Revision

PART I: Data Warehousing

Week 1: Basic Data Warehousing and SQL Revision

- Why data warehousing? Why not the operational database?
- What is DW used for? What is the operational DB used for?
- What is ETL? What is transformation?
- Why explore the data in the operational database?
- Why data cleaning?
- What is transaction? What is OLTP?
- What is OLAP?
- What is the relationship between OLAP and BI? What is BI?

- What is star schema? What is fact measure?
- How to identify fact measures? How to identify dimensions?
- Why fact measures must be numerical values?
- What is an aggregated value? Why fact measure must be an aggregated value?
- How to create a dimension? Why group by is used when creating the fact table?
- Where is the PK (FK)?

- SQL:
  - Create table, PK/FK
  - Insert into
  - Update, Delete
  - Select, Join
  - Group by, Count, Sum
  - Alter

Week 2: Simple Star Schema

- What is a "One-attribute" dim? What is the impact to the fact?
- What is the problem of AVG in the fact?

- What is a two-column table methodology? What is category?
- How can we check whether a simple star schema is correct or not?

- Why do we need a tempfact?
- Why create a dimension manually?
- Why update tempfact?
- What is the granularity of a dimension? What is the level of granularity?

Week 3: Simple Star Schema (Continued)

- How to create a dimension? Why use Select Distinct in create dim?
- What are attributes in dimensions?
- How many different ways in creating a dimension?
- When creating a fact table, why don't use the dimension tables?

- Why use the tables in the operational database when creating a fact table?

- Why use two separate dim? Why not combine into one dim? Is it for drilling down?
- What is hierarchy in dimension? Why use hierarchy?
- What is normalization in dimension?
- Why not have a combined dimension? What is the difference between hierarchy and non-hierarchy in dimension modelling?

- Is tempfact compulsory?
- Can we have multiple fact measures? Do we need to separate into multi facts?

Week 4: Bridge Tables

- What is a bridge table? Why is a bridge table needed?
- What are the three versions of Product-Supplier bridge tables?

- What is weight factor and ListAGG? Where are these two attributes located?
- How to create a dimension with weight factor and ListAGG?
- Is ListAGG needed? When is weight factor needed?
- How to calculate the fact measure after averaging with weight factor?

Week 5: Data Cleaning and Data Exploration

- What types of mistakes are there?
- How to explore data? How to find mistakes?
- How to correct mistakes? Should all mistakes be corrected?

- How to combine several facts into one (the Robcor case study)?
- Why do we need to consider co-pilot?

Week 6: Multi-Fact and Determinant Dimensions

- What is subject-oriented in data warehousing
- Can dimensions be shared?
- How to determine multi-fact?

- What is a determinant dimension?
- When to use a determinant dimension?
- How do we know whether a dimension is a determinant dimension or not?

Week 7: Class Test

PART II: OLAP

Week 8: Temporal Data Warehousing and OLAP

- What is temporal data warehousing?
- Why is a temporal data warehousing needed?

- How to implement a temporal star schema?
- How to calculate the fact measure correctly?

- CUBE, ROLL UP
- Grouping, Decode
- Partial Cube, Partial Roll Up
- Rank, Dense Rank, Percent Rank, Row Number
- Cumulative and Moving Aggregate
- Partition

Week 9: Query Processing

- What is index? Primary and secondary index
- What is the relationship between PK and index?
- What is the relationship between FK and index?
- How to create index in SQL
- How many different ways to process the WHERE clause in SQL? Which one is the most efficient technique
- How many different ways to process JOIN, and which one is the most efficient method?
- Which one is considered more expensive: number of readings or number of comparisons?
- What is the number of readings of Nested Loop Join, Sort-merge and Hash?

- What is query tree?
- How to draw a query tree given an SQL?

- Execution method for Order By, Distinct, Group By, Group By and Order By, Distinct and Order By, and Join

Week 10: Hints and OLAP Optimization

- What is hint?
- PK and index
- Incomplete hints, conflicting hints, same effect hints, etc
- PK and FK

- What is Cartesian product? When is Cartesian product used?
- What are the execution methods for Cartesian product?

- What are the execution methods for Group By, CUBE, ROLL UP, and partial?
- What are the execution methods for cumulative and moving aggregates?
- What are the execution methods for ranking? And partition?

- How to optimize Order By and Group By?
- How to optimize cumulative and moving aggregates?

PART III: Advanced Data Warehousing

Week 11: Data Warehousing Architecture

- What is data warehouse architecture?
- Why do we need multiple level of granularities of star schema?
- What is level of granularity?
- What is drill-down?
- Why is data warehouse architecture needed?
- How can we lower down the granularity of a star schema?
- Can we have multiple star schemas on the same level?
- Can we have more than three levels?
- Can we determine what level a star schema is?
- Why sometimes adding a dimension does not increase the level of granularity?

Week 12: Junk Dimensions and Multi-Input Operational Databases

- What is a junk dimension?
- Compare junk and non-junk method

- What is a surrogate key?
- Why is surrogate key used? When?
- How to match the surrogate key of a dimension and the link to the fact?

- Why do we use multi-input?
- How to create a star schema, with multiple input operational databases?