Transaction Processing

The most common use of relational databases is for operational data.

- Examples:
  - Students enrolling in courses
  - Customers purchasing products
  - Passengers purchasing airline tickets

On-Line Transactional Processing (OLTP)
Databases that support the basic operations of a business are generally classified as OLTP systems.

- Workload characteristics:
  - Simple queries
  - Many short transactions making small changes
  - Systems tuned to maximize throughput of concurrent transactions
Beyond Transaction Processing

More recent uses of operational data:

**Decision Support** Summarizing data to support high-level decision making
- Complex queries with much aggregation

**Data Mining** Searching for trends or patterns in data for a business to exploit
- Simple queries, but very data-intensive

**Data Warehousing**
A data warehouse is a separate copy of the operational data used for executing decision support queries and/or data mining queries.

On-Line Analytical Processing

**On-Line Analytical Processing (OLAP)**
OLAP is a particular type of decision support
- Data is modeled as multidimensional array
- Queries are usually ad hoc
- Queries select and aggregate cells of the array

- OLAP systems are divided into two categories:
  - Special-purpose OLAP systems
    - store data as multidimensional arrays (MOLAP)
    - provide an OLAP-specific query language
  - Relational databases
    - store data in relations (ROLAP)
    - queries written in SQL

Multidimensional Data

- Example: Number of Sales

```
+-----+-----+-----+
|    | A   | B   | C   |
|-----+-----+-----+-----|
|     | 14  | 45  | 22  |
|     | 12  | 20  | 27  |
|     | 11  | 27  | 30  |
|-----+-----+-----+-----|
|     | 1   | 2   | 3   |
+-----+-----+-----+
```

Notes
OLAP Queries

- OLAP queries typically aggregate over one or more dimensions. Examples:
  - Total sales
  - Total sales this year for each product category
  - Total sales for each store per quarter

- OLAP is a tool for ad hoc data exploration/visualization
  - Ad hoc queries tend to be iterative
  - Desirable to express queries using operations over previous result

OLAP Query Operations

- Slicing and Dicing (i.e., equality selection and range selection)

- Roll-up and Drill-down
  (i.e., aggregate at different levels of a dimension hierarchy)
Data Cube

- A data cube extends a multidimensional array of data to include all possible aggregated totals

![Data Cube Diagram]

CUBE operator in SQL:1999

- Generating the data cube:
  - \(\text{SUM(sales)}\) GROUP BY location, product, time (raw cells)
  - \(\text{SUM(sales)}\) GROUP BY location, time
  - \(\text{SUM(sales)}\) GROUP BY product, time
  - \(\text{SUM(sales)}\) GROUP BY product, location
  - \(\text{SUM(sales)}\) GROUP BY product
  - \(\text{SUM(sales)}\) GROUP BY location
  - \(\text{SUM(sales)}\) GROUP BY time
  - \(\text{SUM(sales)}\)

- CUBE operator in SQL:1999 groups by all combinations

  \[
  \text{SELECT lid, pid, tid, } \text{SUM(sales)} \\
  \text{FROM Sales} \\
  \text{GROUP BY CUBE(lid, pid, tid)}
  \]
Creating and Maintaining a Warehouse

Necessary steps when creating a warehouse:

- **Extract**: Run queries against the operational databases to retrieve necessary data
- **Clean**: Delete or repair tuples with missing or invalid information
- **Transform**: Reorganize the data to fit the conceptual schema of the warehouse
- **Load**: Populate the warehouse tables; build indexes and/or materialized views

**Note**
The data in the warehouse needs to be refreshed periodically (typically nightly or weekly). To make this process efficient, the above steps need to be executed incrementally.