

INSTRUCTOR-LED TRAINING COURSE

TERADATA PHYSICAL DATABASE DESIGN

Lecture/Lab

ILT 25968

4 Days

COURSE DESCRIPTION

This course defines the processes and procedures to follow when designing and implementing a Teradata system. It covers Teradata data distribution, access, and use of derived data. Similarities between join and aggregation processing, and the implementation of Referential Integrity are also discussed. Various compression forms are explained in detail. Extensive hands-on labs help reinforce learning.

AUDIENCE

- ~ Database Administrators
- ~ Architects/Designers
- ~ Application Developers

PREREQUISITES

To get the most out of this training, you should have the following knowledge or experience.

- ~ **Introduction to Teradata Database** (ILT #25964 or WBT #26438)
- ~ **Teradata SQL** (Instructor-led (#25965) or WBT (#54458))
- ~ Familiarity with Relational Database Modeling

COURSE OBJECTIVES

After successfully completing this course, you will be able to:

- ~ Distinguish a logical data model from a physical database model
- ~ Explain data distribution and hashing on a Teradata platform
- ~ Describe the structure and characteristics of primary and secondary indexes
- ~ Describe the Teradata file system and how space limits are defined and enforced
- ~ Describe the costs and benefits of Partitioned Primary Indexes and table constraints, including the Referential Integrity constraint
- ~ Determine why, where, and when Statistics should be collected, and how they help the Optimizer develop efficient join plans
- ~ Describe the options available in the CREATE TABLE and CREATE INDEX statements, and how they affect system performance
- ~ Compare the costs and benefits of denormalization

COURSE OUTLINE*

DAY 1	DAY 2	DAY 3	DAY 4
~ Introduction	~ File System Writes	~ Secondary Index Usage	~ Referential Integrity
~ Introduction Physical Database Design Overview	~ DBS Control Settings	~ Partitioned Primary Indexes	~ Optimizer and Collecting Statistics
~ Building the Database Environment	~ Analyze Primary Index Criteria	~ Compression Forms	~ Join Processing Analysis
~ Data Distribution and Hashing	~ Tables without a Primary Index	~ Access Considerations and Constraints	

* Timing and topics covered by day may vary

COURSE CONTENT

Module 0 - Introduction

- ~ Recommended Prerequisite Knowledge
- ~ Course Objectives
- ~ Course Modules

Module 1 - Physical Database Design Overview

- ~ Data Warehouse Usage Evolution
- ~ The Stages of Database Development
- ~ Customer Service Logical Model – ERA Methodology Diagram
- ~ Different Models, Different Terms
- ~ Attributes
- ~ Entities and Relationships in the LDM
- ~ Normal Forms
- ~ Normalization
- ~ Database Design Components
- ~ Understanding Change Rating
- ~ Understanding Access
- ~ Extended Logical Data Model
- ~ Physical Data Model
- ~ The Principles of Index Selection
- ~ Review Questions

Module 2 - Building the Database Environment

- ~ The Teradata Hierarchy
- ~ Owners, Parents, and Children
- ~ Creating Objects
- ~ Permanent Space Terminology
- ~ Spool and Temp Space Terminology
- ~ Assigning Space Limits
- ~ Summary
- ~ Review Questions
- ~ Lab Exercises

Module 3 - Data Distribution and Hashing

- ~ Data Distribution
- ~ Hashing
- ~ Primary Index Hash Mapping
- ~ Hash Maps
- ~ Primary Hash Map
- ~ Hash Maps for Different Systems
- ~ Fallback Hash Map
- ~ Reconfiguration
- ~ MAPS TD 16.10 Reconfiguration
- ~ MAPS TD 16.10 Sparse Maps
- ~ Hash Related Expressions
- ~ Hashing – Numeric Data Types
- ~ Multi-Column Hashing
- ~ Additional Hash Examples
- ~ Using Hash Functions to View Distribution

- ~ Row Retrieval via PI Value – Overview
- ~ Names and Object IDs
- ~ Table ID
- ~ Row ID
- ~ AMP File System – Locating a Row via PI
- ~ Accessing the Row within the Data Block
- ~ General Row Layout
- ~ AMP Read I/O Summary
- ~ Review Questions
- ~ Lab Exercises

Module 4 - File System Writes

- ~ AMP Write I/O
- ~ New Row INSERT
 - ~ Master Index Access
 - ~ Cylinder Index Access
 - ~ Use Free Space within Block
 - ~ Expand the Block
- ~ Block and Row Terminology
- ~ Block Splits (INSERT and UPDATE)
- ~ Cylinder Full
- ~ Mini-Cylpack
- ~ Space Utilization
- ~ 1MB Data Blocks
- ~ 1MB Table Headers (TD14.10)
- ~ Summary
- ~ Review Questions

Module 5 - DBS Control Settings

- ~ System Level Defaults
- ~ DBS Control Record – General Fields
- ~ MaxDecimal
- ~ Date and Time Parameters
- ~ Primary Index Defaults
- ~ DBS Control Record – File System Fields
- ~ Free Space Percent
- ~ MiniCylPackLowCylProd
- ~ DefragLowCylProd
- ~ PermDBSize
- ~ PermDBAllocUnit
- ~ JournalDBSize
- ~ Cylinders Saved for Perm
- ~ Merge Datablocks
- ~ Teradata 14.10 – 1MB Spool Rows
- ~ DBS Control Record – Performance Fields
- ~ DisableSyncScan
- ~ ReDistBufSize
- ~ Support of Hash Joins
- ~ DBS Control Record – Compression
- ~ Enabling Compression
- ~ Controlling Compression Operations

- ~ Temperature Based Block Level Compression
- ~ Summary

Module 6 - Analyze Primary Index Criteria

- ~ Primary Index Choice Criteria
- ~ CREATE TABLE – Indexing Rules
- ~ Primary Index Characteristics
- ~ Multi-Column Primary Indexes
- ~ Primary Index Considerations
- ~ PKs and Duplicate Rows
- ~ Primary Index Demographics
- ~ Column Distribution Demographics for a PI Candidate
- ~ SQL to View Data Demographics
- ~ TableSize View
- ~ SQL to View Data Distribution
- ~ Summary
- ~ Review Questions
- ~ Lab Exercises

Module 7 - Tables without a Primary Index

- ~ What is a NoPI Table?
- ~ Reasons to Consider Using NoPI Tables
- ~ Creating a Table without a PI
- ~ Distribution of Rows (or Blocks) for a NoPI Table
- ~ The Row ID for a NoPI Table
- ~ Loading Data into a NoPI Table (SQL)
- ~ Loading Data into a NoPI Table (Utility)
- ~ Archive and Reconfig Issues
- ~ NoPI Table Options
- ~ Summary
- ~ Review Questions

Module 8 - Secondary Index Usage

- ~ Secondary Indexes
- ~ Choosing a Secondary Index
- ~ Secondary Index Subtables
- ~ USI Subtable General Row Layout
- ~ USI Hash Mapping
- ~ NUSI Subtable General Row Layout
- ~ NUSI Hash Mapping
- ~ Secondary Index Considerations
- ~ Single NUSI Access (Between, Less Than, or Greater Than)
- ~ Dual NUSI Access
- ~ EXPLAIN of ORed NUSI's
- ~ NUSI Bit Mapping
- ~ EXPLAIN of a SELECT (BMSMS)
- ~ Value-Ordered NUSIs
- ~ Hash-Ordered NUSIs
- ~ Hash-Ordered NUSI Example

- ~ Covering Indexes
- ~ NUSI vs. Full Table Scan (FTS)
- ~ Accessing Rows
- ~ Guidelines for Typical Behaviour
- ~ Secondary Index Considerations
- ~ Secondary Index Usage
- ~ Summary
- ~ Review Questions

Module 9 - Partitioned Primary Indexes

- ~ Partitioned Primary Indexes (PPI)
- ~ Logical Example of Row Partitioning
- ~ How is Partitioning Implemented?
- ~ PPI Details
- ~ Primary Index Access (Non-partitioned Table)
- ~ Primary Index Access (Row Partitioned Table)
- ~ Simple Partitioning Example
- ~ Partitioning with CASE_N and RANGE_N
- ~ Special Partitions with CASE_N and RANGE_N
- ~ NO CASE (NO RANGE) or UNKNOWN
- ~ TIMESTAMP PPI
- ~ SQL Use of PARTITION Key Word
- ~ SQL Use of CASE_N
- ~ Using ALTER TABLE
 - ~ With Row Partitioned Tables
 - ~ Populated Tables
 - ~ NO RANGE is not Defined
 - ~ NO RANGE is Defined
 - ~ WITH DELETE
 - ~ DROP/ADD
 - ~ ADD Clause on Partitioning Expression
 - ~ ALTER TABLE TO CURRENT
- ~ EXPLAIN Terminology for PPI Tables
- ~ Example 6 – Partition Elimination with a PPI Table
- ~ Multilevel Partitioning Concepts
- ~ How is the MLPPI Partition # Calculated?
- ~ Summary
- ~ Review Questions
- ~ Lab Exercises

Module 10 - Compression Forms

- ~ Teradata Data Types
- ~ Character Sets
- ~ Table Headers
- ~ Presence Bits
- ~ Multi-value Compression
- ~ Presence Bits and Compression
- ~ Implementing Multi-value Compression
- ~ ALTER TABLE and Compression
- ~ Multi-value Compression Considerations
- ~ Algorithmic Compression
- ~ Block Level Compression Options
- ~ Temperature-Based Block Level Compression
- ~ Defining Temperature Data
- ~ Advantages of Temperature-Based Compression
- ~ BLC Groups
- ~ DDL Compression Options
- ~ Compression Capabilities
- ~ SQL Interface for FERRET
- ~ DBC.CreateFsysInfotable – SHOWBLOCKS
- ~ SHOWBLOCKS (Short) Output Example
- ~ DBC MACROS with SHOWWHERE
- ~ DBC.PopulateFsysInfoTable Content Examples
- ~ Summary
- ~ Review Questions
- ~ Lab Exercises

Module 11 - Access Considerations and Constraints

- ~ Access Method Comparison
- ~ Optimizer Access Scenarios
- ~ Partial Value Searches
- ~ Storing Numeric Data
- ~ Data Conversions
- ~ Matching Data Types
- ~ Counting I/O Operations
- ~ Transient Journal I/O
- ~ Table Level Attributes
- ~ Column Level Constraints
- ~ Table Level Constraints
- ~ Altering Table Constraints
- ~ Summary
- ~ Review Questions
- ~ Lab Exercises

Module 12 - Referential Integrity

- ~ Referential Integrity Overview
- ~ Parent-Child Relationships
- ~ Three Types of Referential Constraints
- ~ Processing Cost of Referential Constraints
- ~ Fixing Referential Integrity Problems

- ~ Standard Referential Integrity
- ~ Batch Referential Integrity
- ~ Soft Referential Integrity
- ~ Referential Integrity
- ~ Join Optimization with RI
- ~ Summary
- ~ Review Questions
- ~ Lab Exercises

Module 13 - Optimizer and Collecting Statistics

- ~ Teradata Optimizer
 - ~ Optimizer – Cost Based vs. Rule Based
 - ~ Optimizer Statistics
 - ~ Optimizer's Search for Statistics
 - ~ Optimizer – Dynamic AMP Samples
- ~ Optional Adjustments for Dynamic AMP Samples
- ~ Dynamic AMP Sampling – How it Works
- ~ Optimizer Estimate without Collected Statistics
- ~ Statistics
 - ~ Statistics Data – What is Collected?
 - ~ COLLECT STATISTICS Command
 - ~ Collecting Statistics
 - ~ Refresh or Re-Collect Statistics
 - ~ COLLECT STATISTICS Command (Index Format with 14.0 Options)
 - ~ COLLECT STATISTICS on a Data Sample
 - ~ Collecting Statistics – 14.0 Examples
 - ~ Statistics
 - ~ Optimizer's Use of Statistics with Uneven NUSI
 - ~ Collecting Statistics on PARTITION
 - ~ Copying STATISTICS
 - ~ Statistics Extrapolation
 - ~ Additional Statistics Options
- ~ Teradata 14.0 Enhancements
- ~ Teradata 14.10 Enhancements
- ~ Teradata 14.10 Re-Collect Enhancements
- ~ COLLECT STATS Using THRESHOLD Options
- ~ Teradata 14.10 AutoStats
- ~ Stats Manager
- ~ Summary
- ~ Review Questions
- ~ Lab Exercises

Module 14 - Join Processing Analysis

- ~ Join Types vs. Join Strategies
- ~ Product Join
- ~ Product Join Explain
- ~ Product Joins
- ~ Merge Join with Duplication and Sorting
- ~ Merge Join Strategy
- ~ Merge Join with Row Redistribution
- ~ Merge Join with Redistribution Explain
- ~ Merge Join with Matching Primary Indexes
- ~ Merge Join on Matching Primary Indexes Explain
- ~ Alternative to Using OR Join Conditions
- ~ Exclusion Joins
- ~ Exclusion Merge Join Explain
- ~ Inclusion Merge Join Explain
- ~ Hash Joins
- ~ n-Table Joins
- ~ Three-Table Join Explain
- ~ Partial Redistribution and Partial Duplication (PRPD)
- ~ PRPD Join Steps Simplified
- ~ 14.10 EXPLAIN Terminology & Modifiers
- ~ Nested Joins
- ~ Nested Join Explain
- ~ Join Considerations with Partitioned Tables
- ~ NPPI to PPI Join – Many Partitions
- ~ NPPI to PPI Join – Sliding Window
- ~ NPPI to PPI Join – Hash Ordered Spool File Join
- ~ PPI to PPI Join – Rowkey-Based Join
- ~ PPI to PPI Join – Unmatched Partitions
- ~ Join Processing Observations
- ~ Index Choice Guidelines
- ~ Summary
- ~ Review Questions
- ~ Lab Exercises