Part I SQL Task (10%)

Sales and inventory data analysis with SQL using Oracle SQL (100 marks)

Using the datasets provided, design a star schema data model and use SQL queries to answer key business questions related to sales and inventory management. You may need to enter dummy data in the above datasets.

TASK 1. In Oracle SQL, import data and create a Star Schema (40 marks)

This task involves both designing the star schema and implementing it **in Oracle SQL Developer**. The **partially completed star schema** is provided below. Your task is to complete the schema by **identifying the missing attributes in the fact and dimension tables** that are necessary to answer the five business questions.

A. **Design a** data warehouse schema to answer the following business questions. The data warehouse schema you create should be capable of answering the **five provided questions** through your design **(30 Marks)**.

- What is the **total sales amount** for *each* **product category**?
- Which **suppliers** contribute the most to **total revenue** and what percentage of **total sales** does each **supplier** contribute?
- How many **products** has each **employee** sold and what is the **average value of sales** per **employee**?
- Which **states** have the *highest number* of **unique customers** and what is the **average sales value** per **customer** in those **states**?
- What are the top 3 **product categories** that generate the highest number of **repeat purchases** and which **customer segments** are responsible for the most **repeat purchases**?

A partially completed Star Schema is provided below. Your task is to complete the schema by identifying the missing fact and dimension table attributes to answer the 5 business questions. Use the datasets provided (InventorySystem.xlsx, Sales-database.xlsx and viewEmployee.xlsx) to finalize the star schema.

Fact table name: Sales_Fact
Sales_ID (Primary Key)
Product_ID (Foreign Key)
???_ID (Foreign Key)
???_ID (Foreign Key)
???_ID (Foreign Key)

Date_ID (Foreign Key) **???** (Measure) **Quantity** (Measure) **Repeat_???? eg (0 = First Purchase, 1 = Repeat Purchase)** **(Add more measures/columns) if needed**

Task: You are required to complete the fact table by adding any missing metrics (measures) such as Total_Sales_Per_???? etc based on the 5 business questions.

Task instructions

1. **Complete the Fact table** - Based on the 5 business questions, fill in the missing attributes in the fact table. You are expected to figure out which metrics and foreign keys are needed to answer these questions (e.g., Total_Sales_Per_???? etc).
2. **Complete the Dimension tables** - Identify and complete the necessary dimension tables (e.g., Product dimension etc) by adding the appropriate dimensions and attributes to support the analysis.
3. **Use the datasets** - The datasets provided (InventorySystem.xlsx, Sales-database.xlsx and viewEmployee.xlsx) contain the information needed to complete the star schema. You are required to analyze the datasets to determine the required attributes.
4. **Create the Star Schema drawing in Visio or a similar tool**: Provide a drawing of your star schema in MS Visio (or any relevant tool), showing the relationships between the fact and dimension tables, including primary keys, foreign keys and appropriate constraints (e.g., NOT NULL, UNIQUE). Ensure that your username and the timestamp of your computer are visible in the drawing.

Provide a drawing of the Star Schema in MS Visio (or another relevant tool). Submit the drawing of the Star Schema with your username and timestamp of your computer showing.

MARKING GUIDE – CHECK THAT YOUR WORK MEETS THE FOLLOWING.

- ☑ **Completion of Fact table (10 Marks)**
  - ☐ Correctly identify and fill in the missing measures and foreign keys in the fact table based on the five business questions.
  - ☐ Accurately determining the necessary measures (e.g., Total Sales etc) aligning with the 5 business questions.

- Correctly identifying the foreign keys linking the fact table to the appropriate dimension tables.

- **Completion of Dimension table (15 Marks)**
  - Complete the necessary dimension tables (e.g., Product, etc) ensuring they contain all the attributes needed to answer the 5 business questions.
  - Including key attributes and correctly linking the dimension tables to the fact table via foreign keys.
  - Ensure all foreign keys and attributes correctly support answering the 5 business questions.

- **Completion of Star Schema drawing (15 Marks)**
  - Clearly show the primary keys and foreign keys in the schema, ensuring that all relationships between the fact and dimension tables are represented.
  - Correct use of constraints and the accurate representation of relationships between tables.
  - Include a legend for the schema, clearly explaining primary keys, foreign keys, measures and constraints (examples of constraints include e.g. NOT NULL, UNIQUE, etc used to ensure data integrity)
  - Ensure your username and timestamp are visible in the drawing.

Key points to check

- Ensure the fact table has the appropriate measures
- Confirm the necessary foreign keys to dimension tables (e.g., Product_ID, etc).

- Include a clearly labelled Legend.
- Validate that the star schema can answer all five business questions.
- The drawing must include clear relationships between tables and be well labelled.
- The drawing includes your username and timestamp. All submitted screenshots and diagrams must show username and timestamp to receive full marks.
- Full marks only if all elements are included.
- Partial marks for missing elements.
- Zero marks for missing username and timestamp.
- Zero marks for emailed work.

B. **Implementation of the data warehouse** schema **in Oracle SQL Developer**. Based on the **Star Schema** you created in Task 1. A create the physical data warehouse structure on rhea.

   Create a fact table and appropriate dimensions tables and appropriate data types, primary keys, foreign keys and indexes where applicable to form a star schema model. Then import the data into the respective dimensions in Oracle SQL Developer and provide screenshots of the resulting output from Oracle SQL Developer **(10 Marks).**

   REMINDER: ALL screenshots must include your username together with your computer's timestamp to receive marks (**zero marks will be awarded** if your username and your computer's timestamp are not included in your screenshots).

MARKING GUIDE – CHECK THAT YOUR WORK MEETS THE FOLLOWING.

- The tables you create in Oracle SQL Developer must match the above design with all appropriate data types and constraints.
- Indexes and keys are correctly implemented.
- Data is accurately imported into the respective tables.
- Provide screenshots of all tables and relations, PKs, FKs and constraints clearly shown.
- Screenshots show the successful import and structure of the tables with username and timestamp.
- Grant the appropriate permissions to the marking account, “MARKERTL”

TASK 2. Extend the star schema in Task 1 based on the following (30 marks)

A. **Extend** the star schema you created in Task 1 to include **an additional dimension or fact table** that you believe would add value to the business analysis. This means **you will need to create and populate your own dataset**, such as by adding data related to marketing campaigns, customer reviews, return rates and add (extend) this dataset to the star schema you created in Task 1 **(20 Marks).**

   HINT: To extend your Star Schema, you could consider adding a **dimension related to customer behavior**, for example customer feedback, loyalty program participation or marketing campaign data to provide additional insights into customer trends.

B. Submit a drawing of the extended star schema. For the screenshot of the star schema, including the additional table in MS Visio (or another relevant tool), ensure that the screenshot not only shows the username with your computer timestamp. The screenshot must include both the original star schema from Task 1 and the extended portion(s) **(10 Marks).**

   Provide a screenshot of the extended star schema drawn in MS Visio with your username and timestamp of your computer indicated. ALL **screenshots <u>MUST</u>** show **your username together with your computer's timestamp** (or zero marks will be awarded)

MARKING GUIDE – CHECK THAT YOUR WORK MEETS THE FOLLOWING.

- This question requires you to add new elements to the star schema you created in Task 1 – think of any additional dimension(s) that would add new insight. It can be anything (see *How to Formulate BI Questions - Tips for Assignment 1*)
- Once you add the new element (dimension(s)) – you may need to modify your fact table to include additional information – this is part of the assessment test.
- For this task you are only required to provide a Drawing of the newly extended schema.
- The drawing must show the original and new elements.
- Provide a brief explanation of how the extended schema will add value and bring new insights (see *How to Formulate BI Questions - Tips for Assignment 1*).

- ☑ Check that the screenshot of your drawing includes your username and computer's timestamp.

TASK 3. Formulate two Business Intelligence questions (30 marks)

A. Formulate two business intelligence questions that can be answered using your **new dataset and extended data warehouse schema**. (Note: This question is asking you to formulate a business intelligence question) **(20 Marks).**

Note: To assist you in formulating a BI question, **here is an example of a business intelligence question from Task 1** "Which **states** have the **highest number** of **unique customers**?" **Hint:** To formulate a good BI question, think about how new dimensions or measures that could lead to new insights such as,

- Which marketing campaigns had the greatest impact on repeat purchases?
- Which geographical regions experience the highest customer churn rates?

These are examples and cannot be used in your assignment.

TASK: List the two business intelligence questions you have formulated.

B. For each question you have formulated, provide a concise explanation how the answer could support business decision-making. A brief paragraph for each question should suffice. **(10 Marks).**

MARKING GUIDE – CHECK THAT YOUR WORK MEETS THE FOLLOWING.

- ☑ This question requires you to 1) formulate two BI questions 2) List the two BI questions and 3) describe how the insights obtained could assist an organization in making informed decisions. For instance, the analysis might reveal

opportunities to increase bicycle sales in warmer cities based on weather patterns and sales data.

- Refer to the document "*How to Formulate BI Questions - Tips for Assignment 1*" for guidance.
- **TIP! Here is another example of a BI Question and how to extend your schema** (What are the **top 10 hottest cities** and what is the **average sales** amount **per customer** in those **states**) – to answer this question you will need to integrate weather data into your schema.
- The two BI questions are well formulated and can be answered using the extended schema. They must directly relate to the added data and insights from the schema extension.
    - For example, using the above example *What are the **top 10 hottest cities** and what is the average sales amount per customer in those states* – since weather data is not initially available, you are required to procure this data from an external source (e.g., Lab01). You will then need to import the relevant tables/columns and integrate this new dimension with your existing schema effectively.
- Each question should reflect an understanding of the data model and its analytical capabilities.
- 10 marks per question.
- 10 marks for explanation – The explanation shows a clear understanding of how the answers to the BI questions could influence business decisions and how the additional data aligns with intended outcomes (see slide 6 of the assignment tip document)
- Refer to the document "How to Formulate BI Questions - Tips for Assignment 1" – for examples of business decisions and how to align with your BI questions.
- Well-articulated reasoning with each explanation evaluated out of 5 marks each.

- The example question (What are the top 10 hottest cities and what is the average sales amount per customer in those states) provided here is for illustrative purposes and **should not** be used in your actual assignment.

Part II Power BI Task (10%)

Sales and Inventory Data Analysis in Power BI (Total 100 marks)

TASK 1. Import data and create a Star Schema (40 marks)

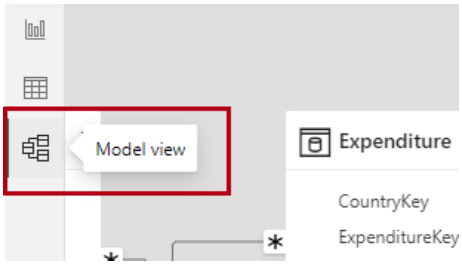
You will need the datasets - *InventorySystem.xlsx*, *Sales-database.xlsx* and *viewEmployee.xlsx*
You may need to enter dummy data in the above datasets.

A. Import the provided datasets into Power BI **(10 marks).**

B. Create a **Star Schema** with fact and dimension tables. **As part of the assignment**, you will need to **decide which fact and dimension tables are needed**. See Task 3 to determine which dimensions and facts table are needed **(10 marks).**

C. In Power BI, define relationships between the fact table and the dimension tables **(20 Marks).**

From the Model View (the Model View menu is shown in the image below) provide a FULL screenshot of the relationships between the fact table and the dimension tables in Power BI. The screenshot MUST show your username, the Model View together with your computer's timestamp.



MARKING GUIDE – CHECK THAT YOUR WORK MEETS THE FOLLOWING.

- ☑ Data from provided datasets correctly imported into Power BI – 10 marks.
- ☑ Provides an accurate representation of the star schema with correct identification of fact and dimension tables – 10 marks.

- Correctly defined relationships between tables in Power BI – 20 marks.
- Screenshot shows the model view with all relationships, username and your computer's timestamp – 0 marks for the entire task for missing username and timestamp.

TASK 2. Data transformation (20 marks)

A. Clean and transform the data, ensuring correct data types are applied, especially dates and any erroneous fields **(10 Marks).**

B. Create new **calculated columns** where necessary (e.g., Total Sale Amount in the 'SALE' table by multiplying 'ProductPrice' and 'Quantity' from the 'SALE_LINE_ITEM' table) **(10 Marks).**

MARKING GUIDE – CHECK THAT YOUR WORK MEETS THE FOLLOWING.

- Correct data types applied and any data errors resolved– 10 marks.
- Necessary calculated columns created and correctly computed – 10 marks.

TASK 3. Data analysis and visualization (40 marks)

A. Create a visualization with appropriate titles, labels showing **total sales amount** *by* **product category**. This should include appropriate titles and labels. Provide a screenshot of the visualization and a brief description or interpretation of the results **(10 Marks).**

B. Create a visualization that displays **total revenue by supplier** and **percentage contribution of each supplier** to the total sales. Provide a screenshot of the visualization and a brief description or interpretation of the results **(10 Marks).**

C. Create a visualization comparing **employee performance by total sales** and **average sales** value **per employee**. Provide a screenshot of the

visualization and a brief description or interpretation of the results **(10 Marks).**

D. Create a visualization showcasing **customer distribution by state and average sales value per customer** in **each state**. Provide a screenshot of the visualization and a brief description or interpretation of the results **(10 Marks).**

NOTE: **Each screenshot of the visualization in Task 3. A, B, C and D)** MUST show your username together with your computer's timestamp (zero marks will be awarded if the screenshots do not include the username and timestamp).
Please submit your Power BI project as a .pbix file. On LMS you are required to upload a single zip file containing only one .pbix file.

MARKING GUIDE – CHECK THAT YOUR WORK MEETS THE FOLLOWING.

- ☑ Each visualization in A, B, C and D is marked out of 10 marks – total 40 marks. Check that, the visualization for
  - □ A, helps understand how product categories contribute to the total sales, allowing them to analyze which product categories are performing best.
  - □ B, aids to see how much revenue each supplier is generating and the relative importance of each supplier to the business, helping to answer the supplier-related business questions.
  - □ C, helps to compare employees not just by total sales but also by their average sales value, providing a more insightful view of employee performance.
  - □ D, provides insights into which states have the highest number of unique customers, along with how much, on average, customers from those states are spending.
- ☑ Visualizations will be assessed for clarity, accuracy and relevance to the task – for example if the tasks require analysis of **customer distribution** by **state** or **postcode** –

submitting an analysis of **customer distribution** without the **state** or **postcode** component is considered incomplete.

- [x] Interpretations will be assessed and awarded for depth of analysis and clarity of explanation for each visualization.