



# Data Warehousing



## Corporate Trainer's Profile

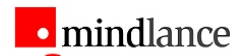
Corporate Trainers are having the experience of 4 to 12 years in development , working with TOP CMM level 5 companies (Project Leader /Project Manager ) qualified from NIT/IIT/IIM and work exp in USA and UK.



## CMM (Capability Maturity Model) level Project Standard:-

The Capability Maturity Model (CMM) is a method for evaluating the maturity of organizations on a scale of 1 to 5. Get the Opportunities to work on Client Projects Of US/UK, which follow the all standard of CMM level 5 Company.

## Projects



# Data Warehousing

**SUMMARY:** CURRENCY:  
9i & 10g

This course focuses on the extract, transform and load (ETL) phase of the data warehouse (DW) development life cycle. After discussing ETL as it relates to all DBMSs (DB2, MS Server, Sybase, etc.) we elaborate on Oracle's ETL offering:

- Change Data Capture
- SQL\*Loader
- Data Pump
- External tables
- Transportable tables
- SQL MERGE
- Table Functions
- Etc.

**AUDIENCE:** DW administrators, DBAs, data architects, DW PL/SQL developers.

**PREREQUISITES:** Completion of our course 'Data Warehouse Terms, Concepts & Architecture' (DWHHTCA) or equivalent knowledge.

**DURATION:** 3 days

**OBJECTIVES:** Upon completion of this presentation, the participant should understand the ETL phase of the data warehouse development life cycle.

## **COURSE CONTENT: I.REVIEW OF DATA WAREHOUSING (DW) TERMS AND CONCEPTS**

- The DW environment
- What is a data warehouse?
- What is a data mart?
- What is Business Intelligence (BI)?
- How do OLTP & OLAP differ?
- What is data mining?
- Operational vs. historical data
- What is a star schema?
- What is a snowflake schema?
- Normalization vs. denormalization
- What are hierarchies?
- What is dimensional modelling?
- What is the Data Warehouse Bus Architecture (DWB)?
- What are surrogate keys?
- What is Extract, Transform, Load (ETL)?
- What are Slowly Changing Dimensions (SCD)?
- What is Metadata?
- What Materialized Views (MV)?

How does logical design differ from physical design?

## **II.EXTRACT, TRANSFORM, LOAD (ETL) TERMS & CONCEPTS**

Options

Extraction options

Transformation options

Loading options

Change Data Capture and publishing

Staging areas

## **III.EXTRACTING**

Logical-to-physical data mapping

Disparate (heterogeneous) data sources

Extracting changes data delta or other

## **IV.DATA CLEANING & CONFORMING**

Data quality criteria

Design methods and alternatives

Cleaning deliverables

Conforming dimension tables

Conforming fact tables

## **V.DIMENSION TABLE DELIVERY**

Dimension table structure

Surrogate key generation

Dimension table grain

Flat (denormalized) or snowflake?

Data and time dimensions

'Big' vs. 'small' dimensions

Dimensional roles

Dimensions as subdimensions

Degenerate dimensions

## **VI.SLOWLY CHANGING DIMENSIONS**

Type 1

Type 2

Type 3

Hybrid

Late arrivals

## **VII.MULTIVALUED DIMENSIONS**

Definition

Bridge tables

## **VIII.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

## **IX.FACT TABLE LOAD CONSIDERATIONS**

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

## **X.ETL TOOLS**

### **XI.ORACLE'S DATA EXTRACTION**

- Synchronous Change Data Capture (CDC)
- Asynchronous Change Data Capture (CDC)

### **XII.ORACLE'S DW LOADING OPTIONS**

- SQL\*Loader
- Optimising SQL\*Loader performance
- SQL\*Loader Direct Path Load
- SQL\*Loader partitioning considerations
- SQL\*Loader and data constraints (e.g., RI)
- SQL\*Loader and parallelism
- SQL\*Loader transformation options
- SQL\*Loader and index optimisation
- Oracle's Data Pump
- External tables
- Transportable tablespaces
- Using SQL MERGE
- Multiple table INSERTs

### **XIII.ORACLE DATA TRANSFORMATIONS & CLEANING**

- SQL updates
- Regular expressions
- Data validation
- DW key lookups
- Table functions
- Moving data from staging to fact
- Exchanging partitions
- Direct path inserts
- CREATE TABLE AS ...

### **XIV.DW PERFORMANCE & GOOD PRACTICES**

- Query rewrite and MVs
- Star vs. snowflake schema
- SQL aggregates (e.g., ROLLUP)
- Parallelism
- Partitioning options
- Index options