DB Design, ETL

1. REVIEW:
2. Relational Database Design
3. Normalization
4. NoSQL[bigdata] → topic: each [assignment]
5. OLTP vs OLAP, Data warehouse, Data mart, Star schema and snowflake schema.
6. ETL
7. Cronjobs, SQL script execution. P/L SQL.
8. Project 1.
9. Class Quiz, Summary, Project review

REVIEW

1. Difference between store procedure and function.
   a. https://stackoverflow.com/questions/1179758/function-vs-stored-procedure-in-sql-server
2. Difference between view and table?
   a. Have you heard of materialized view.
   b. https://stackoverflow.com/questions/93539/what-is-the-difference-between-views-and-materialized-views-in-oracle#:~:text=A%20view%20uses%20a%20query,with%20indexes%20applied%20to%20it.
3. What is outer join? Left outer join? Full outer join?
4. What is cursor? https://www.mysqltutorial.org/mysql-cursor/

Regular Expression

Matching patterns:

. all character

*, +, ? repeat factor * 0-n, + 1-n, ? 0-1

a* (, a, aaaa), b+ (b, bb, bbbb)

\ecape, \\ (\), \n, \t,

\s \d \d{3}

Wild card for * is %: “dav%” (dave, dav, david…)

Wild: dav%

Database design 1 2 3.

Class material:

See lecture 4, 5, 6 Pdf from stanford DB class.

Key concepts: ER diagram [entity relation]

BCNF, 3rd normal form:

Questions to answer: why normalize? Benefits, drawbacks?

-- RDBMS: Normalize? → Update/insert/delete consistency.

-- source of truth, consistency.

-- Query [join many tables, costly for reporting]

Transaction: RDBMS, transaction, OLTP | Reporting, readonly: OLAP

ERP, CRM, SC, HR, Payroll

Fortune 500:

HR, Payroll, CRM, ERP, …..

ERP: Enterprise resource Planning. {SAP, Oracle Applications}

Customer relation management {Salesforce, cloud service, SAP Oracle, …..)

Inventory management

Production planning

Finance Control,

Sale, Distribution, material management, ….

Startup: Payroll, Quickbook, ……, where the data come. ETL, Website. Logs.

OLTP vs OLAP

OLAP vs OLTP

OLTP vs OLAP: What's the Difference?

ERP (Enterprise Resource Planning) [SAP, Oracle], [HR, FI/CO, SD, LOGistics, VM, MM, PP, ….] CRM (Customer Relation Management) [Sales], Peoplesoft, Workday, ….

SAP, Oracle … [Sales, Marketing, Finance, HR, MM, WM, …] Daily operation.

CRM [.... SalesForce, Oracle, SAP, SugarCRM, Intuit, MSFT] daily operation.

Star schema, snowflake schema, galaxy schema: OLAP

ETL ⇒ Extraction, Transformation, Loading ⇒ data warehouse ==? Query:

https://www.guru99.com/star-snowflake-data-warehousing.html

ETL Process

1. MSFT SSIS (ETL), SSAS (OLAP) ⇒ Demo. [TBD]
2. Informatica, AWS glue,
3. Kinesis, airflow,
4. Usage: to build Data warehouse for reporting
5. https://www.guru99.com/top-20-etl-database-warehousing-tools.html
6. OLTP → systems 100 order extraction (join) ---> OLAP
7. Facttable ---> dimension tables → start/snowflake/galaxy schema
    a. Customerid ←--- customer [
    b. Productid ←---
    c. Time
    d. City
    e. Order
    f. Orderqty
    g. Revenue
    h. Tax

Project: ETL

Classimodels:

Fact-table ← transactional data tables [order, orderdetail, Revenue,]

Dimension: customer ← sync, employees, sales, production, dates,

Master data: customer/material/product/date/employee

ETL [

1. Initial load:
2. Delta load: load the changes for daily load. Where [date]
3. Schedule the delta load daily.
4. Fact table and dimension table up to date.

Cron jobs introduction

Crontab -l, -e,

Job scheduling on linux, mac.

https://help.ubuntu.com/community/CronHowto

Windows: at → windows scheduler. ⇒ some dev.

Production deployment Linux (Ubuntu,mint, Susi, Redhat, oracle, Debian, centos)

Linux [--]

Cygwin/ gitbash → linux command.

Window: *.bat cmd. Powershell *.sh.

Shell script:

Home work

Review DB design theory and understand normalization, ER diagram

Assume that we will let you design a transactional sales order system, can you design the database tables for classicmodels? Get familiar with key, foreign key relations, and understand why.

Is it normalized? Why we have products and product lines table?

How to look at the hierarchical data in employee table?

If we want to run report on those data, can you design the star schema?

What is the benefit of running reports against the OLAP vs joining OLTP tables.

Can you write daily extract ETL job

Project 1: DATA ETL, Reporting == Due 8/25.

1. Familiar with the data in classic model.
2. Design a data warehouse solution. OLAP fact table, dimensions (customer, employee, product) in star schema
3. Write extraction, transformation and loading scripts to load your data. Include daily extraction. ⇒ Initial + delta load → fact/dimension.
4. Create cronjobs for ETL process
5. Write reports based on OLAP tables:
6. H2 database. ⇒ Mysql database, then ETL.
7. ---- sql session: exam ---> write sql →
8. Transaction: reports are not allow on complex query
9. https://www.mysqltutorial.org/mysql-stored-procedure-tutorial.aspx/

SQL

Home work: 1. Write sql for the classicmodel.

2. Leetcode database sql query.

3. Project: ETL [build a star schema (fact/x dimension tables), write etl (sql scripts+cron, store procedure + write some reports]

4. Database design.