Course - Data Warehouse from Basic to Master

(can use Keyword planner to find the name)

Prerequisite – Basic SQL and preliminary knowledge of databases.

Good to have – Experience/knowledge on BigQuery, Snowflake, Azure Synapse and other enterprise level data warehouses/databases.

Author – [Abhishek Sharma](mailto:abhisharma8800@gmail.com)[surbhiwalecha@gmail.com](mailto:surbhiwalecha@gmail.com)

Duration – 6-8 hours

Table of Content

1. Warm-up
   1. What is a database and data warehouse? -Abhishek
   2. Why a data warehouse? -Abhishek
   3. OLTP and OLAP -Abhishek
   4. Benefits/Advantages/Applications of data warehouse -Abhishek
   5. References for SQL and Database Courses/ Study Material -Surbhi
2. Basics
   1. CRUD operations - Surbhi
   2. Keys - Surbhi
      1. Primary Key/ Composite Primary Key,
      2. Unique Key,
      3. Foreign Key,
      4. Candidate Key,
      5. Super Key,
      6. Surrogate Keys
   3. Constraints - Surbhi
      1. NOT NULL,
      2. DEFAULT,
      3. AUTO-INCREMENT,
      4. PRIMARY KEY,
      5. UNIQUE KEY,
      6. FOREIGN KEY,
      7. CHECK
   4. Joins - Surbhi
      1. Inner
      2. Left Outer
      3. Right Outer
      4. Full Outer
      5. Cross Join
   5. Indexes (Clustered Index and Non-Clustered Index) - ABhishek
   6. ER Model - Abhishek
   7. Normalization and its form - - Not sure if this needs to be included - Surbhi
   8. Denormalization - Abhishek
   9. Fact and Dimension Tables - Surbhi
   10. Case Study covering Concepts of Basic Section - to be searched / Case study to be in such a manner that can be used in resume as well ; Book My Show ka transactional DB - database design
3. Intermediate
4. Type of Data Warehouse
5. Schemas - Star, Snowflake, Constellation
6. Sharding
7. Scaling
8. ETL / ELT
9. CAP theorem
10. ACID properties
    1. Transaction States in Database - <https://youtu.be/ObwYFVLB_VI>
    2. Schedule
11. Advance
12. Shuffling
13. Order of Execution in SQL Queries
14. How to write queries in Data Warehouse - Best Practices
15. Optimization
    1. Partitioning
    2. Bucketing
16. Master
17. Case Study
18. Data warehouse designing
    * Types of data warehouse
    * Key Components of DW
    * Data Warehouse Architecture
    * Datawarehousing vs DataMining
    * Fact and Dimension tables
    * MetaData
    * Data Cube

### Differentiate between Agglomerative hierarchical clustering and Divisive clustering.

### What do you mean by dimensional modeling in the context of data warehousing?

* + Data Purging

### Explain the architecture of a data warehouse.

* + Top Down vs Bottom Up Pros and Cons
* Intermediate
  + Keys
  + Schemas (Star/ Snowflake/ Constellation)
  + Nested Fields
  + CAP theorem
  + ACID properties
  + Schemas
* Advanced Section
* Design
  + Case Study
* Horizontal Scaling vs Vertical Scaling
* Performance Tuning
* OLAP vs OLTP
* Distributed Data Warehouse
* Streaming data in data warehouse
* Data Warehouse in HIVE
  + HIVE metastore
* IOT related Data Warehouse
* Partitioning and Bucketing
* Datamart, Data mesh, Data farm, DataLake
* New things coming up in Data Warehouse market – like Snowflake, BigQuery (which tool has which advantage)
* DE file types and its compression techniques
* Data Shuffling – minimize shuffle (Performance Optimization)
* ETL related concepts



<https://www.sap.com/insights/what-is-a-data-warehouse.html>