Reading Questions, Lecture 9 (10/6)

Papers: Data Mining Overview, Data Cubes

Data warehouses are an important class of database that is widely used to store and query information about business processes (e.g., sales, customers, etc.) Data warehouse workloads are "OLAP"-style aggregate analysis queries.

Readings for class consist of a thorough overview as well as one of the most influential papers that presents a clever way to simultaneously compute many aggregates for warehouse-style OLAP workloads.

Questions to consider:

1. What types of schemas are typically used in data warehouses? Why? What advantages do those schemas have?
2. For what types of data is a bitmap index an efficient index? When are bitmap indices a bad idea?
3. How does the data cube operator improve upon the GROUP BY expression? What limitations of group by does it address?
4. What mechanisms does the data cube use to improve the performance of grouped aggregates over multiple separate GROUP BY expressions?

Samuel Madden