Lab 1 – Building the TPC-E Order Database

Objective

The objective of this lab is to refresh your skills in MySQL and acquaint you with another tool called MySQL Workbench. MySQL Workbench can be used to manage a MySQL instance and will be used in the remainder of the course.

Description

As you noticed in the description of The Product Company (TPC) and its desire for a financial data warehouse, TPC-E already has a database in place for its order data. In this lab, you are going to create that database using MySQL Workbench and load it with data, which we will provide in a series of comma-separated files (.csv).

This data will become the basis for the data that TPC-E will provide to the corporate data warehouse. The other organizations will provide their feeds of data that will be discussed in future labs. You will develop this database on your machine, or the software is available on the database lab machines.

MySQL Workbench

In the lab, the software is already installed, so you will not need to install it. On your machine, MySQL Workbench can be downloaded from the following website:

http://dev.mysql.com/downloads/workbench/

For MS Windows users, download the MSI installer. For other operating systems, use the appropriate version.

MySQL workbench assumes you already have MySQL installed and running. If you have not done so already, you should have a MySQL server running on your machine. You can download the current GA version, if necessary. It can be found here:

http://dev.mysql.com/downloads/mysql/#downloads

Documentation

There are excellent high-level tutorials here. There are three tutorials to read – read them in the order listed here:

http://wb.mysql.com/?page_id=10

Read the tutorials in this order:

- Get to know MySQL Workbench
- Database Training Blog – Data Modeling Using MySQL Workbench 1 – 3 (links to parts 2 and 3 are in a response at the bottom.)
- Visual Database Design in MySQL Workbench

Finally, if you search for "MySQL Workbench tutorial" (without quotes) on Google, you will find many tutorials, including some YouTube videos.

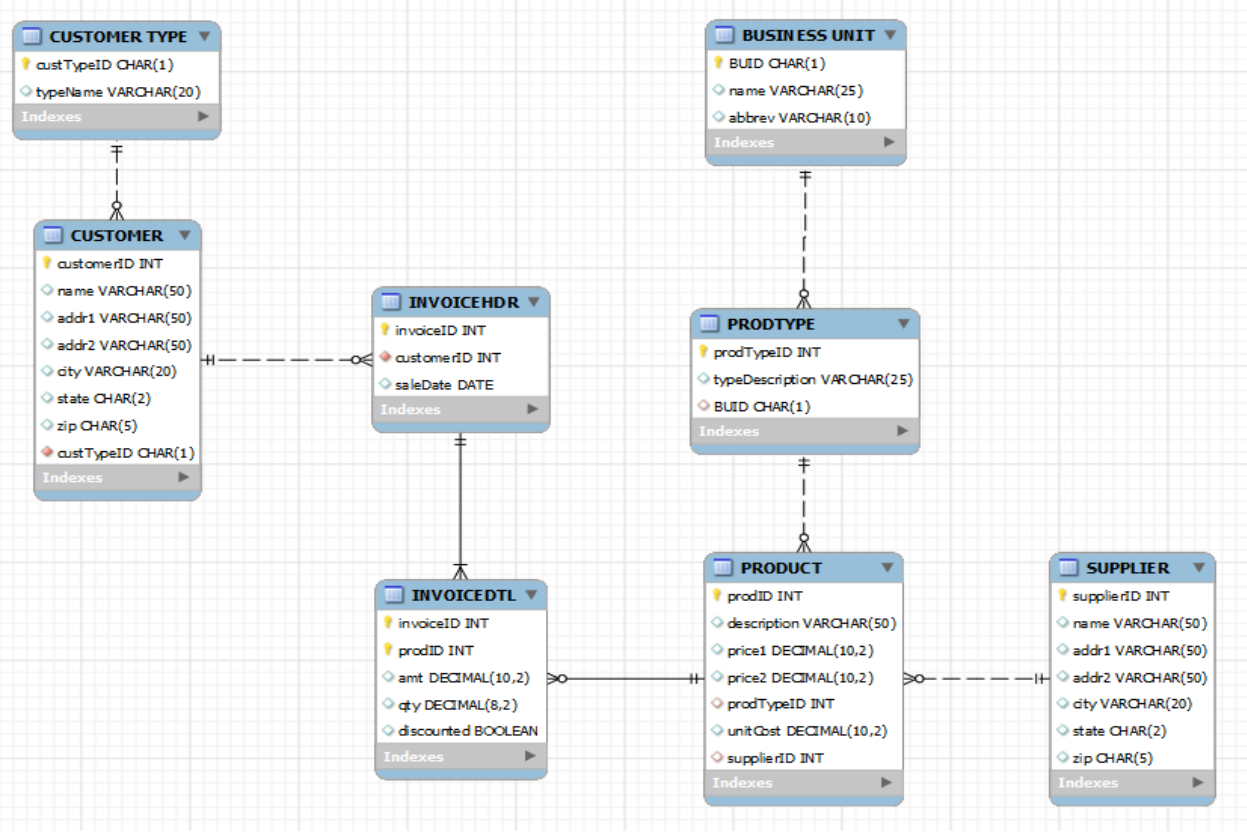
The MySQL Workbench documentation is located here:

http://dev.mysql.com/doc/workbench/en/index.html

The Lab

In this lab, you will build the online transaction processing (OLTP) database used by TPC-E. The following E-R diagram describes the database:

Database name: *YourLastName_SalesOrders*



A data dictionary is provided in **appendix A**. All other information, including primary keys and *foreign keys,* are included on the E-R diagram.

Step 1:

Using MySQL Workbench, build the database as described above, including all tables and attributes. You should name the database *YourLastName*_SalesOrders. Table and attribute names must conform to the E-R diagram. You may want to consider waiting to add the foreign key constraints until the data is loaded. This will help data to load faster. They can be added after the data is loaded using ALTER statements. The InvoiceDtl table is the largest and will have approximately 100,000 rows.

Step 2:

Load the data from the files provided. A description of the data can be found in **Appendix B**. To load the data, and you use the LOAD DATA INFILE … INTO MySQL Statement. If ALTER statements are used for the foreign keys, then you should include them here also. Documentation for the LOAD statement can be found here:

https://dev.mysql.com/doc/refman/5.8/en/load-data.html

Note: If you run into security related errors from using the LOAD DATA statement, you can use the Table Data Import Wizard in the MySQL Workbench.

Step 3:

Write SQL statements to produce the following:

Query #1. Show the information for all invoices for a particular customer (you pick the customer). The result set should include customer name, address, city, state, zip, invoice ID, sales date, product descriptions, amount, and quantity.

Query #2. Show the total sales amount for each business unit by year and business unit abbreviation.

Deliverables

Submit the following to MyCourses as a single .zip file with the following name – *YourLastName*.zip:

1. Submit a pdf file of your MySQL Workbench EER diagram.
2. Submit a copy of your .mwb file from MySQL Workbench.
3. Your SQL Queries from Step 3 above.
4. Screenshot of the first page of the result set for the above two queries.
5. A .sql file (*YourLastName_Lab01dump*.sql) that contains the dumping of the *YourLastName*_SalesOrders database
   (Use mysqldump to create a dump file.
   Ref. https://dev.mysql.com/doc/refman/5.8/en/using-mysqldump.html)

©RIT GCCIS iSchool

Appendix A – Data Dictionary

Attribute Name	Table name	Data type	Description	Sample
abbrev	BusinessUnit	varchar(10)	The business unit abbreviation	
addr1	Customer	varchar(50)	The first address line of the customer	
addr1	Supplier	varchar(50)	The first address line of the supplier	
addr2	Customer	varchar(50)	The second address line of the customer	
addr2	Supplier	varchar(50)	The second address line of the supplier	
amt	InvoiceDtl	decimal(10,2)	The dollar amount of the sale	22222222.22
buID	BusinessUnit	char(1)	The product type ID	
buID	ProdType	char(1)	The product type ID	
city	Customer	varchar(20)	The city of the customer	
city	Supplier	varchar(20)	The city of the supplier	
customerID	Customer	int	The ID of the customer	
customerID	InvoiceHdr	int	The ID of the customer	
custTypeID	CustomerType	char(1)	The ID for the CustomerType table.	
custTypeID	Customer	char(1)	This is the ID for the CustomerType table.	
description	Product	varchar(50)	The descriptioion of the product	
discounted	InvoicdeDtl	boolean	Indicate whether the product is discounted or not.	
invoiceID	InvoiceHdr	int	The ID of the invoice (synonomous with invoice number)	
InvoiceID	InvoiceDtl	int	The ID of the invoice (synonomous with invoice number)	
name	Customer	varchar(50)	The name of the customer	"Chemfix Technologies Inc"
name	BusinessUnit	varchar(25)	The business unit name	
name	Supplier	varchar(50)	The name of the supplier	"Chemfix Technologies Inc"
price1	Product	decimal(10,2)	The standard price of the product per unit	22222222.22
price2	Product	decimal(10,2)	The discounted price of the product per unit	22222222.22

©RIT GCCIS iSchool

prodID	InvoiceDtl	int	The ID of the product	
prodID	Product	int	The ID of the product	
prodTypeID	ProdType	int	The ID for the product type	
prodTypeID	Product	int	The ID for the product type	
qty	InvoiceDtl	decimal(8,2)	The quantity purchased	222222.22
saleDate	InvoiceHdr	Date	The date the invoive was produces	
state	Customer	char(2)	The offical USPS state abbreviation	NY
state	Supplier	char(2)	The offical USPS state abbreviation	NY
supplierID	Product	int	The ID of the supplier	
supplierID	Supplier	int	The ID of the supplier	
typeDescription	ProdType	varchar(25)	The description got the profuct type	
typeName	CustomerType	varchar(20)	The type of customer	"Commercial", "US Govt"
unitCost	Product	decimal(10,2)	The standard cost of the product per unit	22222222.22
zip	Customer	char(5)	The 5 digit USPS zip code	14623
zip	Supplier	char(5)	The 5 digit USPS zip code	14623

©RIT GCCIS iSchool

Appendix B – Description of the Data Files

The data files contain data that can be loaded into the database once it is created. Each of the data files has a first row that describes the columns in the file.

The following is the layout of the data files:

Business Unit – business_unit.csv

Business unit data describes a company unit that is responsible for a certain collection of product types.

buID
name
abbrev

4 rows

Product Type – prod_type.csv

Product type data describes types of products. There are several products for each type.

prodTypeID
typeDescription
buID

14 rows

Product – product.csv

Product data provides a master list of the products sold by the company.

prodID
description
price1
price2
prodTypeID
unitCost
supplierID

100 rows

©RIT GCCIS iSchool

Supplier – supplier.csv

Supplier data lists the supplier of the products sold by the company.

supplierID
name
addr1
addr2
city
state
zip

10 rows

Invoice Header – invoice.csv

The invoice header contains information common to the entire invoice such as date and customer.

invoiceID
customerID
salesDate

20,000 rows

Invoice Detail – invoice_details.csv

Each invoice has a series of detail lines indicating the individual products on the invoice.

invoiceID
prodID
amt
qty
discounted

95,134 rows

Customer Type – customer_type.csv

custTypeID
typeName

©RIT GCCIS iSchool

4 rows

Customer – customer.csv

The customer file contains the list of customers serviced by the company.

custID
name
addr1
addr1
city
state
zip
custTypeID

41 rows

©RIT GCCIS iSchool