

## COMP37332 tutorial 2 (Data warehouses and OLAP)

1. Explain why data warehouses are *time-variant*?
2. What are the main challenges in building a data warehouse? Explain and discuss them using an example. Explain the main architectural components in a data warehouse.
3. Describe the main 'clues' to identify measures and dimensions while designing a data-warehouse. Give an example using the star schema.
4. Explain the nine steps in modelling, constructing and managing a data warehouse.
5. Discuss the differences between OLTP and OLAP.
6. Describe the role of ranking functions (e.g. *rank()*, *dense\_rank()*) in SQL.
7. Explain slice and dice operations in OLAP. Given a fact table with sales data (for example *sales(market#, product#, time#, amount)* – see the lecture notes) and relevant dimension tables, write an SQL statement that slices the cube to select sales only in week 2, and dice it by regions.
8. The videotape company has stores in several regions. We would like to track profit information across different departments (Video Sales and Video Rentals) and regions (East, West, Central) in different years (e.g. 2000 and 2001). Create an appropriate data warehouse schema using the star multi-dimensional model and discuss the fact and dimension tables you would need.
9. Explain ROLLUP and CUBE extensions in SQL. Calculate the following query given the table (sales) below:

```
SELECT      Time, Region, Department, sum(Profit) AS Profit
FROM        sales
GROUP BY ROLLUP(Time, Region, Department)
```

Time	Region	Department	Profit
2000	Central	VideoRental	75,000
2000	Central	VideoSales	74,000
2000	East	VideoRental	89,000
2000	East	VideoSales	115,000
2000	West	VideoRental	87,000
2000	West	VideoSales	86,000
2001	Central	VideoRental	82,000
2001	Central	VideoSales	85,000
2001	East	VideoRental	101,000
2001	East	VideoSales	137,000
2001	West	VideoRental	96,000
2001	West	VideoSales	97,000

10. **Palo** is a leading open-source BI solution for corporate performance management and OLAP-based planning, analysis and reporting. Go to OLAP demo web-site <http://www.jedox.com/en/Demos/Online-Demos/OLAP-demos.html> and get familiarised with the options and queries that you can perform in both demos. For example, using the first demo (<https://www.worksheetserver.com/kd569271159/mis/run.html>), find out what were the top 10 products based on turnover in South Europe in March 2007. What kind of OLAP operations you can perform with Polo (e.g. what about regional or product analyses)?
11. **Dundas** reporting services is another example of DW and OLAP software. The software is briefly described in the movie available at <http://www.dundas.com/Products/Chart/NET/OLAP/index.aspx>. The full details on the software features are available here: <http://www.dundas.com/Products/Chart/NET/OLAP/QuickStart1.aspx>. Get familiarised with the demo at <http://demos3.dundas.com/OlapDemo62/>, and find out the amount of Customer Count and Internet Order Count in first half of 2003 FY (fiscal year) (hint: use the Grid tab and select the two measures from the Sales Summary). Then, using the Geography dimension, present the data on these two measures by country, and then pivot the data so that you can browse it based on countries (and their regions).

12. Given are the fact table *PropertySale*(*branchNo*, *propertyType*, *yearMonth*, *saleAmount*), dimension table *Branch*(*branchNo*, *city*) and the following SQL statement:

```
SELECT      propertyType, yearMonth, city, SUM(saleAmount)
FROM        Branch, PropertySale
WHERE       Branch.branchNo = propertySale.branchNo
           AND propertySale IN ('2007-01', '2007-02')
           AND branch.city IN ('Manchester', 'Edinburgh', 'Birmingham')
GROUP BY   CUBE (propertyType, yearMonth, city).
```

Explain (in plain English) which data would the above query retrieve. If the query resulted in the table below, what would be the result of the following query:

```
SELECT      propertyType, yearMonth, city, SUM(saleAmount)
FROM        Branch, PropertySale
WHERE       Branch.branchNo = propertySale.branchNo
           AND propertySale IN ('2007-01', '2007-02')
           AND branch.city IN ('Manchester', 'Edinburgh', 'Birmingham')
GROUP BY   ROLLUP (propertyType, yearMonth, city).
```

propertyType	yearMonth	city	saleAmount
flat	2007-01	Manchester	115432
flat	2007-01	Edinburgh	236573
flat	2007-01	Birmingham	7664
flat	2007-01		359669
flat	2007-02	Manchester	123780
flat	2007-02	Edinburgh	323100
flat	2007-02	Birmingham	8755
flat	2007-02		455635
flat		Manchester	239212
flat		Edinburgh	559673
flat		Birmingham	16419
flat			815304
house	2007-01	Manchester	77987
house	2007-01	Edinburgh	135670
house	2007-01	Birmingham	4765
house	2007-01		218422
house	2007-02	Manchester	76312
house	2007-02	Edinburgh	166503
house	2007-02	Birmingham	4889
house	2007-02		247713
house		Manchester	154308
house		Edinburgh	302173
house		Birmingham	9654
house			466135
	2007-01	Manchester	193419
	2007-01	Edinburgh	372243
	2007-01	Birmingham	12429
	2007-01		578091
	2007-02	Manchester	2001001
	2007-02	Edinburgh	489603
	2007-02	Birmingham	13644
	2007-02		703348
		Manchester	393520
		Edinburgh	861846
		Birmingham	26073
			1281439