CSE 544
Principles of Database Management Systems

Alvin Cheung
Fall 2015
Lecture 8 - Data Warehousing and Column Stores
Announcements

• Shumo office hours change
  – See website for details

• HW2 due next Thurs
  – Please start soon!

• Project meetings this week
  – Sign up on doodle if you haven’t already
Where We Are

• Data models
  – Relational
  – IMS / Codasyl
  – Unstructured

• Query processing
  – Algorithms for relational operators
  – Indexing and physical design

• Queries that real-world users write
  – Data warehousing
  – Transaction processing
Where We Are

• What queries do real people write?
  – Data warehousing
    • Column stores (today)
    • Parallel databases (Thursday)
    • Programming models (next week)
  – Transaction processing
References

• Data Cube: A Relational Aggregation Operator Generalizing Group By, Cross-Tab, and Sub-Totals. 
  Jim Gray et. al. Data Mining and Knowledge Discovery 1, 29-53. 1997

• Database management systems. 
  Ramakrishnan and Gehrke. 
  Third Ed. Chapter 25
Why Data Warehouses?

- **DBMSs designed to manage operational data**
  - Goal: support every day activities
  - Online transaction processing (OLTP)
  - Ex: Tracking sales and inventory of each Wal-mart store

- **Enterprises also need to analyze and explore their data**
  - Goal: summarize and discover trends to support decision making
  - Online analytical processing (OLAP)

- To support OLAP consolidate all data into a warehouse
The Origin of Data Warehouses

Walmart, 90s

Sale transactions

Sales DB

Nightly Backups

Cashiers
The Origin of Data Warehouses

Amazon, 00s

Operational DB

Sale transactions

Nightly Backups

Sales DB

Users
Data Warehouse Overview

• Consolidated data from many sources
  – Must create a single unified schema
  – The warehouse is like a materialized view

• Very large size: terabytes of data are common

• Complex read-only queries (no updates)

• Fast response time is (not as) important
  – Compared to transaction processing
Creating a Data Warehouse

- **Extract** data from distributed operational databases
- **Clean** to minimize errors and fill in missing information
- **Transform** to reconcile semantic mismatches
  - Performed by defining views over the data sources
- **Load** to materialize the above defined views
  - Build indexes and additional materialized views
- **Refresh** to propagate updates to warehouse periodically
- This is known as the **ETL pipeline**
Alternative: Distributed DBMS

• User submits a query at one site

• Query is defined over data located at different sites
  – Different physical locations
  – Different types of DBMSs

• Query optimizer finds the best distributed query plan
  – Query executes across all the locations
  – Results shipped to query site and returned to user

• (Stay tuned for the next lecture!)
Back to Warehouses: Outline

- Multidimensional data model and operations
- Data cube & rollup operators
- Data warehouse implementation issues
- Other extensions for data analysis
Data Analysis Cycle

- Formulate query that extracts data from the database
  - Typically ad-hoc complex query with group by and aggregate

- Visualize the data (e.g., spreadsheet)
  - Dataset is an N-dimensional space

- Analyze the data
  - Identify “interesting” subspace by aggregating other dimensions
  - Categorize the data and compare categories with each other
  - Roll-up and drill-down on the data
Multidimensional Data Model

- Focus of the analysis is a collection of measures
  - Example: Wal-mart sales

- Each measure depends on a set of dimensions
  - Example: product (pid), location (lid), and time of the sale (timeid)

Slicing: equality selection on one or more dimensions

Dicing: range selection
Star Schema

Representing multidimensional data as relations (ROLAP)

- Product
  - pid
  - pname
  - category
  - price

- Location
  - locid
  - city
  - state
  - country

- Sales
  - pid
  - timeid
  - locid
  - sales

- Times
  - timeid
  - date
  - week
  - month
  - quarter
  - year

Facts table: Sales
In BCNF

Dimensions tables
- Product
- Location
- Times
Not necessarily normalized
Dimension Hierarchies

Dimension values can form a hierarchy described by attributes

- Product
  - category
    - pname
- Time
  - year
  - quarter
  - week
  - month
  - date
- Location
  - country
  - state
  - city
Desired Operations

• Histograms (agg. over computed categories)
  – Problem: awkward to express in SQL (paper p.34)

• Summarize at different levels: roll-up and drill-down
  – Ex: total sales by day, week, quarter, and year

• Pivoting
  – Ex: pivot on location and time
  – Result of pivot is a cross-tabulation
  – Column values become labels

<table>
<thead>
<tr>
<th>Year</th>
<th>WI</th>
<th>CA</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>500</td>
<td>200</td>
<td>700</td>
</tr>
<tr>
<td>2006</td>
<td>150</td>
<td>850</td>
<td>1000</td>
</tr>
<tr>
<td>2007</td>
<td>250</td>
<td>400</td>
<td>650</td>
</tr>
<tr>
<td>Total</td>
<td>900</td>
<td>1450</td>
<td>2350</td>
</tr>
</tbody>
</table>
Challenge 1: Representation

• Problem: How to represent multi-level aggregation?
  – Ex: Table 3 in the paper need $2^N$ columns for N dimensions!
  – Ex: Table 4 has even more columns!
Challenge 1: Representation

The representation suggested by Table 5 and unioned GROUP BYs “solve” the problem of representing aggregate data in a relational data model. The problem remains that expressing roll-up, and cross-tab queries with conventional SQL is daunting. A six dimension cross-tab requires a 64-way union of 64 different GROUP BY operators to build the underlying representation.

There is another very important reason why it is inadequate to use GROUP BYs. The resulting representation of aggregation is too complex to analyze for optimization. On most SQL systems this will result in 64 scans of the data, 64 sorts or hashes, and a long wait.

3. CUBE and ROLLUP operators

The generalization of group by, roll-up and cross-tab ideas seems obvious: Figure 3 shows the concept for aggregation up to 3-dimensions. The traditional GROUP BY generates the N-dimensional data cube core. The N-1 lower-dimensional aggregates appear as points, lines, planes, cubes, or hyper-cubes hanging off the data cube core. The data cube operator builds a table containing all these aggregate values. The total aggregate using function f() is represented as the tuple: ALL, ALL, ALL, ((, . . .), ALL, f(*))

Points in higher dimensional planes or cubes have fewer ALL values.
Creating a data cube requires generating the power set (set of all subsets) of the aggregation columns. Since the \texttt{CUBE} is an aggregation operation, it makes sense to externalize it by overloading the SQL \texttt{GROUP BY} operator. In fact, the cube is a relational operator, with \texttt{GROUP BY} and \texttt{ROLL UP} as degenerate forms of the operator. This can be conveniently specified by overloading the SQL \texttt{GROUP BY}.

Figure 4 has an example of the cube syntax. To give another, here follows a statement to aggregate the set of temperature observations:

\begin{verbatim}
SELECT day, nation, MAX(Temp) FROM Weather GROUP BY CUBE Day((Time)) AS day, Country((Latitude, Longitude)) AS nation;
\end{verbatim}

The semantics of the \texttt{CUBE} operator are that it first aggregates over all the \selectlist{} attributes in the \texttt{GROUP BY} clause as in a standard \texttt{GROUP BY}. Then, it \texttt{UNIONs} in each super-aggregate of the global cube—substituting \texttt{ALL} for the aggregation columns.

If there are $N$ attributes in the \selectlist{}, there will be $2^N - 1$ super-aggregate values. If the cardinality of the $N$ attributes are $C_1 \hookrightarrow C_2 \hookrightarrow \ldots \hookrightarrow C_N$, then the cardinality of the...
Challenge 1: Representation

- Problem: How to represent multi-level aggregation?
  - Ex: Table 3 in the paper need $2^N$ columns for N dimensions!
  - Ex: Table 4 has even more columns!
  - And that’s without considering any hierarchy on the dimensions!

- Solution: special “all” value

<table>
<thead>
<tr>
<th>T.year</th>
<th>L.state</th>
<th>SUM(S.sales)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>WI</td>
<td>500</td>
</tr>
<tr>
<td>2005</td>
<td>CA</td>
<td>200</td>
</tr>
<tr>
<td>2005</td>
<td>ALL</td>
<td>700</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>ALL</td>
<td>ALL</td>
<td>2350</td>
</tr>
</tbody>
</table>

Note: SQL-1999 standard uses NULL values instead of ALL
Challenge 2: Computing Aggregations

- Need $2^N$ different SQL queries to compute all aggregates
  - Expressing roll-up of a single column and cross-table queries is thus daunting
  - Cannot optimize all these independent queries

- Solution: CUBE and ROLLUP operators
Outline

- Multidimensional data model and operations
- Data cube & rollup operators
- Data warehouse implementation issues
- Other extensions for data analysis
Data Cube

- CUBE is the N-dimensional generalization of aggregate

- Cube in SQL-1999
  
  ```sql
  SELECT T.year, L.state, SUM(S.sales)
  FROM Sales S, Times T, Locations L
  WHERE S.timeid=T.timeid and S.locid=L.locid
  GROUP BY CUBE (T.year,L.state)
  ```

- Creating a data cube requires generating the power set of the aggregation columns
Rollup

- Rollup produces a subset of a cube

- Rollup in SQL-1999
  
  ```sql
  SELECT T.year, T.quarter, SUM(S.sales)
  FROM Sales S, Times T
  WHERE S.timeid=T.timeid
  GROUP BY ROLLUP (T.year, T.quarter)
  ```

- Will aggregate over each pair of (year,quarter), each year, and total, but will **not** aggregate over each quarter
Computing Cubes and Rollups

• Naive algorithm
  – For each new tuple, update each of $2^N$ matching cells

• More efficient algorithm
  – Use intermediate aggregates to compute others
  – Relatively easy for distributive and algebraic functions

• Updating a cube in response to updates is more challenging
Outline

• Multidimensional data model and operations

• Data cube & rollup operators

• Data warehouse implementation issues

• Other extensions for data analysis
Indexes

- **Bitmap indexes**: good for sparse attributes (few values)
  - M | F | custid | name | gender | rating
  - 0 | 1 | 10    | Alice | F      | 3
  - 1 | 0 | 11    | Bob   | M      | 4
  - 1 | 0 | 12    | Chuck | M      | 1

- **Join indexes**: to speed-up specific join queries
  - Example: Join fact table F with dimension tables D1 and D2
  - Index contain triples of rids <r₁,r₂,r> from D₁, D₂, and F that join
  - Alternatively, two indexes, each one with pairs <v₁,r> or <v₂,r>
    where v₁, v₂ are values of tuples from D₁, D₂ that join with r
Materialized Views

• How to choose views to materialize?
  – Physical database tuning

• How to keep view up-to-date?
  – Could recompute entire view for each update: expensive
  – Better approach: incremental view maintenance
  – Example: recompute only affected partition

  – How often to synchronize? Periodic updates (at night) are typical
    • Think back in the case of Walmart
Outline

- Multidimensional data model and operations
- Data cube & rollup operators
- Data warehouse implementation issues
- Other extensions for data analysis
Additional Extensions for Decision Support

• Window queries

```
SELECT L.state, T.month, AVG(S.sales) over W AS movavg
FROM Sales S, Times T, Locations L
WHERE S.timeid = T.timeid AND S.locid=L.locid
WINDOW W AS (PARTITION BY L.State
              ORDER BY T.month
              RANGE BETWEEN INTERVAL ‘1’ MONTH PRECEDING
              AND INTERVAL ‘1’ MONTH FOLLOWING)
```

• Top-k queries: optimize queries to return top k results

• Online aggregation: produce results incrementally
Leveraging Column Stores
References

Main Idea

• Most DBMS (up till now) store each tuple using row-major order

![Row diagram]

• Store tuples by column-major order instead?

![Column diagram]

Why?
From Row-Store to Column-Store

Rows stored contiguously on disk (+ tuple headers)

Columns stored contiguously on disk (no headers needed)
Recall Record Formats in Row Stores

Variable length records

Field 1   Field 2   ...   ...   Field K

Record header
More Detailed Example

Row-based (4 pages)

<table>
<thead>
<tr>
<th>Page</th>
<th>Column-based (4 pages)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A 1</td>
<td>A 1</td>
</tr>
<tr>
<td>A 2</td>
<td>A 2</td>
</tr>
<tr>
<td>A 2</td>
<td>A 2</td>
</tr>
<tr>
<td>A 2</td>
<td>A 2</td>
</tr>
<tr>
<td>B 2</td>
<td>B 2</td>
</tr>
<tr>
<td>B 4</td>
<td>B 2</td>
</tr>
<tr>
<td>C 4</td>
<td>C 4</td>
</tr>
<tr>
<td>C 4</td>
<td>C 4</td>
</tr>
</tbody>
</table>

C-Store also avoids large tuple headers
Column-Store Optimizations

Numbers from earlier paper and C-Store system: “Column-Stores vs. Row-Stores: How Different Are they Really?” Abadi et. al. SIGMOD’08.

• Vectorized processing / Block iteration (1.5X)
  – Pass blocks of values between ops instead of individual tuples
• Compression: e.g., run-length encoding of columns (10X)
• Late tuple materialization (3X improvement)
  – Process individual columns as long as possible
  – Merge columns into complete tuples as late as possible
• Invisible joins (1.5X)
## Compression Example

<table>
<thead>
<tr>
<th>Page</th>
<th>Row-based (4 pages)</th>
<th>Column-based (4 pages)</th>
<th>Compressed (2 pages)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A 1 A 2 A 2 A 2</td>
<td>A A A 1 2 2</td>
<td>4XA 1X1</td>
</tr>
<tr>
<td></td>
<td>A 2 A 2 A 2</td>
<td>A A A 2 2</td>
<td>2XB 4X2</td>
</tr>
<tr>
<td></td>
<td>B 2 B 2 B 4</td>
<td>B B C 2 4 4</td>
<td>2XC 5X4</td>
</tr>
<tr>
<td></td>
<td>C 4 C 4 C 4</td>
<td>C C C 4 4</td>
<td></td>
</tr>
</tbody>
</table>

CSE 544 - Fall 2015
Late Tuple Materialization

Ex: SELECT R.b from R where R.a=X and R.d=Y

Early materialization

Late materialization

CSE 544 - Fall 2015
Late Tuple Materialization

Figure 4.1: An example of a select-project-join query with late materialization.
Figure 4.2: An example multi-column block containing values for the SHIPDATE, RETFLAG, and LINENUM columns. The block spans positions 47 to 53; within this range, positions 48, 49, 52, and 53 are active (i.e., they have passed all selection predicates).
Joins

```sql
SELECT emp.age, dept.name
FROM emp, dept
WHERE emp.dept_id = dept.id
```
Simulating a Column-Store DBMS in a Row-Store DBMS

• Vertical partitioning
  – Two-column tables: (key, attribute)

• Index-only plans
  – Create a B+ tree index on each attribute
  – Answer queries using indexes only, without reading actual data

• Materialized views
  – Each view contains a subset of columns
Conclusion

• Column-store DBMS outperforms row-store DBMS
  – Measured on a data warehousing benchmark (SSBM)

• Late materialization and compression are key factors

• Difficult to simulate a column-store in a row-store
  – Tuple overheads cause data blow-up
  – Column joins are expensive
  – Hard to get the DBMS to “do the right thing” (e.g., index plans)

• Not the end of the story, however, … see CIDR’09 paper