1. Summary of the data what are they used for.

2. Describe relational schema, draw ER Diagram.

3. Describe start schema, draw ER Diagram.

Tables include:

calendar

customer locations

shippedOrders: calendar, customer, employee, product key, productline

employees

customers

productLine

4. App for analysis:

a. Select employee, productLine then print the sales of each month

b. Select productLine then show the sales of each quarter

c. Sales of each productLine per location

c. Draw diagrams? (Optional)

d. Pie chart compare sales of each product line each quarter? (Optional)

1. **Summary of the data**

For our data, we use the **classicmodels** database from MySQL sample databases. The original can be found using this link: [MySQL Sample Database (mysqltutorial.org)](https://www.mysqltutorial.org/mysql-sample-database.aspx). We did some modifications to the database in order to delete some irrelevant information. For example, the productlines table contained columns like text descriptions, html descriptions and image of each product line. We cannot do meaningful analysis on such descriptions, and both html descriptions and image columns only contain null values anyways. So, we decided to drop the entire productlines table, as well as any foreign keys associated with the table. We also deleted columns from other tables, such as credit limit from the customers table, phone extension from the employees table, comments from the order table. To briefly summarize the data, we have the “customerNumber” as the primary key, customers’ names, phone numbers, and addresses in the corresponding table. It has a foreign key “salesRepEmployeeNumber” linked to the employees. The employees have names, email addresses, office code, which links to the offices table, and job titles. Since job positions are hierarchical, there is a “reports to” foreign key that links to the primary key “employeeNumber” in the table itself. The orderdetails table contains the primary key “orderNumber”, foreign key “productCode”, as well as quantity and price per item. Similarity, the orders table contains dates and status, and the foreign key “customerNumber”. The payments table uses “customerNumber” as the primary key, and has columns checkNumber, payment date and the payment amount. Last but not least, the products table uses “productCode” as the primary key, and multiple feature columns. The major focus of that table should be productLine, quantityInStock, and buyPrice.

1. **Relational Schema**

**Before:**

**Diagram

Description automatically generated with medium confidence**

**After:**

**Diagram

Description automatically generated**

1. **Star Schema**

**Warehouse:**

**Table

Description automatically generated with medium confidence**

For our analytical data warehouse, the first dimension table we made was the calendar table. We generated a primary key that by incrementing automatically. We used the orders table from before, and we only selected the required dates of the shipped products. In addition, we broke up the date information into several columns like day of the week, quarter, and year. Remember that the employees table was self-referred, such that sales representatives report to a manager. For our sales representatives employee dimension table, we replaced the “reportsTo” foreign keys with the actual managers’ names, and their respective email addresses. The productLine table is self-explanatory, where we just made a primary key for each product line. For the products table, we filtered out useless information from the relational table, and we were left with productCode, productName, productDescription, and most importantly buyPrice. Notice that we only selected products that were actually ordered when constructing this table. For the customers table, we filtered out addresses, and we were left with city and country as our geographic data. In addition, we were only interested in customers that were helped by sales representatives. Lastly, for the sales fact table, we joined all the dimension tables described above. At the end, it consisted of all the primary keys and two target columns: quantityOrdered and PriceEach.