At least six (6) months in a DB2 environment or its equivalent.

**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 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?
- What is Extract, Transform, Load (ETL)?

**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.

# DB2 Data Warehouse Fundamentals
# Core Topics

WISE LTD.

## 1. DATA WAREHOUSE OVERVIEW
- Overview
- Typical uses

## 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
- Massive size implementation
- 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
- Needs assessment
- Project definition
- Interview and requirement gathering
- Data modelling
- Warehouse design
- Implementation
- Follow-up and review

## 5. MULTIDIMENSIONAL DESIGN
- Overview
- Metadata properties
- Star schema
- Snowflake schema
- Measures and facts
- Attributes and relationships
- Dimension
- Hierarchies
- Joins
- Summary tables and aggregation
- Case study

## 6. ETL – EXTRACTION

- Integrating heterogeneous data sources
- Mainframe sources
- Flat files
- XML sources
- Web log sources
- ERP system sources
- Extracting changed data

## 7. ETL – DATA CLEANING AND CONFORMING

- Data source types
- Data destination types
- Aggregates
- Audits
- Conditional splits
- Copy columns
- Data conversion
- Derived columns
- Export columns
- Lookup
- Merge and merge joins
- Multicast
- Sampling
- Pivot and unpivot

## 8. ETL DESTINATION – DIMENSION TABLES

- Basic structure of a dimension
- Grain of a dimension
- Basic load plan for a dimension
- Flat dimensions and snowflaked dimensions
- Date and time dimensions
- Big dimensions
- Small dimensions
- One dimension or two
- Dimensional roles
- Dimensions as subdimensions of another dimension
- Degenerate dimensions
- Slowly changing dimensions
- Type 1 slowly changing dimension
- Type 2 slowly changing dimension
- Type 3 slowly changing dimension
- Precise time stamping of a Type 2 slowly changing dimension

 

## 9. ETL DESTINATION – FACT TABLES

- Basic structure of a fact table
- Guaranteeing referential integrity
- Surrogate key pipeline
- Fundamental Grains
- Preparing for loading fact tables
- Factless fact tables
- Augmenting a Type 1 fact table with Type 2 history
- Graceful modifications
- Multiple units of measure in a fact table
- Collecting revenue in multiple currencies
- Late arriving facts
- Aggregations

## 10. PHYSICAL DESIGN CONSIDERATIONS

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

## 11. DATA WAREHOUSE PERFORMANCE DESIGN

- Automatic Summary Tables (AST)
- Large concurrent reports
- Short running queries
- Long running queries
- Random queries
- Occasional updates
- On-line utilities
- Partitioning and parallelism (e.g., LOADs)
- Query Patroller
- Configuration parameters
- Table spaces and buffer pools

## 12. QUERY PERFORMANCE

- DB2 compiler
- Join strategies
- Sort and aggregation parallelism

 

## 13. HIGH AVAILABILITY

- Multidimensional Clustering (MDC)
- Online loads
- Load from cursor
- Batch window elimination
- Elimination of table reorganisation
- Online load and MQT maintenance
- MQT staging tables
- Online table reorganisation
- Dynamic bufferpool management
- Dynamic database configuration
- Database managed storage considerations
- Logging considerations

## 14. INTRODUCTION TO ANALYTIC AND OLAP SQL QUERIES

- AVG
- CORRELATION
- COUNT
- COUNT_BIG
- MAX
- MIN
- RAND
- STDDEV
- SUM
- GROUPING, ROLLUP & CUBE

 