Focal Points Data Warehouse Technologies

Fundamentals

- Terms: Data Warehousing, Data Warehouse, DWH Requirements (FASMI/Codd)
- OLAP vs. OLTP
- Distinction DWH vs. Big Data vs. Parallel Systems vs. Federated Systems

Architecture

- Sketch of DWH Architecture
- Definition/Characteristics of all Components of DWH Architecture

Multidimensional Data Model

- Terms: Cube, Operations on Cube, Summation Types, Types of Aggregate Functions,
- Types of Hierarchies
- Star vs. Snowflake Schema
- Slowly Changing Dimensions

Simple Hierarchies – Present a straightforward classification with each level aggregating directly into the next. For instance, "Product Category → Product Family → Product Group → Item" shows successive levels leading up to a top node.
Parallel Hierarchies – Within a single dimension, multiple, independent paths coexist without hierarchical dependence, such as a time dimension that includes separate paths for "Year → Quarter → Month → Day" and "Week," allowing flexible aggregation levels - can use roll up or drill down to navigate in between the levels

ETL

- Terms: Bulk Loading, ETL, ELT
- Monitoring and Extraction Techniques for Different Source Types
- Differential-Snapshot Problem
- Duplicate Detection
- Transformation on Example
- Identify Data Quality Issues (via SQL)

OLAP Queries

- Terms: CUBE, ROLLUP, GROUPING SETS, OVER-clause, Special Aggregate Functions
- Cube on Example
- Iceberg Cube

```
SELECT A, B, C, COUNT(*), SUM(X) FROM R GROUP BY CUBE(A, B, C) HAVING COUNT(*) >= N;
```

Storage Structures

- Terms: Partitioning, Materialization Strategies
- ROLAP vs. MOLAP vs. HOLAP
- Data Dwarf on Example
- Column vs. Row Stores

Index Structures

- Characterization of Index Structures
- B-Tree Specialties
- Grid File
- Bitmap Indexes
- UB-Tree
- R-Tree

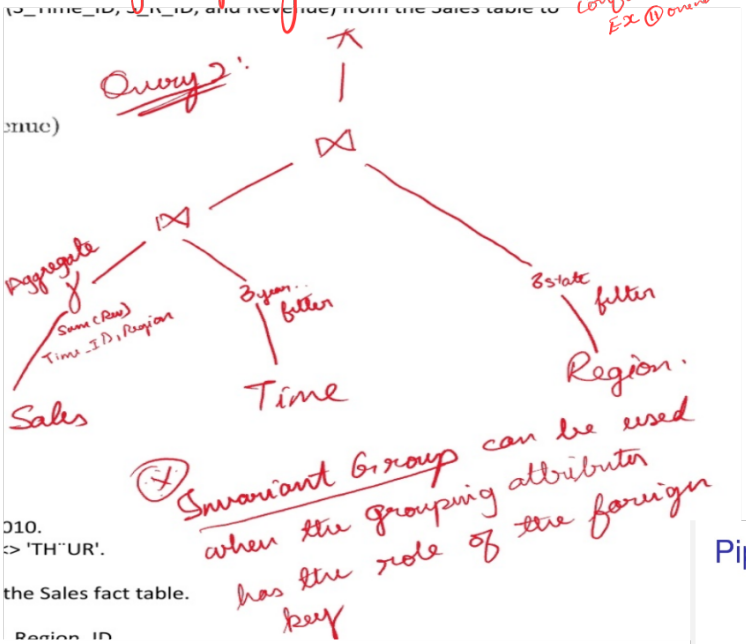
Query Optimization

- Star Join
- Invariant Grouping/Early Pre-Grouping
- Data Cube Lattice & Pipesort

Business Intelligence

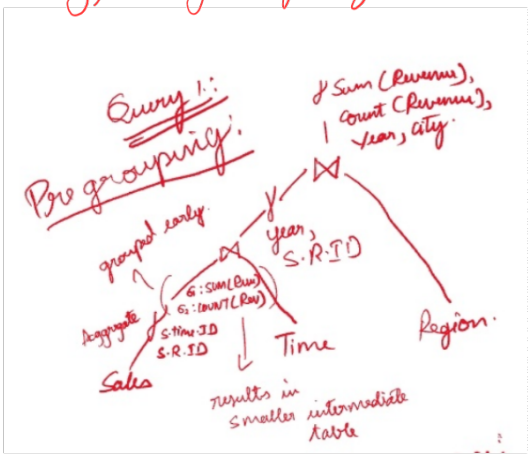
- Terms: Business Intelligence, Business Intelligence Process, Report Types
- MCQ: K-Means Clustering, Association Rule Mining, Decision Trees

Invariant grouping:

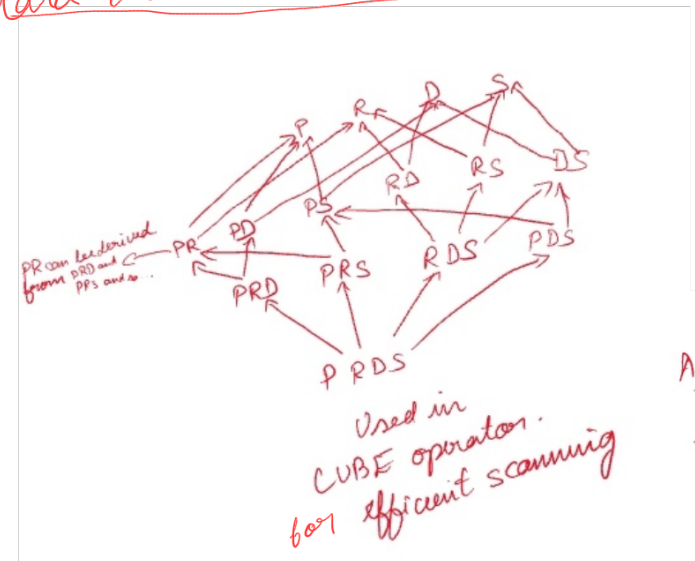


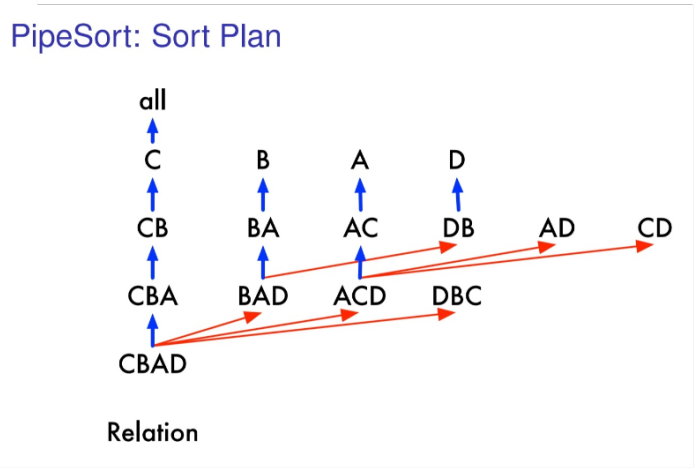
Invariant Group can be used when the grouping attributes has the role of the foreign key

Early/Pre grouping



data cube Lattice:





Assignment 4:

→ Goal: find a subgraphs with minimal sum of edge costs

- edge costs: →Sorting: Sort data based on order.
  ⇒Pipelining: Additional grouping on subsets of attributes

Example: