

Data Warehousing and Data Mining

FoodMart Exercise 1 (Non-Coursework)

The FoodMart data warehouse contains information about the supply and sale of food products in stores of a company. Among the tables in the warehouse are:

```
SALES_FACT_1998
SALES_FACT_DEC_1998
TIME_BY_DAY
```

SALES_FACT_1998 and SALES_FACT_DEC_1998 each contain the following columns:

```
SQL> describe sales_fact_1998
```

Name	Null?	Type
PRODUCT_ID		NUMBER(11)
TIME_ID		NUMBER(11)
CUSTOMER_ID		NUMBER(11)
PROMOTION_ID		NUMBER(11)
STORE_ID		NUMBER(11)
STORE_SALES		NUMBER(15,2)
STORE_COST		NUMBER(15,2)
UNIT_SALES		NUMBER(11)

Each row records information about the sale of one or more units of a product to a customer in a store. This includes the amount paid by the customer i.e. the store's sales receipt for the sale - STORE_SALES, the cost to the company - STORE_COST, and the number of units sold - UNIT_SALES. Two separate tables hold the sales information for 1998: SALES_FACT_1998 holds the rows representing sales in January to November 1998, while SALES_FACT_DEC_1998 holds the rows representing sales in December 1998.

The time of the sale is represented by an identifier TIME_ID. A separate table TIME_BY_DAY records information about the time represented by each TIME_ID value:

```
SQL> describe time_by_day
```

Name	Null?	Type
TIME_ID	NOT NULL	NUMBER(11)
THE_DATE		DATE
THE_DAY		VARCHAR2(15)
THE_MONTH		VARCHAR2(15)
THE_YEAR		NUMBER(5)
DAY_OF_MONTH		NUMBER(5)
WEEK_OF_YEAR		NUMBER(5)
MONTH_OF_YEAR		NUMBER(5)
QUARTER		VARCHAR2(2)
FISCAL_PERIOD		VARCHAR2(255)

1. `SALES_FACT_1998` and `SALES_FACT_DEC_1998` have not had primary keys created. Should they have primary keys? If so what should they be?
2. What time period is represented by a particular `TIME_ID` value in `TIME_BY_DAY`? What are the functional dependencies which hold between the columns of `TIME_BY_DAY`?
3. What would be the advantages and disadvantages of replacing the current design of three tables with alternative designs where:
 - `TIME_BY_DAY` is dropped and the information about the time of a sale currently in `TIME_BY_DAY` is held instead in `SALES_FACT_1998` and `SALES_FACT_DEC_1998`. Would there be any reason to retain the `TIME_ID` column?
 - `SALES_FACT_1998` and `SALES_FACT_DEC_1998` are replaced with a single table holding the sales information for all the months in 1998.
4. Formulate SQL queries to answer the following:
 - i. In which months in 1998 did store 24 have total sales receipts exceeding 5000.00? In each case display the month and the total sales receipts for that month.
 - ii. In which month in 1998 did store 12 have the highest total sales receipts? Display the month and the total sales receipts for that month.