Data Mining

Associate Professor Dr. Raed Ibraheem Hamed

University of Human Development,
College of Science and Technology
Department of CS

2016 – 2017

Introduction

- What is OLAP
- Purpose of OLAP
- Why need OLAP over Relational Database
- OLAP Implementation
- Relational Database Model
- Two dimensions
- Specialised Multidimensional tool

What is OLAP

- Basic idea: converting data into information that decision makers need.

- Concept to analyze data by multiple dimension in a structure called **data cube**.

Purpose of OLAP

- To derive summarized information from large volume database
- To generate automated reports for human view
- Consistently fast response

Why need OLAP over Relational Database

- Provide analysis functions that are difficult or impossible to express in DBMS
- DBMS was developed primarily for transaction systems, not for reporting applications

OLAP Implementation

- Multidimensional OLAP (**MOLAP**)
- Relational OLAP (**ROLAP**)
- Hybrid OLAP (**HOLAP**)

MOLAP

- The database is stored in a special structure that is optimized for multidimensional analysis.
- Very fast query response time because data is mostly pre-calculated.
- Size is limited depending on the **time** taken to calculate the database and the **space** required to hold these pre-calculated values.

Relational Database Model

	Attribute 1 Name	Attribute 2 Age	Attribute 3 Gender	Attribute 4 Emp No.
Row 1	Anderson	31	F	1001
Row 2	Green	42	M	1007
Row 3	Lee	22	M	1010
Row 4	Ramos	32	F	1020

The table above illustrates the employee relation.

Two dimensions

City	Time	Total Revenue
Glasgow	Q1	29726
Glasgow	Q2	30443
Glasgow	Q3	30582
Glasgow	Q4	31390
London	Q1	43555
London	Q2	48244
London	Q3	56222
London	Q4	45632
Aberdeen	Q1	53210
Aberdeen	Q2	34567
Aberdeen	Q3	45677
Aberdeen	Q4	50056
.....	.....	.....
.....	.....	.....

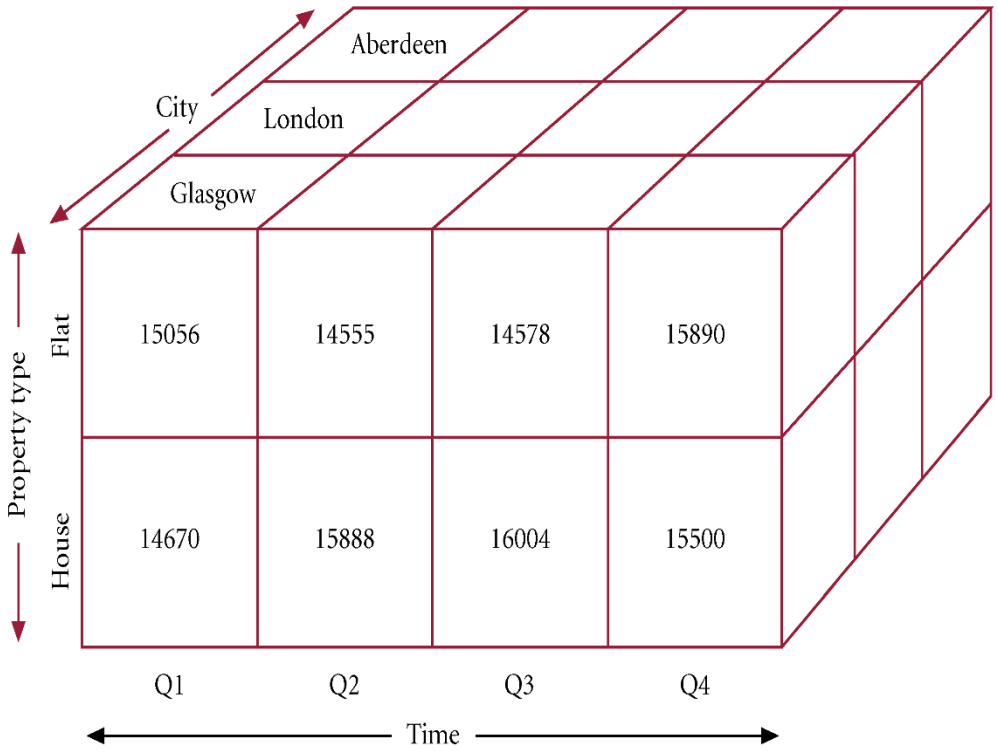
City

Time

City / Quarter	Glasgow	London	Aberdeen	.....
Q1	29726	43555	53210	.....
Q2	30443	48244	34567	.....
Q3	30582	56222	45677	.....
Q4	31390	45632	50056	.....

Three dimensions

Property Type	City	Time	Total Revenue
Flat	Glasgow	Q1	15056
House	Glasgow	Q1	14670
Flat	Glasgow	Q2	14555
House	Glasgow	Q2	15888
Flat	Glasgow	Q3	14578
House	Glasgow	Q3	16004
Flat	Glasgow	Q4	15890
House	Glasgow	Q4	15500
Flat	London	Q1	19678
House	London	Q1	23877
Flat	London	Q2	19567
House	London	Q2	28677
.....	.....	.....	.....
.....	.....	.....	.....



MOLAP tool Advantages and Disadvantages

- Advantages:
  1. Quick access to very large volumes of data
  2. Extensive and comprehensive libraries of complex functions
  3. Can access multidimensional and relational database structures
  4. Provide with calculated values
- Disadvantages:
  1. Difficulty of changing model
  2. Lack of support for very large volumes of data
  3. May require significant processing power

The MOLAP Cube

Fact table view:

sale	prodId	storeId	amt
	p1	s1	12
	p2	s1	11
	p1	s3	50
	p2	s2	8

Multi-dimensional cube:

	s1	s2	s3
p1	12		50
p2	11	8	

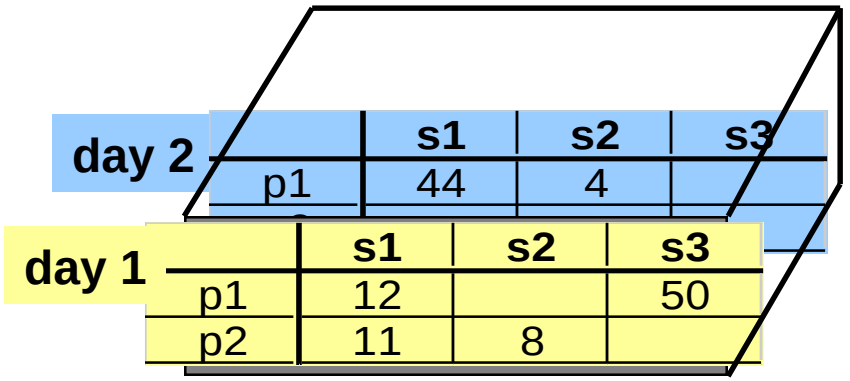
dimensions = 2

3-D Cube

Fact table view:

sale	prodId	storeId	date	amt
	p1	s1	1	12
	p2	s1	1	11
	p1	s3	1	50
	p2	s2	1	8
	p1	s1	2	44
	p1	s2	2	4

Multi-dimensional cube:



dimensions = 3


Thank You