

# Unit-2

Dr. Shalini Gambhir

# OLTP Systems

- Every large company has some sort of BI system to analyze operational data
- OLTP system vendors are constantly improving their ability to integrate with BI systems

# OLAP

- Modern BI systems designed to follow OnLine Analytic Processing (OLAP) model
- Named by IBM's E.F. Codd (inventor of SQL and relational databases)
- All OLAP systems have to meet three key criteria

# Three Key OLAP Criteria

1. Must support multidimensional analysis
  - Top managers/analysts have always thought multidimensionally
  - View “by” qualifiers are usually dimensions
  - OLAP systems organize data into multidimensional structures
  - Provide tools for users to examine/filter dimensional data

# Three Key OLAP Criteria

## 2. Fast retrieval times

- Answer more questions in less time
- “Infinite Question Syndrome”

## 3. Calculation engine that can handle specialized multidimensional math

# Three Key OLAP Criteria

## 2. Fast retrieval times

- Answer more questions in less time
- “Infinite Question Syndrome”

## 3. Calculation engine that can handle specialized multidimensional math

- Lets analysts use simple formulas that are auto-performed across dimensions

# Dimensions

- **Dimension**: categorically consistent view of data
- Two tests for dimensionality:
  1. Can data about members be compared?
    - Sales numbers of one product compared to sales numbers of another product
  2. Can data from members be aggregated into summaries?
    - Jan, Feb, Mar aggregate together as Q1

# Slicing & Dicing

- Dimensions let you “slice and dice” multidimensional data





# Slicing & Dicing



# Slicing & Dicing



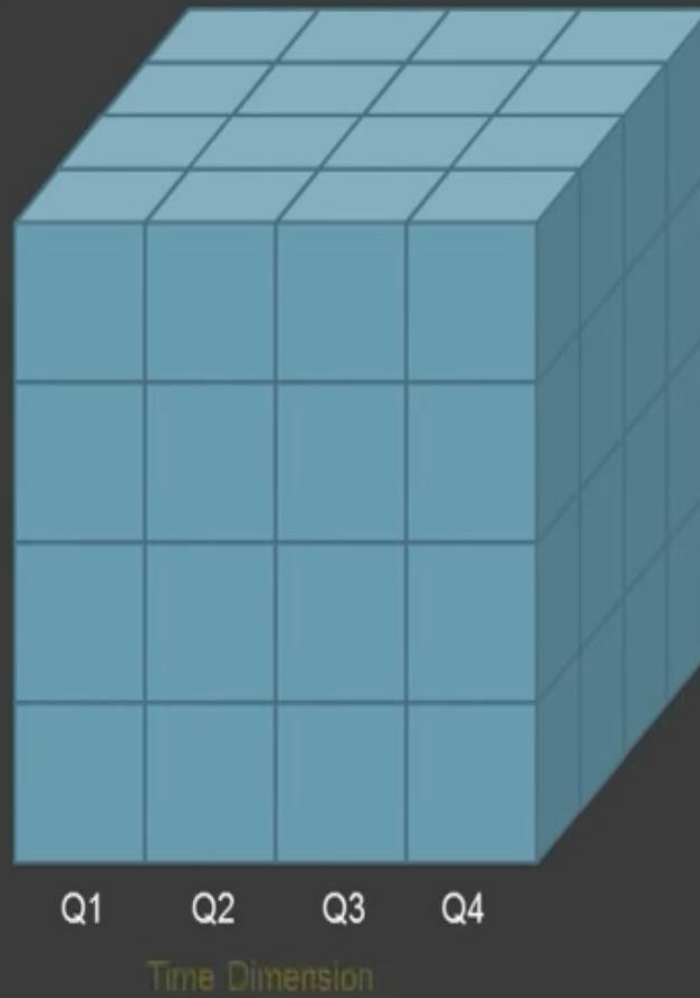
# Pivoted Soda Data

		Cola	Cherry	Grape	Lem-Lime	Total
Munich	Qtr 3	\$ -	\$ -	\$1,000	\$2,000	\$3,000
	Qtr 4	\$4,000	\$1,000	\$ -	\$ -	\$5,000
	<b>Total</b>	<b>\$4,000</b>	<b>\$1,000</b>	<b>\$1,000</b>	<b>\$2,000</b>	<b>\$8,000</b>
Frankfurt	Qtr 3	\$ -	\$ -	\$3,000	\$2,000	\$5,000
	Qtr 4	\$ -	\$3,000	\$ -	\$ -	\$3,000
	<b>Total</b>	<b>\$ -</b>	<b>\$3,000</b>	<b>\$3,000</b>	<b>\$2,000</b>	<b>\$8,000</b>
Cologne	Qtr 3	\$2,500	\$2,000	\$ -	\$ -	\$4,500
	Qtr 4	\$ -	\$ -	\$1,500	\$2,000	\$3,500
	<b>Total</b>	<b>\$2,500</b>	<b>\$2,000</b>	<b>\$1,500</b>	<b>\$2,000</b>	<b>\$8,000</b>
Berlin	Qtr 3	\$1,500	\$2,000	\$ -	\$ -	\$3,500
	Qtr 4	\$ -	\$ -	\$2,500	\$2,000	\$4,500
	<b>Total</b>	<b>\$1,500</b>	<b>\$2,000</b>	<b>\$2,500</b>	<b>\$2,000</b>	<b>\$8,000</b>
<b>Grand</b>	<b>Total</b>	<b>\$8,000</b>	<b>\$8,000</b>	<b>\$8,000</b>	<b>\$8,000</b>	<b>\$32,000</b>

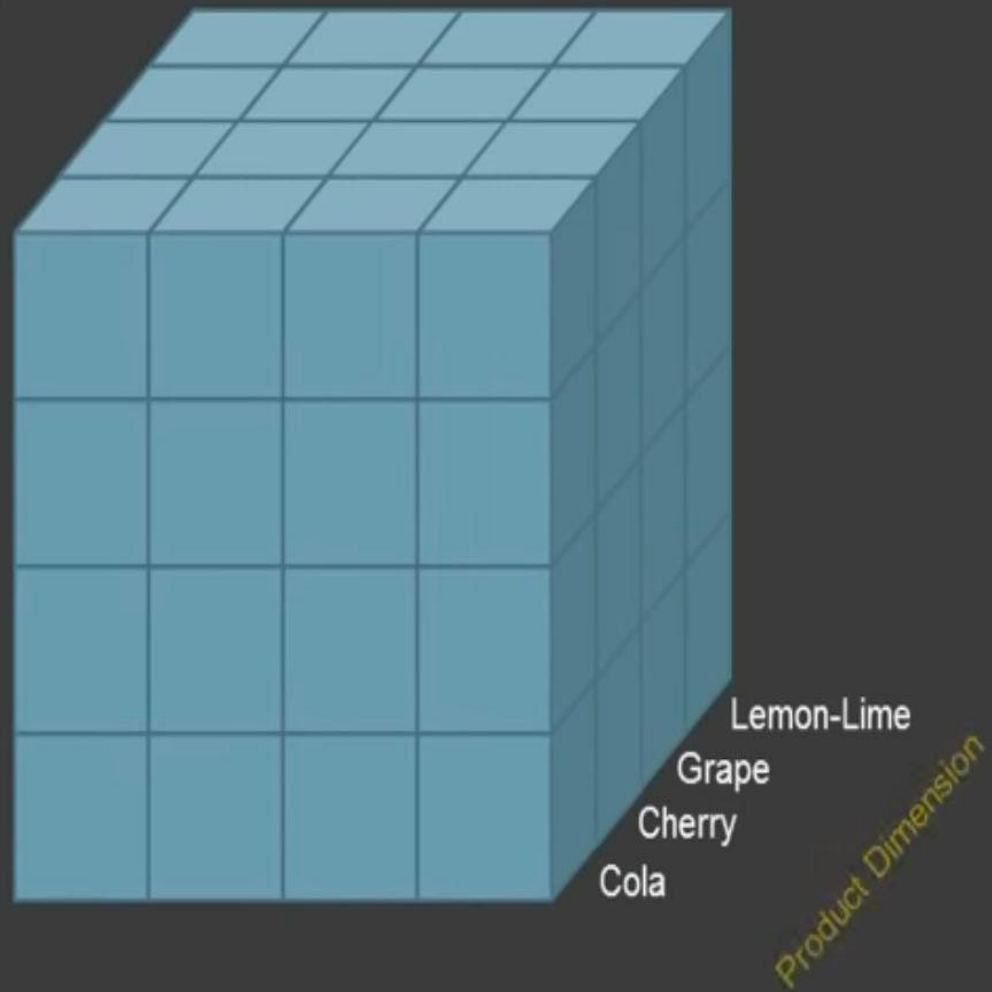
# OLAP



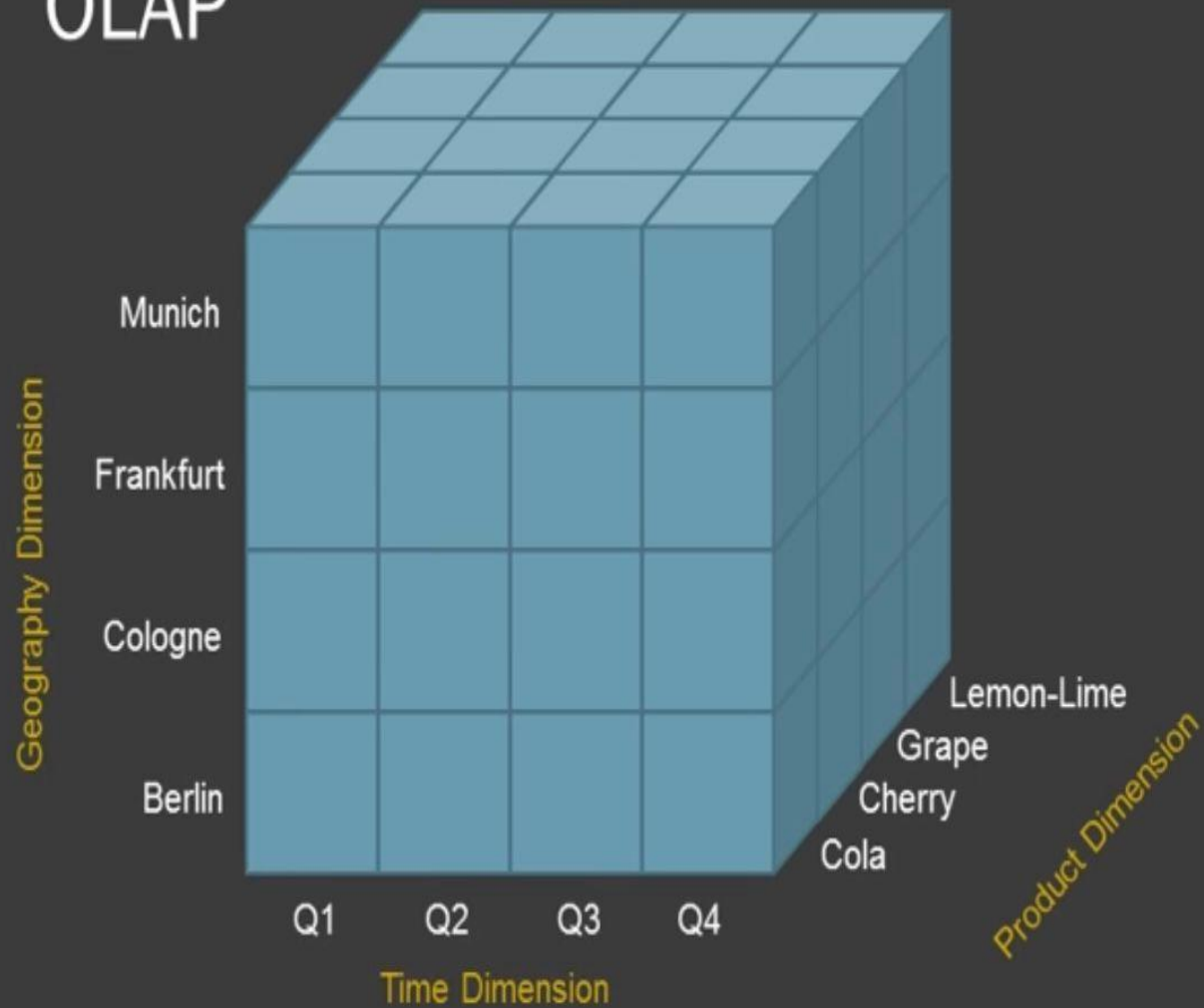
# OLAP



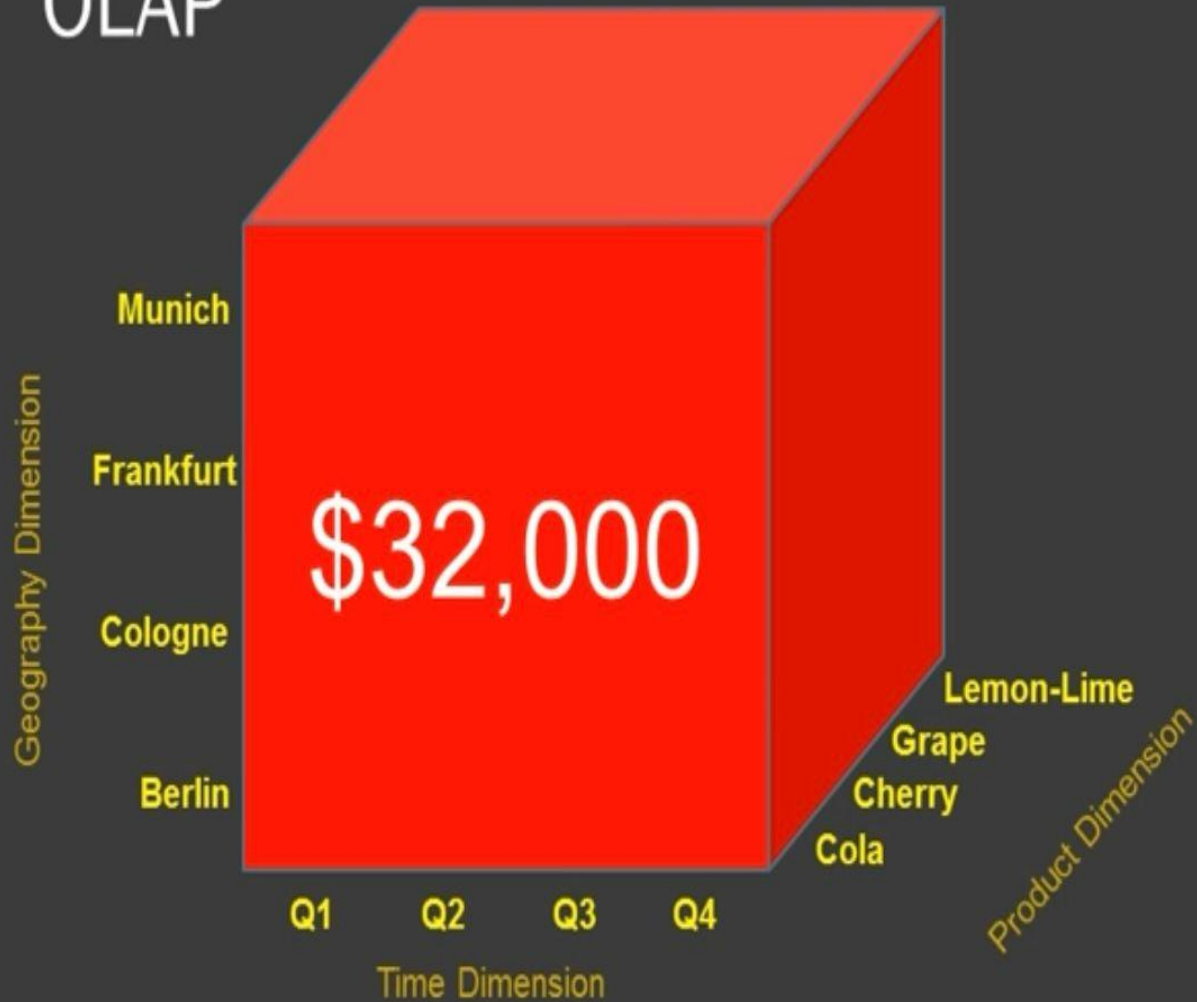
# OLAP



# OLAP

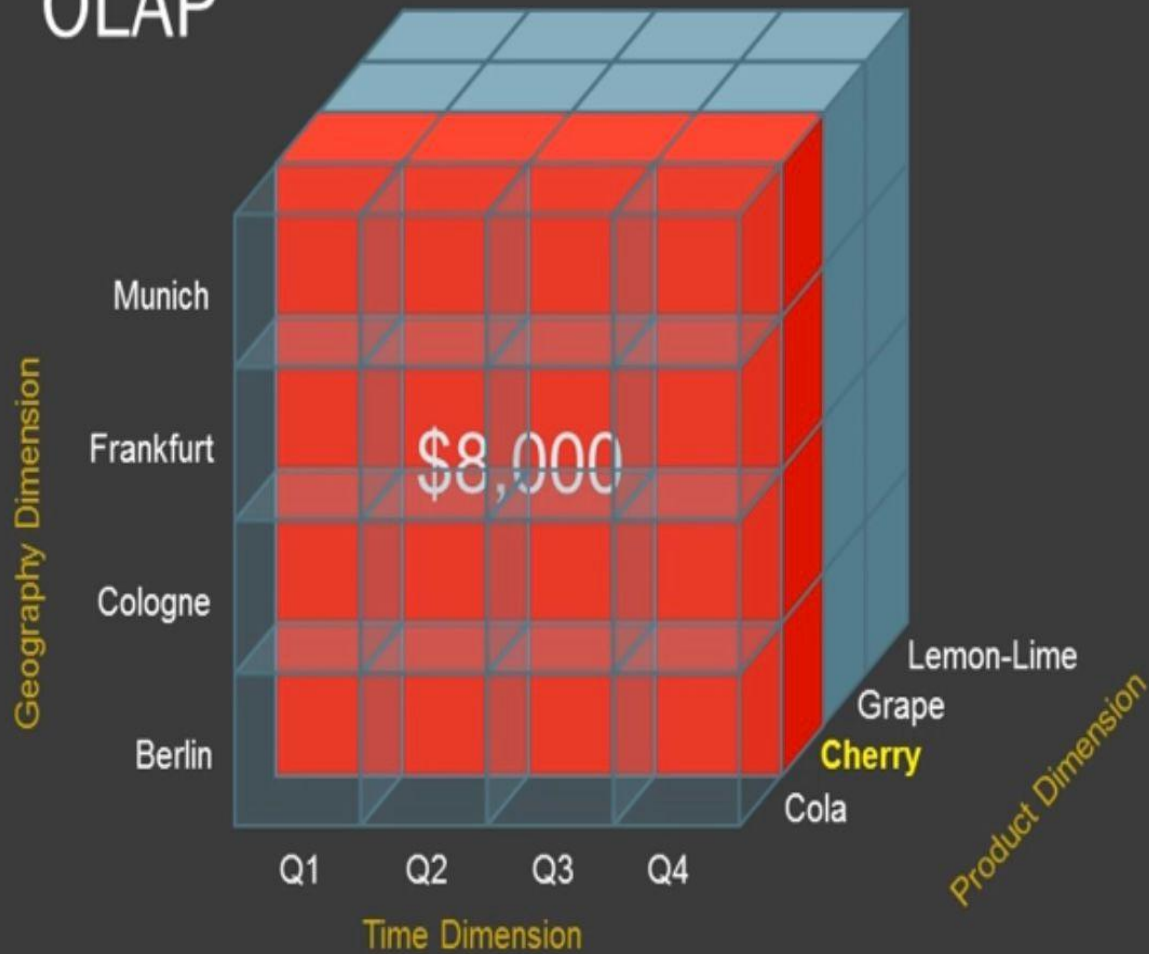


# OLAP





# OLAP



# OLAP

- Data cubes can have very large numbers of members
- **OLAP Cube**: multidimensional structure that stores and maintains discrete intersection values
- Some OLAP systems let cubes intersect with each other

# Hierarchies

- ⦿ Typical analysis task:
  - Units Sold, Average Price, Dollar Sales
  - 100 products
  - 24 months
  - 200 major cities
- ⦿ Total data points: 1,440,000
- ⦿ Not all products sold in all cities during all months

# Hierarchies

- **Hierarchy** – organizes data by levels
- Each level in the hierarchy is the aggregate of the levels beneath it
- Examples:
  - Monthly data rolls up to quarters and years
  - Cities roll up to regions and states
  - Products roll up to product lines and groups
- Calculations, like Average Price, can be back-calculated at each hierarchy level

# Hierarchies

- Hierarchies let you drill-down into data to explore interesting patterns and anomalies
- Top-down approach is like “20 Questions”
  - Start by exploring broad trends
  - Become more focused as analysis progresses
- Top-down thinking is natural way for humans to organize complex info

# Ad hoc Analysis

- Point-and-click drill-down is made usable by OLAP's rapid response model
- Lets managers and analysts perform ad hoc analysis
- Paper-based reporting gives fixed answers to fixed questions
- OLAP-based ad hoc analysis lets virtually any question be answered quickly

# Ad hoc Analysis

- Virtually any report can be formatted multidimensionally (pivoting & nesting dimensions)
- Virtually anyone can be taught how to do their own analysis work with minimal training

# Sample Hierarchy

2013

Q1

Q2

Q3

Q4

Jan

Feb

Mar

Apr

May

Jun

Jul

Aug

Sep

Oct

Nov

Dec



# Attributes

- **Attribute**: descriptive non-hierarchical information
- Examples:
  - Model number
  - Size
  - List price
  - Color
  - Flavor
  - Street address

# Measures

- **Measure**: any quantitative expression contained in an OLAP system
- A measure is the data that's being analyzed across multiple dimensions
- Example: Dollar Sales of soda by month, by product, and by city

# Measures

- Four important properties of a measure:
  1. Always a quantity or expression that yields a quantity
  2. Can take any quantitative format
  3. Can be derived from any original data source or calculation
  4. At least one measure required to perform OLAP analysis

# Measures

- The measures to be analyzed depend on the purpose of the OLAP system
- In BI, measures known by different names depending on application:
  - Metric/Key Performance Indicator (KPI)
  - Benchmark
  - Ratio

# Summary

- Analysis gap between raw data and BI can be bridged by combining OLTP systems with BI systems
- OLAP systems provide ad hoc analysis, slicing and dicing, pivoting dimensions, and drilling down through hierarchies