**Database**

**Design Project**

DESCRIPTION

A national retailer company specialised in selling electronic products wants to increase their sales on Boxing Day. More discounts would be given to customer this year for the most sold products. Find out the top 9 electronics products best sold in quantity on Boxing Days over all the past five years. Provide and design an expandable star schema that can address the company’s concerns.

TASK   
• Design a relational schema that meets requirements.   
• Provide SQL queries which run against your designed model.   
• Provide explanations to the your queries.

# Design

A screenshot of a cell phone

Description automatically generated

# Explanation

The following high-level business goals the project must meet are:

* Increase overall company profit through increasing sales of popular electronic products on Boxing days over a five-year-period
* The Sales and Marketing department objectives are to:
  + Analyse customer product consumption trends and target specific market segments
  + Identify product trends and create strategies for developing promotion channels
  + Analyse sales channels and increase profits

In specific, this database design is used to record daily transactions, access reports to improve production, strategic forecasting and marketing forecasting JD Hub’s products. It will be used for:

* Entering sales order details
* Recording product inventory
* Entering and updating customer contacts
* Entering and updating promotion code for each order
* Recording of daily sales orders
* Listing recurring customer sales orders
* Listing recurring product favourites
* Listing high frequency stores
* Listing customer product tax amount and gross profit

## Technicalities

Facts: sales (ie. SALES\_FACT)

Measures: quantity, sales\_amount, gross\_profit, sales\_tax, timestamp, (unit\_cost)

Dimensions: date of sales (ie. DATE\_DIM), promotion used (ie. PROMOTION\_DIM), delivery method (ie. DELIVERY\_DIM), sales customer (ie. CUSTOMER\_DIM), product (ie. PRODUCT\_DIM), location of sales (ie. STORE\_DIM)

The transactional fact table SALES\_FACT record measurements for sales. It contains five foreign key columns, which together comprise a multi-column primary key. The foreign keys are related to the primary keys of the dimension tables below. A surrogate key is embedded into SALES\_FACT for data operations (ie. ETL) purposes.

|  |  |  |  |
| --- | --- | --- | --- |
| **Column** | **Datatype** | **Role** | **Description** |
| sales\_ID | NUMBER | Surrogate key | Key for extract, transform and load operations |
| store\_ID | NUMBER | Key | Related to STORE\_DIM |
| delivery\_ID | NUMBER | Key | Related to DELIVERY\_DIM |
| date\_ID | DATE | Key | Related to DATE\_DIM |
| promotion\_ID | NUMBER | Key | Related to PROMOTION\_DIM |
| product\_SKU | NUMBER | Key | Related to PRODUCT\_DIM |
| quantity | NUMBER | Fact | Number of units sold |
| unit\_cost | NUMBER | Fact | Cost per unit |
| sales\_amount | NUMBER | Fact | Total sales amount |
| gross\_profit | NUMBER | Fact | Gross profit after tax |
| sales\_tax | NUMBER | Fact | Tax per sales |
| timestamp | TIMESTAMP | Fact | Timestamp of sales |

**Dimension Table**: DATE\_DIM

The DATE\_DIM table defines a time dimension with six levels. The surrogate key holiday\_indicator is used to segregate a normal day with a holiday. In a standard hierarchy, the rollup sequence from the base to the top level is:

Date –> Week –> Month –> Quarter–> Half Year –> Year

Descriptions of DATE\_DIM are:

|  |  |  |
| --- | --- | --- |
| **Column** | **Datatype** | **Role** |
| date\_ID | DATE | Primary key  Surrogate key |
| date\_week | VARCHAR(10) | Attribute |
| calendar\_month | NUMBER | Attribute |
| calendar\_weekmonth | NUMBER | Attribute |
| calendar\_year | NUMBER | Attribute |
| calendar\_quarter | NUMBER | Attribute |
| calendar\_halfyear | NUMBER | Attribute |
| holiday\_indicator | BOOLEAN | Surrogate key |

**Dimension Table**: PROMOTION\_DIM

The PROMOTION\_DIM table defines a single rollup level: Promotion

Descriptions of PROMOTION\_DIM are:

|  |  |  |
| --- | --- | --- |
| **Column** | **Datatype** | **Role** |
| promotion\_ID | NUMBER | Primary key  Surrogate key |
| promotion\_name | VARCHAR(100) | Attribute |
| promotion\_media | VARCHAR(100) | Attribute |
| begin\_date | DATE | Attribute |
| end\_date | DATE | Attribute |
| promotion\_descr | VARCHAR(255) | Attribute |
| promotion\_code | VARCHAR(10) | Attribute |

**Dimension Table**: DELIVERY\_DIM

The DELIVERY\_DIM table defines a single rollup level: Delivery.

delivery\_type describes the method of delivery either chosen automatically or by the customer:

|  |  |  |
| --- | --- | --- |
| **Column** | **Datatype** | **Role** |
| delivery\_ID | NUMBER | Primary key  Surrogate key |
| delivery\_type | VARCHAR(10) | Attribute |
| delivery\_date | DATE | Attribute |
| received\_date | DATE | Attribute |
| delivery\_status | VARCHAR(10) | Attribute |

**Dimension Table**: CUSTOMER\_DIM

CUSTOMER\_DIM contains standard customer details such as name, customer type in the system (ie. customer\_type), customer address and payment method:

|  |  |  |
| --- | --- | --- |
| **Column** | **Datatype** | **Role** |
| customer\_ID | NUMBER | Primary key  Surrogate key |
| first\_name | VARCHAR(50) | Attribute |
| middle\_name | VARCHAR(50) | Attribute |
| last\_name | VARCHAR(50) | Attribute |
| customer\_type | VARCHAR(50) | Attribute |
| customer\_phone | VARCHAR(15) | Attribute |
| customer\_email | VARCHAR(255) | Attribute |
| customer\_address | VARCHAR(255) | Attribute |
| customer\_streetname | VARCHAR(255) | Attribute |
| customer\_suburb | VARCHAR(50) | Attribute |
| customer\_city | VARCHAR(50) | Attribute |
| customer\_postcode | NUMBER | Attribute |
| customer\_payment\_method | VARCHAR(15) | Attribute |
| other\_customer\_details | VARCHAR(255) | Attribute |

**Dimension Table**: PRODUCT\_DIM

PRODUCT\_DIM describes product\_SKU as its primary key while categorising each product based on its various descriptions:

|  |  |  |
| --- | --- | --- |
| **Column** | **Datatype** | **Role** |
| product\_SKU | NUMBER | Primary key  Surrogate key |
| product\_name | VARCHAR(255) | Attribute |
| product\_descr | VARCHAR(255) | Attribute |
| brand\_descr | VARCHAR(255) | Attribute |
| model\_descr | VARCHAR(255) | Attribute |
| category\_descr | VARCHAR(50) | Attribute |
| department\_descr | VARCHAR(50) | Attribute |
| unit\_cost | NUMBER | Attribute |
| import\_date | DATE | Attribute |

**Dimension Table**: STORE\_DIM

STORE\_DIM describes details such as the sales location, its manager, store address and maintenance details.

|  |  |  |
| --- | --- | --- |
| **Column** | **Datatype** | **Role** |
| store\_ID | NUMBER | Primary key  Surrogate key |
| store\_managerID | NUMBER | Attribute |
| store\_name | VARCHAR(255) | Attribute |
| store\_number | VARCHAR(255) | Attribute |
| store\_streetname | VARCHAR(255) | Attribute |
| store\_suburb | VARCHAR(50) | Attribute |
| store\_city | VARCHAR(50) | Attribute |
| store\_postcode | NUMBER | Attribute |
| store\_phone | VARCHAR(15) | Attribute |
| store\_email | VARCHAR(255) | Attribute |
| open\_date | DATE | Attribute |
| maintenance\_date | DATE | Attribute |
| maintenance\_descr | DATE | Attribute |

# SQL Queries

Suppose we have sufficient data for data analysis, the CEO’s request to find out the top 9 electronics products that are the best sold in quantity on Boxing Days for the past five years in NSW is:

**SELECT** product\_SKU, product\_name, **SUM** (quantity) **as** sales\_total

**FROM** SALES\_FACT sf **JOIN**

PRODUCT\_DIM pd

**ON** sf.prodct\_SKU = pd.product\_SKU

**WHERE** MONTH(sf.date\_ID) = 12 **AND**

DAY(sf.date\_ID) = 26 **AND**

YEAR(sf.date\_ID) > YEAR(CURRENT\_DATE) – 5

**GROUP BY** sf.product\_SKU, sf.product\_name

**ORDER BY** quantity

**FETCH** FIRST 9 **ROWS** ONLY;