Tutorial 4: Data Warehousing Design

Semester 1, 2021

Question 1

Suppose that a data warehouse for *AllElectronics* consists of the following three dimensions *time*, *item*, and *location*, and one measure *sales*. The dimension hierarchies used are [day < month < quarter < year] for *time*; [item name < brand < type] for *item*; and [street < city < province or state < country] for *location*.

(a) Draw a design of the data warehouse on sales data using the Star Schema.

(b) Construct a Snowflake Schema for the above data warehouse, with the dimension tables normalized to 3NF.

(c) Compare and contrast the above two schemas, Star and Snowflake.

Question 2

The following table is a sample of the *AllElectronics* sales data for one year. This report identifies a multi-dimensional model using dimension hierarchies defined in Question 1. Answer the following questions:

	location = "Chicago"				location = "New York"				location = "Toronto"				location = "Vancouver"			
	item				item				item				item			
	home				home				home				home			
time	ent.	comp.	phone	sec.	ent.	comp.	phone	sec.	ent.	comp.	phone	sec.	ent.	comp.	phone	sec.
Q1	854	882	89	623	1087	968	38	872	818	746	43	591	605	825	14	400
Q2	943	890	64	698	1130	1024	41	925	894	769	52	682	680	952	31	512
Q3	1032	924	59	789	1034	1048	45	1002	940	795	58	728	812	1023	30	501
Q4	1129	992	63	870	1142	1091	54	984	978	864	59	784	927	1038	38	580

(a) Perform a Roll-up operation on the *location* dimension from cities to countries.

(b) Perform a Drill-down operation on the *time* dimension from quarters to months (You may assume sales at each month).

(c) Perform a Slice operation for *time* = "Q1".

(d) Perform a Dice operation for (*location* = "Toronto" or "Vancouver") and (*time* = "Q1" or "Q2") and (*item* = "home entertainment" or "computer").

(e) Perform a Pivot operation on *location* and *item* dimensions.

Question 3

OLAP queries can be implemented using standard SQL queries. Consider the sample *AllElectronics* data in Question 2 and answer the following questions:

(a) Write SQL queries to implement the Roll-up operation in Question 2(a), the Dice operation in Question 2(d), and the Pivot operation in Question 2(e).

(b) Oracle supports CUBE extension to standard SQL for building a data cube. For the following SQL query, it equivalents to how many SQL queries with GROUP BY clauses? Why?

> **SELECT** item, time, location, sum(sales)
> **FROM** AllElectronics
> **GROUP BY CUBE**(item, time, location);

Question 4

What are the main differences between an operational database and a data warehouse? Given raw data from multiple sources, what are the main issues we need to consider when designing and constructing a data warehouse?

---ooo000O000ooo---