Data Mining:

Concepts and Techniques

(3rd ed.)

— Final Review (Chapter 2,4 and 5)—

Slides Courtesy of Textbook

Chapter 2: Getting to Know Your Data

- Data attribute types: nominal, binary, ordinal, interval-scaled, ratio-scaled
- Many types of data sets, e.g., numerical, text, graph, Web, image.
- Gain insight into the data by:
  - Basic statistical data description: central tendency, dispersion, graphical displays
  - Data visualization: map data onto graphical primitives
  - Measure data similarity and dissimilarity

Chapter 4: Data Warehousing and On-line Analytical Processing

- Data Warehouse: Basic Concepts
  - (a) What Is a Data Warehouse?
  - (b) Data Warehouse: A Multi-Tiered Architecture
- Data Warehouse Modeling: Data Cube and OLAP
  - (a) Cube (A Lattice of Cuboids): Calculate the Number of Cuboids in a Cube
  - (b) Conceptual Modeling of Data Warehouses
  - (c) Stars, Snowflakes, and Fact Constellations: Schemas for Multidimensional Databases
  - (d) Dimensions: The Role of Concept Hierarchy
  - (e) Measures: Their Categorization and Computation: distributive, algebraic and holistic
  - (f) Cube Definitions in Database systems
  - (g) Typical OLAP Operations: drill up, drill down, slice and so on…
- Data Warehouse Usage
- Data Warehouse Implementation
  - (a) Efficient Data Cube Computation: Cube Operation, Materialization of Data Cubes, and Iceberg Cubes
  - (b) Efficient Processing of OLAP Queries

Chapter 5: Data Cube Technology

- Data Cube Computation: Preliminary Concepts
  - Cube Materialization: Full Cube, Iceberg Cube, Closed Cube, Closed Cells, Base Cells vs. Aggregated Cells, Ancestor Cells vs. Descendant Cells
- Data Cube Computation Methods
  - Multiway Array Aggregation
  - BUC

Multiway Array Aggregation

- Using multi-dimensional chunk to materialize cube
- High Efficiency for Full Cube Computation for data set with small number of dimensions:
  - Simultaneous aggregation on multiple dimensions
  - Intermediate aggregate values are re-used for computing ancestor cuboids
- No Iceberg Optimization: Cannot Apriori Prunning
- How to calculate the minimum memory requirement?

BUC

- Introduction:
  • BUC (Beyer & Ramakrishnan, SIGMOD'99) • Bottom-Up cube computation
  - (Note: **Top-Down** in our view!)

- Partition and Optimization: Divides dimensions into partitions and facilitates iceberg pruning
  - If a partition does not satisfy *min_sup*, its descendants can be pruned
  - No simultaneous aggregation: Difference to Multiway Array Aggregation
  - How to do iceberg prunning on a small cube?