|  |  |
| --- | --- |
| Name: Mithila Nigudkar \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | **Rochester Institute of Technology**  Due: 09/25/2021  **Golisano College of Computing and Information Sciences**  **School of Information** |

**PE03: Dimensional Modeling**

Exercise 1: Identifying Facts and Dimensions

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | | | |  |
| - Suppose a regional dairy products company employs you, and your task is to build the data marts for the overall data warehouse  - The company sells products to grocery stores, convenience stores, gas stations, and mass merchandisers  - You will be extracting data from the **Product Returns** operational system and **Sales Forecast spreadsheet** | | | | |  |
| - Identify each field | |  |  |  |  |
|  | D - a dimensional attribute |  |  |  |  |
|  | F - a fact |  |  |  |  |
|  | O - operational only, not to be included | | |  |  |
|  |  |  |  |  |  |
| **D/F/O** | **Product Returns** |  | **D/F/O** | **Sales Forecast** |  |
| D | Customer Account Number |  | D | Account Rep |  |
| D | Product Category |  | D | Month |  |
| D | Product Brand |  | D | Item # |  |
| D | Customer Name |  | D | Item Description |  |
| D | Product Expiration Date |  | F | Forecast Units |  |
| D | Product # |  | F | Forecast Amount |  |
| D | Product Description |  | D | Valid Forecast Flag |  |
| D | Package Type |  |  |  |  |
| D | Plant Number |  |  |  |  |
| D | Manufacturing Line |  |  |  |  |
| O | Regular/Low fat |  |  |  |  |
| D | Customer Ship to Street Address |  |  |  |  |
| D | Customer Ship to City |  |  |  |  |
| D | Customer Ship to State |  |  |  |  |
| D | Customer Ship to Country |  |  |  |  |
| D | SKU (Stock Keeping Unit) |  |  |  |  |
| F | Returned Quantity |  |  |  |  |
| D | Returned Reason |  |  |  |  |
| F | Expired Quantity |  |  |  |  |
| F | Damaged Quantity |  |  |  |  |
| D | Damaged Code |  |  |  |  |
| D | Returned Date |  |  |  |  |
| D | Sales Rep |  |  |  |  |
| D | Sales Region |  |  |  |  |
|  |  |  |  |  |  |

Exercise 2: Identifying Dimensions and Fact Groups

Using the same extract files that you used in Exercise #1, identify the following:

* Possible dimensions
* Possible fact groups (facts in each data mart)

**Possible Dimensions**

Customer\_Dinension , Product\_Dimension, Return\_Dimension, Sales\_Dimension , Expiry\_Dimension, Forecast\_Date\_Dimension , Return\_Date\_Dimension.

**Fact Groups**

Product\_Return\_Fact , Sales\_Forecast\_Fact

Exercise 3: Designing Dimensions

Design (i.e., draw a diagram of) each of the dimensions that were identified in Exercises 1 & 2. Follow the dimension representation shown on slide #8 of Week 4 Lecture notes). Specifically:

* Identify dimension attributes
* Identify all hierarchies of the attributes within a dimension

1. **Customer\_Dimension**

Customer\_SK(PK) , Customer\_account\_number(NK), Customer\_name, Shipping\_address, Shipping\_state, Shipping\_city, Shipping\_country, Package\_type.

**Hierarchies of customer dimension are as follows:**

Shipping\_country ----🡪 Shipping\_state ---🡪 Shipping\_city ---🡪 Shipping\_address.

1. **Product\_Dimension**

Product\_SK(PK) , Item#(NK), Description, Catoegry, Brand, SKU, Plant\_number, Manufacturing\_line.

**Hierarchies of product dimension are as follows:**

Category ---🡪 Brand ---🡪 SKU ---🡪 Item#

1. **Sales\_Dimension**

Sales\_Sk(PK), Sales\_region, Sales\_rep, Account\_rep.

**Hierarchies of sales dimension are as follows:**

Sales\_region ---🡪 Sales\_rep ----🡪 Account\_rep

1. **Forecast\_Dimension**

Forecast\_date\_SK(PK), Forecast\_month, Forecast\_year, Forecast\_day

**Hierarchies of forecast dimension are as follows:**

Forecast\_year ---🡪 Forecast\_month ---🡪 Forecast\_day

1. **Expiry\_Dimension**

Expiry\_date\_SK(PK) , Expiry\_month, Expiry\_day, Expiry\_quarter, Expiry\_year, Expiry\_date

**Hierarchies of expiry dimension are as follows:**

Expiry\_year ---🡪 Expiry\_quarter ---🡪 Expiry\_month ---🡪 Expiry\_day --🡪 Expiry\_date

1. **Return\_Dimenion**

Return\_Sk (PK) , Damaged\_code, Return\_reason.

**Hierarchies of return dimension are as follows:**

There are no hierarchies for return dimension.

1. **Return\_Date\_Dimension**

Return\_date\_SK(PK), Return\_quater, Return\_month, Return\_day, Return\_year, Return\_date.

**Hierarchies of return date dimension are as follows:**

Return\_year ---🡪 Return\_quarter ---🡪 Return\_month --🡪 Return\_day --🡪 Return\_date.

Exercise 4: Designing Fact Groups

Design each of the fact groups that were identified in Exercises #1 & #2. Specifically, for each fact group (data mart):

* list the facts that relate to the process that the fact group represents
* write a description for the fact – i.e., define it
* state the default aggregation rule (“sum” if additive; “semi-additive over time” if semi-additive; “N/A” if non-additive)

|  |  |  |
| --- | --- | --- |
| Fact Group: Sales\_Forecast\_Fact \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | | |
| Fact Name | Fact Description | Default Aggregation Rule |
| Forecast\_amount | Forecasted sales amount | SUM |
| Forecast\_units | Forecasted sales unit | SUM |
|  |  |  |
|  |  |  |
|  |  |  |

|  |  |  |
| --- | --- | --- |
| Fact Group: Product\_Return\_Facts\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | | |
| Fact Name | Fact Description | Default Aggregation Rule |
| Expired\_quantity | Number of items expired | SUM |
| Damaged\_quantity | Number of items damaged | SUM |
| Returned\_quantity | Number of items returned | SUM |
|  |  |  |
|  |  |  |

|  |  |  |
| --- | --- | --- |
| Fact Group: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | | |
| Fact Name | Fact Description | Default Aggregation Rule |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

Exercise 5: Create the Data Mart Matrix

The data mart matrix shows the relationship between the possible data marts and dimensions. Any dimension (column) with more than one X implies that this dimension must be conformed across multiple data marts to fit into the Data Warehouse Bus Architecture.

Fill in the data mart matrix using the following table:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Data Mart** | Customer | Sales | Return | Product | Return\_Date | Expiry\_Date | Forecast\_Date |  |  |  |  |  |  |  |  |
| Product\_return | X |  | X | X | X | X |  |  |  |  |  |  |  |  |  |
| Sales\_Forecast |  | X |  | X |  |  | X |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

Exercise 6: Logical Table Design

Use the dimensional models that you have created so far to:

* Design the actual star schema for each of the fact groups that you defined in Exercise #4.
* Create your Dimensional Models using MySQL Workbench and save it as a pdf file.
* Submit a zip file containing 1) a copy of the answered PE03 and 2) pdf file of EER diagrams to MyCourses PE03 Dropbox. Check the deadline on the dropbox.

Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Graded By: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**PE03: Dimensional Modeling Grade sheet**

|  |  |  |
| --- | --- | --- |
| **Requirements** | **Grade** | **Grade Earned** |
| **Exercise 1: Identifying Facts and Dimensions**   * Identify all the attributes correctly   **Exercise 2: Identifying Dimensions and Fact Groups**   * Dimensions * Fact Groups   **Exercise 3: Designing Dimensions**   * Identify all dimensions’ attributes * Identify all hierarchies of the attributes   **Exercise 4: Designing Fact Groups**   * Fact groups * All the facts are included with the fact groups * Fact descriptions and aggregation rules   **Exercise 5: Create the Data Mart Matrix**  **Exercise 6: Logical Table Design**   * Dimensions * Fact tables * Conformed dimension * Correct primary keys and foreign key constraints * Relationships between fact tables and dimensions * Submit star schema | 10  6  4  8  5  2  5  5  9  16  8  8  5  9  -20 |  |
| **Total Grade:** | **100** |  |