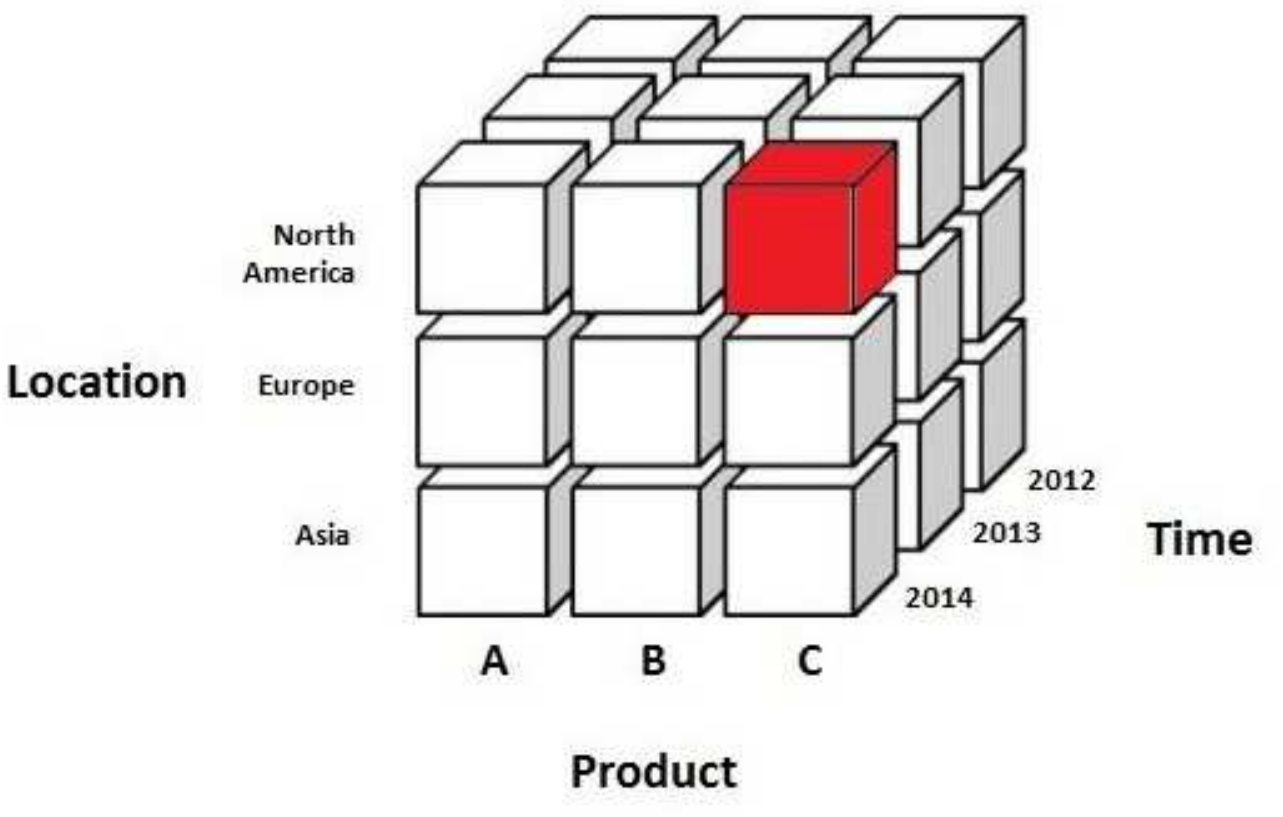
CS2102 Lecture 11b
DBMS Landscape

DBMS Landscape

- OLTP (On-Line Transaction Processing)
  - Characteristics:
    - Large data (Gigabytes to Terabytes)
    - ACID transactions
    - Workload: small update transactions
- Data Warehouse
  - a.k.a. OLAP (On-Line Analytical Processing)
  - Characteristics:
    - Very large data (Terabytes to Petabytes)
    - Workload: compex read-mostly queries (e.g., business intelligence, data mining)
- Everything else

OLAP: Multidimensional Data Model

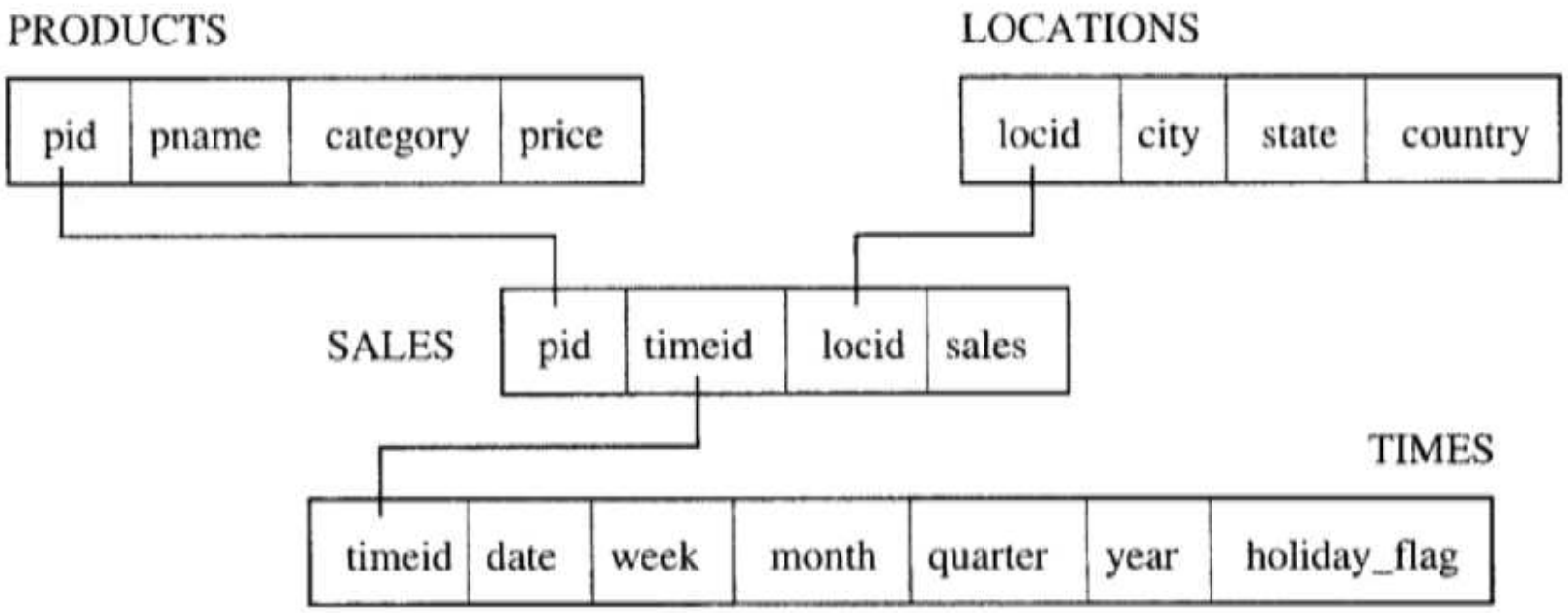
- Stores a collection of numeric measures
- Each measure depends on a set of dimensions



http://www.openit.com/faster-analysis-with-olap/

OLAP: Star Schema

- Data is modeled using a fact table & dimension tables



(Ramakrishnan & Gehrke, 2003)

OLAP: Multidimensional Aggregation

Find the total sales

```
SELECT SUM(sales) FROM Sales
```

Find the total sales for each state

```
SELECT L.state, SUM(S.sales)
FROM Sales S JOIN Locations L ON S.locid = L.locid
GROUP BY L.state
```

Find the total sales for each city and year

```
SELECT L.city, T.year, SUM(S.sales)
FROM Sales S JOIN Locations L ON S.locid = L.locid
JOIN Times T ON S.timeid = T.timeid
GROUP BY L.city, T.year
```

Find the total sales for each city, year, category

```
SELECT L.city, T.year, P.category, SUM(S.sales)
FROM Sales S JOIN Locations L ON S.locid = L.locid
JOIN Times T ON S.timeid = T.timeid JOIN Products P ON S.pid = P.pid
GROUP BY L.city, T.year, P.category
```

OLAP: Multidimensional Aggregation

Find the total sales for each city, year, category
Find the total sales for each city, year
Find the total sales for each city, category
Find the total sales for each year, category
Find the total sales for each city
Find the total sales for each year
Find the total sales for each category
Find the total sales

```
SELECT      L.city, T.year, P.category, SUM(S.sales)
FROM        Sales S JOIN Locations L ON S.locid = L.locid
            JOIN Times T ON S.timeid = T.timeid
            JOIN Products P ON S.pid = P.pid
GROUP BY    CUBE (L.city, T.year, P.category)
```

`https://www.postgresql.org/docs/current/static/queries-table-expressions.html`

OLAP: Analytic Window Functions

For each state and month, compute its moving average sales over three months

```
SELECT   L.state, T.month, AVG(S.sales) OVER W AS movingAvg
FROM     Sales S, Times T, Location 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)
```

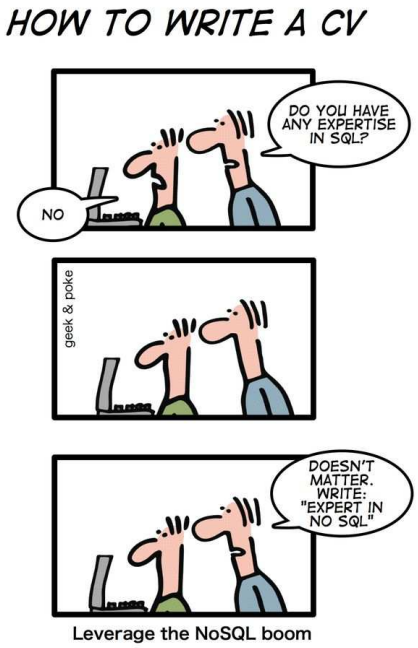
`https://www.postgresql.org/docs/current/static/tutorial-window.html`

Modern DBMSs: NoSQL & NewSQL

http://www.informationweek.com/big-data/big-data-analytics/16-nosql-newsql-databases-to-watch/d/d-id/1269559

NoSQL Systems

- Early NoSQL Systems
  - Google's Bigtable
  - Amazon's Dynamo
  - Yahoo!'s PNUTS
- Data Models:
  - Key-value
  - Document
  - Graph
  - etc.



- Features of many early NoSQL systems
  - Schema-less data
  - Simple access API (put & get) instead of query language
  - Limited/No ACID transactional support
  - Weak consistency for replicated data

NewSQL Systems

- Targeted at OLTP workloads
- Features
  - Relational data model
  - SQL query language
  - ACID transactions
  - Runs on distributed cluster of shared-nothing nodes
- Some examples:
  - Clustrix
  - CockroachDB
  - Google's Spanner
  - MemSQL
  - Microsoft's Azure Cosmos DB
  - VoltDB