

Database Creation and Dimension Table / PostgreSQL

Didem B. Aykurt

Colorado State University Global

MIS541; Data Warehousing in Enterprise Environments

Dr.Peter Salemi

April 2, 2023

Create the Northwind Data Warehouse

This project explains the data warehouse and dimensional table with Northwind_DW database example.

First, let's define what a data warehouse is, how a data warehouse works, how many types of data warehouses there are, and why/who needs a data warehouse. A data warehouse is best used for storing, collecting, and managing structure data from diverse sources such as application, business, transactional data, and batch reporting that have provided meaningful business realization with a specific purpose in mind, like data mining for BI or data warehouse professionals, or business analyst.

There are three types of data warehouses: an enterprise data warehouse centralizes a business' information from multiple sources, and an application warehouse helps know about the company, employees, customers, and more. The operational data store (ODS) is primarily used in current operations and housed before being transferred to the data warehouse for long-term storage, archiving, and real-time refresh. Finally, data mart sets each business purpose data warehouse in an independent data mart, such as sales or finance data, which can be stored directly from sources.

Second, dimensional tables contain textual structures like product dimensions with names, product details, unit prices, colors, weights, and other text details. The time dimensions could have year, quarter, month, week, day, and hour to be accurate since the company started operating. Thus, the dimension table answers the following questions about the event: who, what, where, when, how, and why.

The data warehouse is mainly used for supplying, transporting, marketing, etc. I will create the Northwind_DW data warehouse used for retaining chains in PostgreSQL. Right-click on the Databases (2)-Create-Database, then pop up the new page and name Northwind_DW. And I created two-dimension tables "NW_Customer_DIM" and "NW_Employee_DIM."

Identify the "NW_Customer_DIM" Dimensions:

- Customer_id (PK)

- Company_name
- Contact_name
- Contact_title
- Address
- City
- Region
- Postal_code
- Phone
- Fax

Identify the “NW_Employee_DIM” Dimensions:

- Employee_id (PK)
- Last_name
- First_name
- Title
- Title_of_courtesy
- Birth_date
- Hire_date
- Address
- City
- Region
- Postal code
- Country
- Home_phone
- Photo

- Notes
- Reports_to(FK)
- Photo_path

SQL code command

Creating a dimensional table: open a new query below the list showing all SQL syntax and statements I used to create the table.

'--' is writing a comment to explain the query, and SQL can't read when code running will not be executed.

';' semicolon is the end of each SQL statement.

The 'Drop Table If Exists' statement deletes an existing table in a database to eliminate the duplicate table.

The 'Create Table()' statement creates a new table in a database.

The 'Alter Table' statement changed the definition of an existing table and is used to add, drop, and modify different constraints on a current table.

'character varying (n)' is a character type that can store string values up to n characters in length.

'text' is a string function that contains text type values.

'smallint' is a numeric type that can store numeric values ranging from -32768 to +32767.

'date' is the date type that only contains the date part of the date column in the format YYYY-MM-DD.

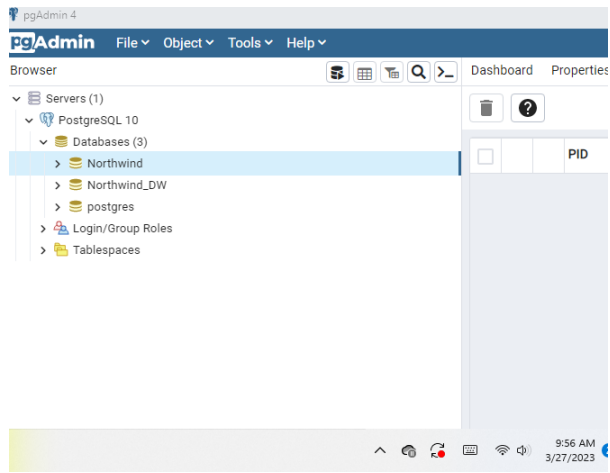
'bytes' is a binary string function containing binary strings; that value type is bytea.

The 'Not Null' operator tests it and selects a non-empty value.

'Primary Key' contains a unique value that individually identifies each row in a table.

'Foreign Key' connects the child to the parent table. If the table has a foreign key, it is called the child table, and the table with the primary key is called the parent or reference table.

Figure 1: Create a Northwind_DW data warehouse in PostgreSQL.



Create the Dimensional Tables

Figure 2: Create “NW_Customer_DIM” dimensional table.

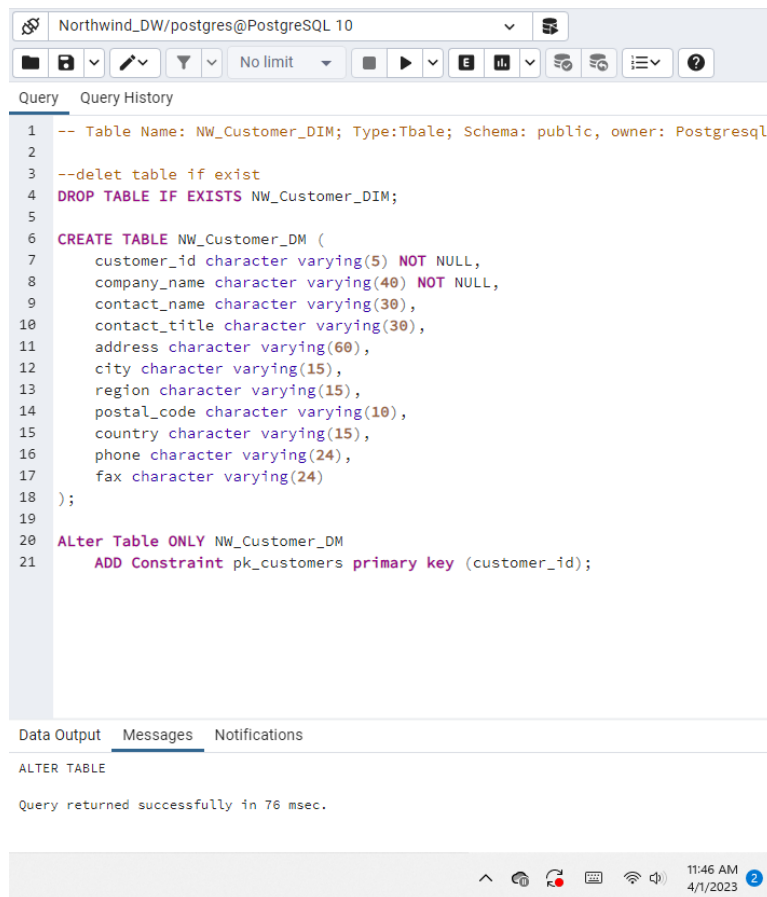


Figure 3: Create an “NW_Employee_DIM” dimensional table.

```

1  -- Name:NW_Employee_DIM; Type:Table; schema:public, owner:Postgresql
2
3  --drop table
4
5  DROP TABLE IF EXISTS NW_Employee_DIM;
6
7  CREATE TABLE NW_Employee_DIM
8  (
9      employee_id smallint NOT NULL,
10     last_name character varying(20) NOT NULL,
11     first_name character varying(10) NOT NULL,
12     title character varying(30),
13     birth_date date,
14     hire_date date,
15     address character varying(60),
16     city character varying(15),
17     region character varying(15),
18     country character varying(15),
19     home_phone character varying(24),
20     photo bytea,
21     notes text
22 );
23
24 ALTER TABLE NW_Employee_DIM
25     ADD CONSTRAINT pk_employees Primary Key (employee_id);

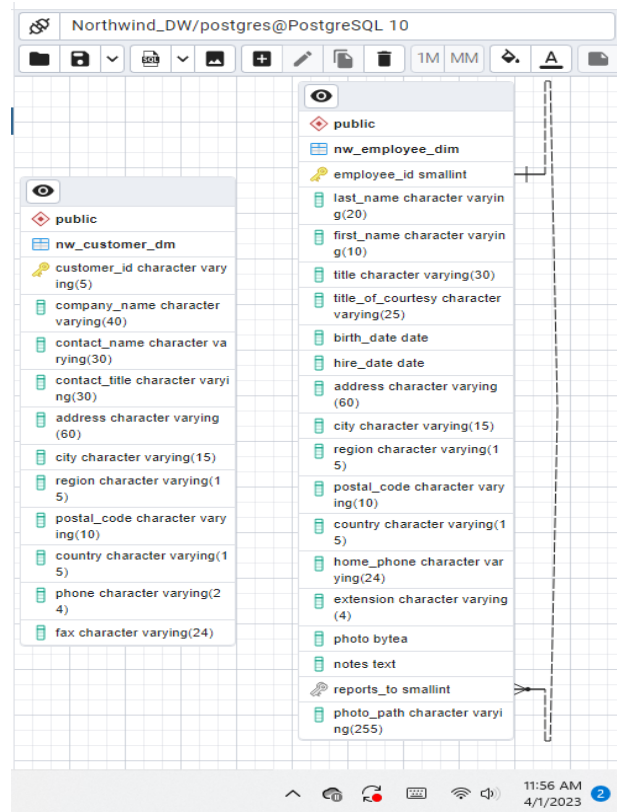
```

Data Output Messages Notifications

ALTER TABLE

Query returned successfully in 68 msec.

Figure 4: Entity-Relationship Diagram for both two tables.



References

David Taylor, 2023. What is the data warehouse? Type, Definition & Example. <https://www.guru99.com/data-warehousing.html>

Azura.Microsoft.com. What is a data warehouse? <https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-a-data-warehouse/#data-warehouse-definition>

Precisely.com. Enterprise Data Warehouse (EDW). <https://www.precisely.com/glossary/enterprise-data-warehouse>

Ben Lutkevich, 2023. The operational data store (ODS).
<https://www.techtarget.com/searchoracle/definition/operational-data-store>

Rick Sherman, 2015. Business Intelligence Guidebook. <https://www.sciencedirect.com/topics/computer-science/customer-dimension>

PostgreSQL.org <https://www.postgresql.org>

Ralph Kimball & Margy Ross, 2013. The Data Warehouse Toolkit The Definitive Guide to Dimensional Modeling. 3rd edition.

Pascal Thomet , 2022. Northwind_psql/ northwind.sql.
https://github.com/pthom/northwind_psql/blob/master/northwind.sql