Data Modelling for Analytics and Business Intelligence

Duration: 1 Week

Day 1: Data Modelling Fundamentals & Relational Concepts

Module 1: Data Modelling Concepts & ER Modelling

- Conceptual, logical, and physical data models
- Entity-Relationship (ER) modelling: entities, attributes, relationships
- Cardinality: one-to-one, one-to-many, many-to-many
- Primary keys, foreign keys, and referential integrity
- Data modelling tools: ERwin, dbdiagram.io, Lucidchart

Lab Exercises:

1. Build a conceptual ER model for a sales or HR domain from a set of business requirements
2. Convert the conceptual ER model to a logical model with data types and key definitions

Module 2: Normalisation & Denormalisation

- Normalisation: 1NF, 2NF, 3NF, BCNF — rules and examples
- Anomalies: insertion, update, and deletion anomalies in unnormalised data
- Denormalisation: trading redundancy for query performance
- Partial and transitive dependencies: identifying and resolving
- Normalisation vs denormalisation trade-offs in analytics contexts

Lab Exercises:

3. Normalise a flat dataset to 3NF; document each step and explain anomalies resolved
4. Denormalise a 3NF model for a reporting use case; justify each denormalisation decision

Day 2: Dimensional Modelling for Analytics

Module 3: Dimensional Modelling Concepts

- Kimball vs Inmon approaches: bottom-up vs top-down
- Dimensional modelling concepts: facts, dimensions, grain
- Star schema vs snowflake schema: structure and trade-offs
- Fact table types: transaction, periodic snapshot, accumulating snapshot
- Degenerate dimensions, junk dimensions, role-playing dimensions

Lab Exercises:

5. Design a star schema for a retail sales domain: fact table + 5 dimensions with correct grain
6. Convert a star schema to snowflake; compare query complexity and storage trade-offs

Module 4: Dimension Table Design & Conformed Dimensions

- Dimension attributes: descriptive, hierarchical, and derived
- Surrogate keys vs natural keys in dimension tables
- Hierarchies in dimensions: product category, geography, time
- Conformed dimensions: sharing dimensions across multiple fact tables
- Enterprise data bus: aligning dimensions for cross-domain analytics

Lab Exercises:

7. Design a Customer dimension with hierarchies, surrogate key, and 12 descriptive attributes
8. Identify conformed dimensions across HR and Finance domains; design a shared Date dimension

Day 3: Slowly Changing Dimensions & Time Intelligence

Module 5: Slowly Changing Dimensions (SCD)

- SCD Type 1: overwrite — no history retained
- SCD Type 2: add new row — full history with effective dates and active flag
- SCD Type 3: add attribute column — limited history
- Hybrid SCDs: combining types for complex scenarios
- Choosing the right SCD type: business requirements and storage impact

Lab Exercises:

9. Implement SCD Type 2 for a Customer dimension: design table, write load logic, test history
10. Apply SCD Type 1 and Type 3 to two other dimensions; compare history behaviour

Module 6: Date Dimension & Time Intelligence

- Date dimension design: calendar date, fiscal date, working days
- Date dimension attributes: year, quarter, month, week, day, holiday flag
- Fiscal calendar: fiscal year, period, and quarter offsets
- Time dimension for intraday analytics
- Time intelligence patterns in BI: YTD, MTD, rolling averages, period comparison

Lab Exercises:

11. Build a comprehensive date dimension table covering 5 years with fiscal and holiday attributes
12. Write SQL time intelligence queries: YTD sales, MTD comparison, 3-month rolling average

Day 4: Modern Data Modelling with dbt

Module 7: dbt Project Structure & Layered Modelling

- dbt project structure: models, seeds, snapshots, tests, macros
- Layered modelling approach: staging, intermediate, mart layers
- Materialisation strategies: table, view, incremental, ephemeral
- Sources and ref() function: building dependency graphs
- dbt documentation: model descriptions, column descriptions, lineage DAG

Lab Exercises:

13. Set up a dbt project; build staging, intermediate, and mart models for a sales domain
14. Configure materialisation for each layer; run dbt docs generate and explore the lineage DAG

Module 8: dbt Testing, Incremental Models & Best Practices

- dbt schema tests: not_null, unique, accepted_values, relationships
- Custom singular tests for business rule validation
- Incremental models: is_incremental() filter and merge strategies
- dbt snapshots for SCD Type 2 using built-in snapshot functionality
- dbt best practices: naming conventions, modularity, DRY principles

Lab Exercises:

15. Add schema tests and write 2 custom singular tests; run dbt test and fix all failures
16. Build an incremental dbt model for a fact table; verify only new rows are processed on re-run

Day 5: BI-Ready Models, Performance & Capstone

Module 9: Optimising Models for BI & Analytics

- BI tool consumption patterns: Power BI, Tableau, Looker
- Calculated columns vs measures: modelling in warehouse vs BI tool
- Aggregate tables and pre-aggregation strategies
- Partitioning and clustering for analytical query performance
- Semantic layer: metrics, dimensions, and reusable business definitions

Lab Exercises:

17. Optimise a complex SQL model using partitioning and clustering; measure query cost improvement
18. Build a semantic layer definition for a sales mart; connect to Power BI and validate measures

Module 10: Capstone Project & Model Review

- End-to-end data model for a multi-domain analytics use case
- Requirement analysis: identifying facts, dimensions, grain, and metrics
- Model design: ER diagram, star schema, SCD strategy, dbt layers
- Implementation: dbt models, tests, documentation, and lineage
- Presentation: model walkthrough, design decisions, and trade-offs

Lab Exercises:

19. Design and implement a complete dimensional model for an assigned domain using dbt end-to-end
20. Present model to cohort: ER diagram, star schema, SCD choices, dbt lineage, and key metrics