Mbarara University of Science and Technology Faculty of Computing and Informatics

MASTER OF SCIENCE IN INFORMATION SYSTEMS

TUSHABE BRUCE - 2022/MSIS/017/PS

DATA WARE HOUSING ASSIGNMENT

An organization comprises at least two units that contain employee data and sales important to the organization among others as shown below:

Employee Personal Information:

Last Name: STRING (40)

First Name: STRING (15)

Sales: STRING (3) with the following permissible values:

“NR” – New Recruit

“ESP“ - Experienced Sales Person”

“NSP” – Not Sales Person

Information on Organizations Part Timers:

Last Name: STRING (25)

First Name: STRING (10)

Employee Status: STRING (10) with the following permissible values:

“Newly joined the company”

“Has been in company for a long time”

Academic Status: STRING (13) with the following permissible values:

“Good standing”

“Probation”

a) You are responsible for the integration work necessary to build a single “master list of students” in the data warehouse. To do this:

i. Specify the data transformations necessary as part of the ETL.

ii. Add any 3 to 4 units to the organization (Data Marts)

b) Use fact and dimentional tables (Or any other means) to simulate the resulting data warehouse

ANSWERS

a) To build a single "master list of employees" in the data warehouse and perform the necessary data transformations as part of the ETL process, the following steps can be taken:

i. **Data Transformations:**

1. Merge Employee Personal Information and Information on Organizations Part Timers based on a common unique identifier, such as an employee ID or a combination of first name and last name.

2. Standardize the data types for consistent storage and analysis. For example, convert the Last Name and First Name fields to a uniform string length as specified.

3. Convert the Sales field into a numerical representation for better analysis. Map the permissible values "NR," "ESP," and "NSP" to corresponding numerical values (e.g., 1, 2, 3).

4. Transform the Employee Status field into a binary representation. Map the permissible values "Newly joined the company" and "Has been in company for a long time" to binary values (e.g., 0 and 1).

5. Transform the Academic Status field into a binary representation. Map the permissible values "Good standing" and "Probation" to binary values (e.g., 0 and 1).

6. Handle missing or null values by assigning default values or applying appropriate data imputation techniques.

ii**. Data Marts:**

1. Sales Performance Data Mart: This data mart can store information related to sales performance, including sales figures, salesperson details, and relevant metrics.

2. Employee Information Data Mart: This data mart can contain comprehensive employee information, including personal details, academic status, employee status, and any other relevant attributes.

3. Time Dimension Data Mart: This data mart can store time-related information, such as dates, months, years, and other temporal attributes, to facilitate time-based analysis.

4. Organization Hierarchy Data Mart: This data mart can store information about the organizational structure, including departments, teams, reporting relationships, and hierarchical levels.

b) Simulating the resulting data warehouse can be achieved using fact and dimensional tables. Here's a simplified representation:

1. Fact Table: EmployeeSalesFact

- EmployeeID (foreign key referencing Employee dimension)

- SalesID (foreign key referencing Sales dimension)

- DateID (foreign key referencing Time dimension)

- SalesAmount

2. Dimension Tables:

a. EmployeeDim

- EmployeeID (priy key)

- LastName

- FirstName

- EmployeeStatus

- AcademicStatus

b. SalesDim

- SalesID (primary key)

- SalesType

c. TimeDim

- DateID (primary key)

- Date

- Month

- Year

d. OrganizationDim

- OrganizationID (primary key)
- Department
- Team
- ReportingRelationships
- HierarchicalLevel

These tables form the basis of the data warehouse structure. Additional attributes and tables can be added based on specific requirements and data sources within the organization. The fact table captures the sales information, while the dimension tables provide descriptive information related to employees, sales, time, and organization hierarchy, enabling comprehensive analysis and reporting capabilities.

The code for this is in the zip