

**Design rationale**

We have decided to design the data mart model as a star schema to emphasise performance and simplicity to cater to “Bits&Bobs,” as from the assignment brief, they have stated that they are not tech-savvy people. Star schema has fewer joins and is more denormalised than snowflake schema, thus providing a more efficient query performance while maintaining simplicity, which makes this design the most appropriate for “Bits&Bob” business intelligence future use.

**Fact table (Sales)**

We have developed this fact table (sales) to cover the areas of the sale transactions to derive meaningful insights for helping Jenny Landers identify her best-performing staff. The fact table consists of multiple foreign keys to establish a relationship between the dimensions table to provide relevant additional information.

The fact table foreign keys consist of the following:

* Date: The time dimension table “date” primary key is referenced by the “date” column in the fact table as a foreign key to associate the sales transaction to the specific date. This would enable analysts to derive insights into customer behaviour patterns, staff sales patterns, sale patterns based on location, and popular items sold based on dates.
* receiptId: The receipt dimension table “receiptId” primary key is referenced by the “receiptId” column in the fact table as a foreign key to associate the sales transaction to the specific receipt. This would ensure that guarantee that every transaction can be traced back to the relevant receipt, making it more convenient for data analysts.
* customerId: The customer dimension table “customerId” primary key is referenced by the “customerId” column in the fact table as a foreign key to associate the sales transaction to the specific customer. This would enable analysts to analyse their purchase patterns based on dates, what are their frequently purchased items, and the average total amount spent per receipt.
* officeId: The office dimension table “officeId” primary key is referenced by the “officeId” column in the fact table as a foreign key to associate the sales transaction to the office. This would enable analysts to derive insights into which office has the highest number of items sold and what are the popular items in their area based on the dates.
* staffId: The staff dimension table “staffId” primary key is referenced by the “staffId” column in the fact table as a foreign key to associate the sales transaction to the specific staff. This would enable analysts to analyse the staff performance based on sales and analyse their sales performance behaviour. The information can be further used to offer feasible solutions to the company to improve its staff performance.
* itemId: The item dimension table “itemId” primary key is referenced by the “itemId” column in the fact table as a foreign key to associate the sales transaction to the specific item. This would enable analysts to derive meaningful insights further to determine the most popular and least popular products based on the dates and locations.

The fact table metrics consist of the following:

* Transaction row: Transaction row metrics as a field would allow the analysts to see the transaction row in the receipt belonging to which customer and staff.
* Item price: Putting the price of the item would allow the analyst to see the staff. with the highest sales and potentially derive insights on product pricing based on the dates and whether it affects the sales. Furthermore, it stores the historical prices of the item, which can be used to derive insights into strategic plans for improving sales revenue.
* Quantity: The quantity of products metric would allow the analyst to see which staff have the most items sold under their staff ID based on the dates and potentially derive insights on the staff. working behaviour and customer purchase behaviour.
* Discounted row total: A discount metric would allow the analyst to see the customer behaviour on whether their current marketing tactic is working and which can be used to improve their current strategy. These metrics would automatically discount the row total of 12.5 percent if the customer is a loyalty member, and if not, it would just put the row total amount. This way, the analyst can see the total amount of discounted items sold.
* Row total of each transaction row: A row total metric would allow the analyst to see the total amount of items sold in a transaction row to enable them to derive customer behaviour patterns and staff sales patterns based on the dates.

**Time dimension table**

The time dimension contains a collection of information about the date. The information in the data fields is important as it can provide more comprehensive information to be used for trend analysis.

Data fields:

* Date {Primary key}: Unique identifier for each date
* Day: Day of the date
* Month: Month of the date
* Year: Year of the date

**Receipt dimension table**

receiptId: The receipt dimension table serves as a reference to associate the sales transaction with the receipt.

Data fields:

* receiptId {Primary key}: Unique identifier for each receipt

**Customer dimension table**

The customer dimension table contains a collection of information about customers in the company, which enables the examination of sales transactions relating to the specific customer. It would enable the data analysts to extract meaningful insights into customer behaviour and item preferences to help with marketing tactics.

Data fields:

* customerId {Primary key}: Unique identifier for each individual customer
* name: Name of the customer
* surname: Surname of the customer
* loyalty: Whether the customer is a loyalty member or not.

**Office dimension table**

The office dimension table contains a collection of information about the various locations within the company. It would enable data analysts to extract meaningful insights into which office has the highest sales and can investigate why certain office sales are lacking. From there, the company can perform an investigation into whether it is an internal or external problem.

Data fields:

* officeId {Primary key}: Unique identifier for each individual office
* location: Location of the office

**Staff dimension table**

The staff dimension table contains a collection of information about the staffs in the company. It acts as a reference link to the sales transaction and the specific staff. This would enable data analysts to analyse each individual performance, item popularity, and customer behaviour based on the dates.

Data fields:

* staffId {Primary key}: Unique identifier for each individual staff
* name: Name of the staff
* surname: Surname of the staff

**Item dimension table**

itemId: The item dimension table contains information about a variety of items offered by the company. It acts as a reference link to the sales transaction and specific item. This would enable data analysts to analyse the popularity of each product.

Data fields:

* ItemId {Primary key}: Unique identifier for each item
* Description: Item name

Before designing the data mart schema, we investigated the “Bits&Bobs” business requirements and rules and have developed several assumptions, limitations, and questions appropriate for this design.

Assumptions before designing the data mart:

* Clean data
* Data consistency
* Types of user requirements

Limitations of designing the data mart:

* Redundancy of data: since our data mart is designed in a star schema, which is more denormalised than the snowflake schema, it might lead to inefficient data storage. To address this problem, we will implement only foreign keys and metrics in the fact table. This would avoid data repetition stored in the fact table.
* Data complexity: Star schema does not naturally fit hierarchical data. It might be a problem to display these structures effectively. To address this problem, we have structured them into relevant dimension tables. This would maintain star schema simplicity while being able to handle hierarchical data.

Our data mart can be used to answer the questions below:

* Which staff has the lowest/highest item sold per month or year?
* Which staff has the highest total amount of items sold monthly or annually?
* Which customer is the highest spender?
* Which staff is the least productive based on sales?
* Is there any pattern work of staff based on the month?
* Which is the most popular item?
* Which is the least popular item?
* Is there any customer pattern based on time, products, and staff?
* Which month has the highest sales?
* Which month has the least sales?