Querying the Data Warehouse Using Microsoft Access

I. What is a Data Warehouse?

A data warehouse stores data from current and previous years that have been extracted from the various operational databases of an organization. It is a central source of data that have been screened, edited, standardized, and integrated so they can be used by various end users for a variety of forms of analysis, research, and decision support.

Data warehouses provide information to data without affecting online transaction processing systems.

Some characteristics of data warehouses include:

a. Data warehouses are typically large collections of meaningfully interrelated databases. By bringing these databases together in one central warehouse, users can pick and choose among a variety of data.
b. Data warehouses are often updated frequently, some are not. Data warehouses are usually unsuitable for information that must be updated within a few minutes.
c. Databases are usually organized around a specific department. Data warehouses, in contrast, are designed to get a view of the entire organization.[2]

II. The Ohio University Data Warehouse

The Ohio University Data Warehouse was first rolled out in 1996, running on the Informix database application. For the first time, users were able to create their own reports, charts, and spreadsheets, as well as download data from the Data Warehouse to address their specific information needs.

In 1999, the data warehouse was migrated to an Oracle environment, environment running on an IBM OS/390 mainframe. Moving the warehouse to the same server as the legacy systems allowed for the reduction of the load time by avoiding the need to transfer data through the network.

As of June 20, 2000 the data warehouse consisted of more than 75 tables including the Student Information, Admissions, Student Accounts, Student Financial Aid, Housing, Pre-College and Human Resources data. Currently the Data Warehouse is being expanded, to include more data, and being optimized to offer a better performance to our users.

Current information on the Ohio University Data Warehouse may be obtained from the following web site:

http://www.ohiou.edu/datawarehouse/

III. Open Database Connectivity

Open Database Connectivity (OBDC) is a set of standards that ensures that software written to comply with these standards can be used with any OBDC complaint database. If software conforms to this standard, data can be can be imported or linked to other applications.

Both Oracle, the database application on the Data Warehouse server, and Microsoft Access, a desktop database application are OBDC compliant. Therefore, users can view and query tables residing in the warehouse on their desktop using MS-Access.

IV. Creating a Database linked to the Data Warehouse

The Ohio University Data Warehouse Utility (OUDW Utility) allows you to create a database, stored on your computer, that links to the data warehouse using ODBC. The utility program includes all accessible tables from the data warehouse in your database.

To create a new database, perform the following:

1) Trace from the **Start** button to **Programs**, then to **OUDW Utility** folder, then to the **ODUW Utility** program. The program's menu will appear as illustrated in **figure 001**

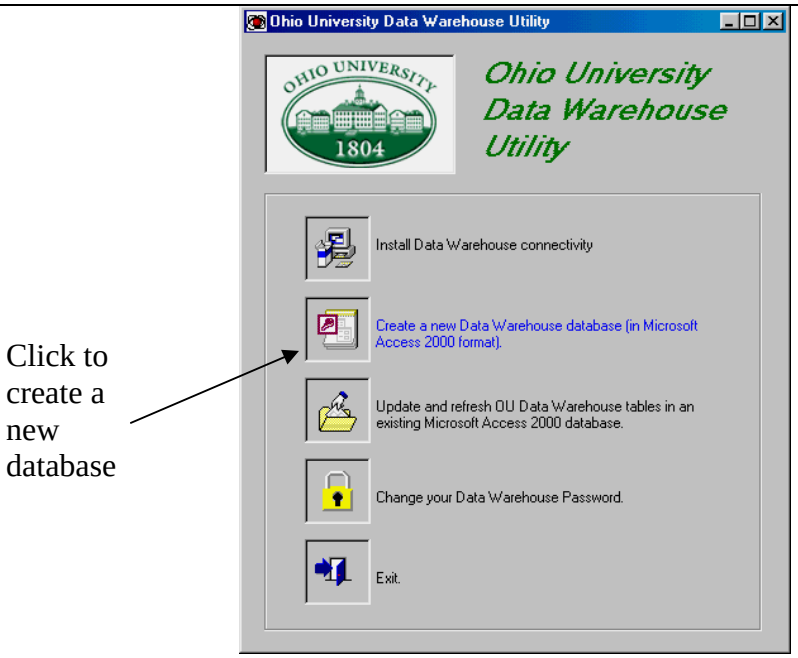
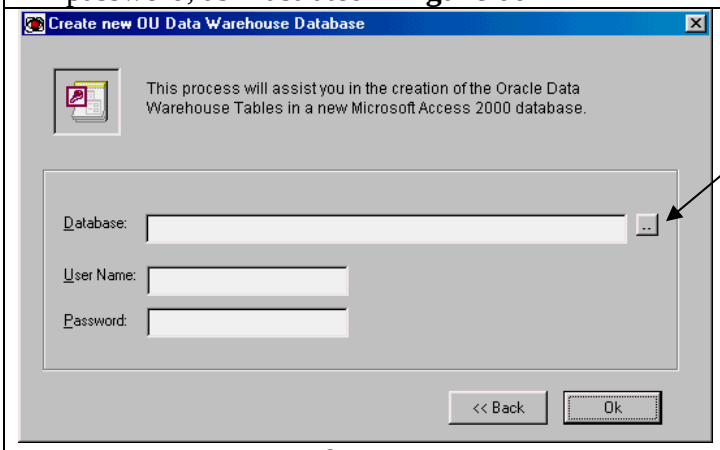


figure 001

2) Using your mouse, left click on the selection "**Create a new DATA WAREHOUSE database (in Microsoft Access 2000 format).** You will be prompted to name the table and location as well as enter your username and password, as illustrated in **figure 002**



Click to browse your hard-drive and choose a directory location for the database

figure 002

3) Click on the browse button as illustrated above. Choose a file location and name.

4) Enter your **User Name** and **Password**, then click **Ok**.

5) You will see a status bar indicating that the database is being created. Click **Ok** when the dialogue box appears indicating that the database was successfully created.

6) Close the OUDW Utility program

7) Start Access from the **Start** menu and open the file you just created.

V. The Access/Data Warehouse Environment

When you create a database using the OUDW Utility program, you are building a local file that resides on your hard-drive. This file contains linked tables. These tables contain data that is linked to the data contained on the Data Warehouse server. In effect, these tables are mirror copies of data in the warehouse.

As data is updated on the server, it is automatically updated in the tables contained in your database. When you perform an action that uses data, like opening a table, query, or report, the table is refreshed so as to contain the most current data from the warehouse. The first time you perform such an operation upon opening the database, you will be prompted to enter your username and password.

You may interact with the data warehouse on a "read-only" basis. That is, you may view and extract information in the warehouse, but you may not change the data. If you try to perform an action which will modify the data, you will receive an error

message indicating that you cannot modify the data. An example of this error message appears below in **figure 004.**

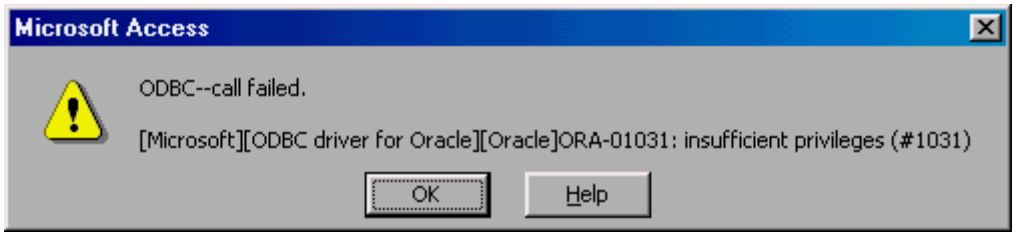


figure 004

Tables in your database are in effect images of data which is stored in the data warehouse. However, you will create objects like queries and reports which are stored only on your computer in the database file you create. Table Fundamentals

All of the data in the warehouse resides in tables. There are 25 tables in the database you created. Of those tables, 17 are look up tables and provide lists of items for particular fields. Each of these look up tables contain the letters XREF in the table name. For example, the XREF_DEGREE table contains full descriptions for degree programs referred to in other tables by degree code.

The primary table in the SIS Data Warehouse is the "PERSONAL_INFO" table. This table contains such demographic data including Name, Social Security Number, Marital Status, Date of Birth, and Handicap status. Other tables contain related information in a detailed format.

In order to extract information from the data warehouse you must have an understanding of where data is located. Selecting information involves querying one or more tables in the data warehouse. You must understand the structure of the data in order to build the correct queries. In effect, you need to ask the right questions in order to get the right answers.

Note for SIS Users

Many of the screens on the Student Information System (SIS) are really combinations of tables. They do not correspond directly to tables in the data warehouse. Rather, they are really queries that combine information from two or more tables. Any screen whose name starts with a "Z" is really a table. For example, the "ZPUB" screen in SIS has no corresponding table in the data warehouse, as it is a combination of data from the XXXXXX and XXXXXXXXX tables.

VI. Relationships

Relational databases, like Oracle and Access subdivide information into different tables. This significantly reduces data redundancy and enforces data integrity. For example, information on a student's address, phone should only appear in a table (or

list) one time. This information would not need to be repeated for every class scheduled in a list of classes taught. So demographic information is split into one table, and class schedule history into another.

The information is joined together, or related on one common field between the two tables. When using Access to query the data warehouse, you will have to manually establish relationships between two or more related tables.

The key table in the data warehouse is the SISWAREHOUSE.PERSONAL_INFO. This table contains demographic information on a student and is related to many other tables in the warehouse.

The PERSON_ID field is the primary key field for the SISWAREHOUSE.PERSONAL_INFO table. By joining that field to other tables which contain also contain the PERSON_ID field, the two tables can share information. The PERSON_ID field links to all other data tables (excepting XREF tables). It is the field which links these related tables together.

ESTABLISHING RELATIONSHIPS

1. Trace from the **Tools** menu to **Relationships.**

2. The Show Table Dialogue Box will appear as illustrated in **figure 003**

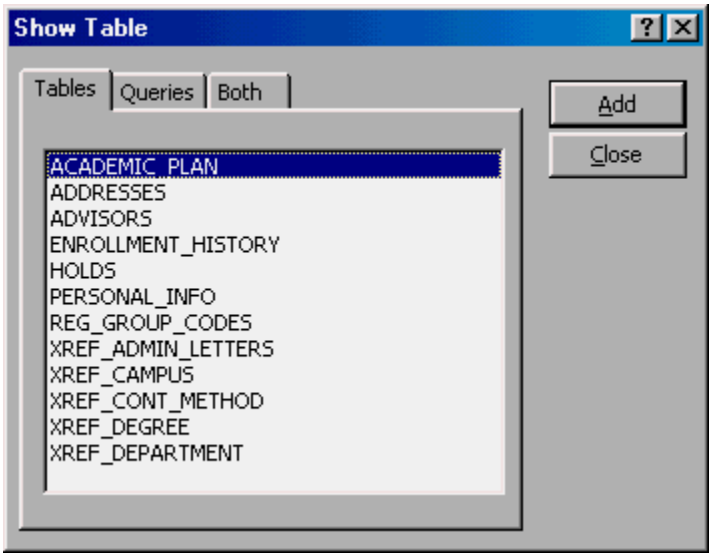


figure 003

3. Choose relevant tables by selecting them then clicking **Add.** When finished **Close** the Show Table Dialogue Box.

4. Note that a representation of each table you added appears in the Relationships window. All fields are listed for each table. You can expand the field lists for each table as necessary by positioning your cursor over a border to get a resizing arrow. Click and hold with the left mouse button to enlarge the table.

5. Access uses a primary key to make unique records in a table. The PERSON_ID field is the primary key for the PERSONAL_INFO table. It appears boldfaced in the field list. Because it is the primary key, no two records in the PERSONAL_INFO table will have the same PERSON_ID number. However, each individual listed in PERSONAL_INFO may have more than one address (home or local) in the ADDRESSES table. We will establish a relationship between the two tables so that they can share information.

6. Click and hold the left mouse button over the PERSON_ID field in the PERSONAL_INFO field list. Drag the mouse to cover the PERSON_ID field in the ADDRESSES table. A dialogue box will appear as illustrated in **figure 005**

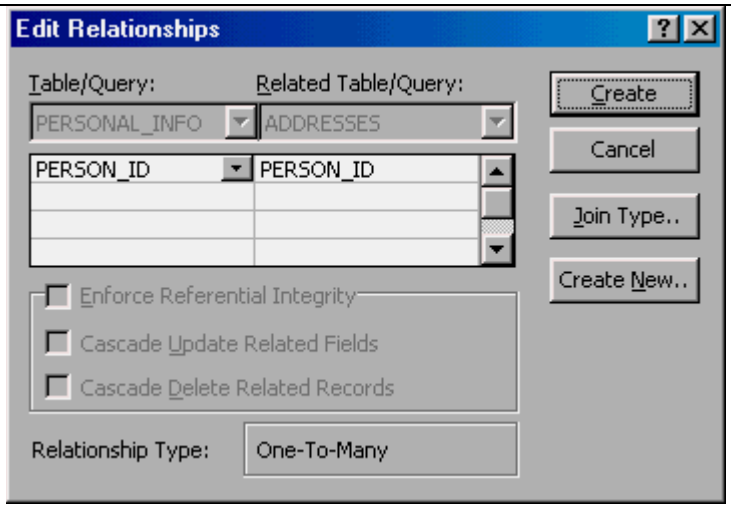


figure 005

7. Click with the left mouse button on **Create** to link the two tables on their related fields.

8. Close the relationships window.

VII. Querying the data warehouse

Queries allow you to extract certain records from a table, based on criteria that you specify. Queries create subsets of data, extracted from the entire set of the table. For example, the PERSONAL_INFO table contains a list of all students. A query would allow you to list only those students with residential state code of TN.

Because of the nature of relational databases, Access allow you to extract information from one or more tables, by building and running a query. Sometimes all of the information needed is located on one table. Often times information may reside in multiple related tables.

Running Single Table Queries

1. From the main Access Window, left-click the QUERIES button

2. Double-click with the left mouse button on the object titled "Create a Query In Design view. The "Show Table" Dialogue box appears as listed below in **figure 006**.

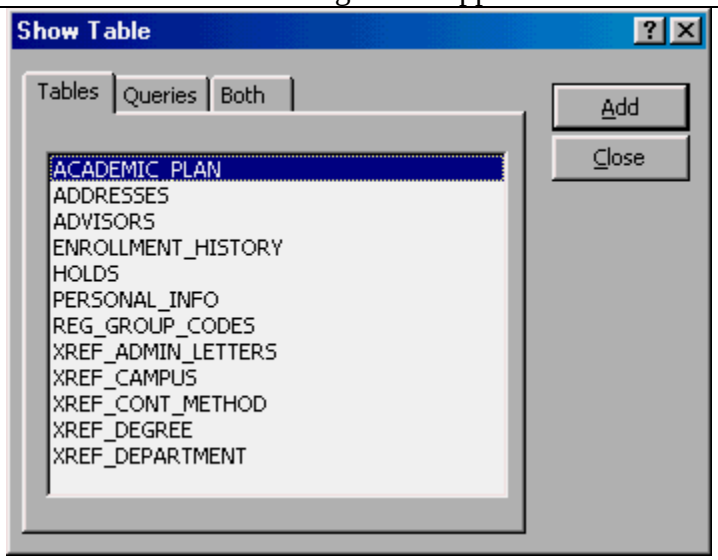


figure 006

1. Select the PERSONAL_INFO table, and click **Add**. (You may alternately, double-click with the left mouse button on PERSONAL_INFO). Then click the **Close** button.

2. The Field List appears as illustrated below in **figure 007**. Double-click on the following fields in the field list:

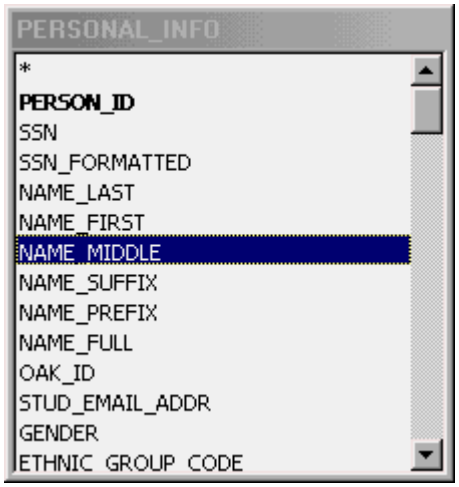


figure 007

NAME_LAST
NAME_FIRST
NAME_MIDDLE
GENDER
MARITAL_STATUS
RESI_STATE_ CODE

(you may also select fields for the query by clicking and dragging from the field list, or using the drop-down arrows in the query design grid)

3. You can run the query by clicking on the change view button, or the query run button as pictured in **figure 008**

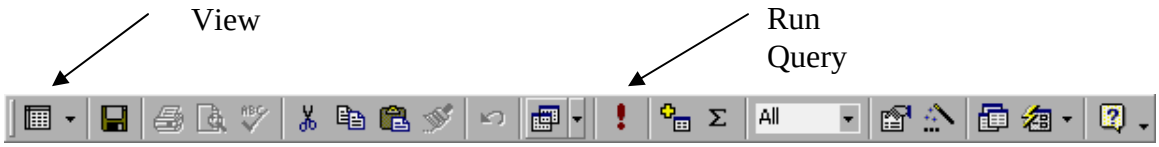


figure 008

At this point, the query will return approximately 26,000 records. The contents of this query (recordset) include the selected fields for every record in the table.
4. To refine our search, we need to specify criteria. Return to design view by clicking the view button.
5. Once in design view, type in the text TN in the criteria row in the Residency Code column. Now run the query again. The resulting recordset will only display those records which contain TN in the RESI_STATE_CODE field.

Often times, you need to extract records that meet more than one criteria. Access allows you to choose multiple criteria for records in a query. It is important to determine in the design of the query whether or not a query result should yield records that contain only one, or all of the multiple criteria. The distinction requires explaining the difference between AND criteria and OR criteria.

- AND – criteria written in a query on the same line specifies the AND criteria. The query recordset will contain only those records that meet all of the criteria statements..
- OR – criteria written on different lines will produce an OR statement in the query. The resulting recordset will contain records that meet one of the criteria selected.

Specifying Multiple Criteria

3. Return to the query design row. Enter the text M, in the same criteria row as TN, but under the MARITAL STATUS field.
4. Run the query. The resulting recordset will return values for those students who have a residency code of TN and are married. This demonstrates the AND property of the query.
5. Now, return to query design view. Delete the M from the AND line of the criteria design and insert it on the next line down. This will return an OR statement.
6. Return to Query Design view and change the Marital STATUS criteria back to the AND line. Then, modify the criteria for RESI_STATE_CODE to make it “TN or KY”. Run the query. When you’ve reviewed the query results, close and save the query as “SELECTED SOUTHERN STUDENTS”.

Running Multiple Table Queries

Since we really need more information than just student name, let's add another table to our query.

1. Open the SELECTED SOUTHERN STUDENTS QUERY in design view. Click on the SHOW TABLE icon as illustrated below in **figure 009**

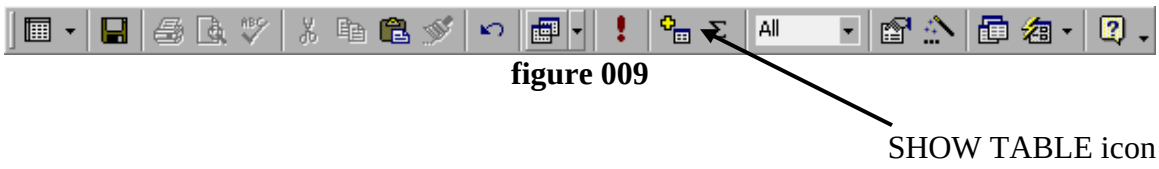


figure 009

(alternatively you can choose **Show Table** from the **View** Menu)

2. The Show Table Dialogue Box appears as it did when we first created a query in design view. Select the ADDRESSES table and click ADD. Then close the Show Table Dialogue box.

3. Note that the ADDRESSES table is joined to the PERSONAL_INFO table by a join line as shown in **figure 010**. This means that the two tables share a common field (PERSON_ID) and thusly share information between related records. We created the relationship between the PERSONAL_INFO and ADDRESSES table earlier in these exercises.

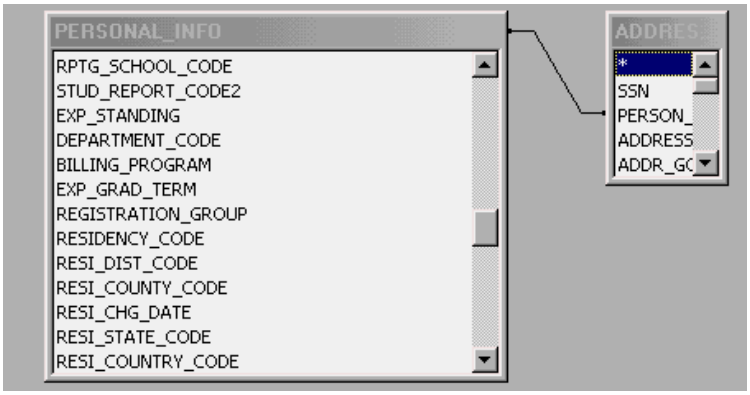


figure 010

4. Add the following fields to the query design grid:
   - ADDRESS_LINE1
   - ADDRESS_LINE2
   - CITY_NAME
   - STATE_NAME
   - ZIP_CODE
   - ADDRESS_CODE

5. Type "LO" in the criteria line for ADDRESS_CODE. This will return only local addresses for these students. Run the query and view the recordset.

Join Types

The join type defines the relationship between two tables, and how the related fields are connected. When a multiple table query is run and a Recordset is generated, the results of the query depend on the type of join between the tables. Defining the join types allows you to manipulate the data in the related tables to obtain the information that you want.

There are two join types in Access: inner join, and outer join.

With an inner join, information is selected in a query from related tables only if the information in the joined field is the same in both tables. This is the default setting for Access. An outer join will allow records to be listed even if they don't have a related record in the corresponding table.

For example, a query extracting data from the PERSONAL_INFO and ADDRESSES table will only return a record for a student who has a related record in the ADDRESS table. A student without an address will not be included in the query recordset if the tables are related with an inner join.

It is important to keep the relationships in mind when you are extracting information. You may need to see certain records, even if they don't have a related record in another table. For example, you may need to know if a student meets certain criteria, even if they don't have an address in the address table. You can change the relationships in a query to return different results.

To Change Join Types

1. View the query in Design View.
2. **2.** Double click with the left mouse button on the line joining the PERSONAL_INFO and ADDRESSES tables. A dialogue box will appear as illustrated in **figure 011**

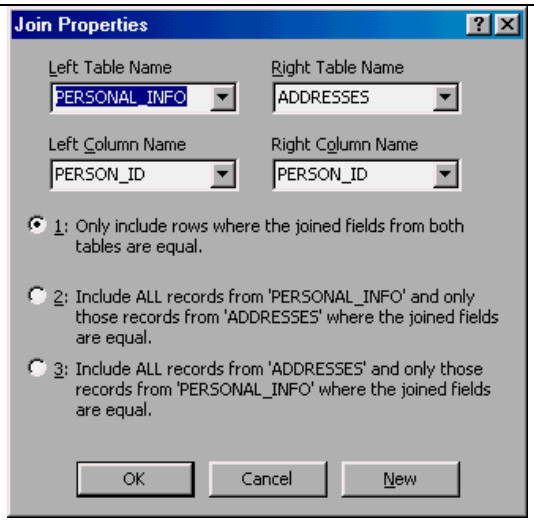


figure 011

3. Choose the second option "Include ALL records from PERSONAL_INFO…". and click **OK**. This creates an "outer-join".

4. Re-run the query and view the recordset. Previously, records with no address did not appear. Now that there is an "outer-join" established between the two tables, student records will appear even if they do not contain a matching address.

VIII. Creating Labels

Access provides a relatively easy way to take the results of a table or query and print them in a label format. For example, we may want to create mailing labels to mail a brochure to those students selected from our previous query example.

To Run the Label Wizard

1. Select the **Reports** section from the main Access window.

2. Left-click the **New** icon at the top of the screen

3. Choose **Label Wizard** as the type of report, and select **state marital name** from the drop down list of tables and queries

4. From the first screen in the label wizard, change the unit of measurement to "**English**" and the label manufacturer to "**Avery**" as illustrated in **figure 012**. For our purposes choose Avery Product number 5160

Change to English

Change to Avery

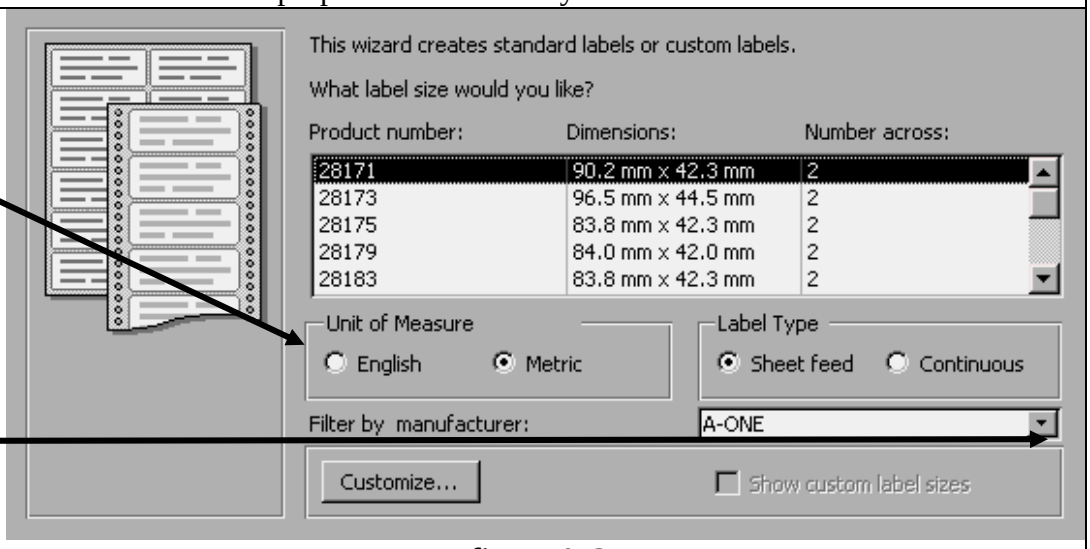


figure 012

5. Click the **Next** button

6. Select the Font, Size, Style and Color for the entire label. Note that larger font sizes restrict the amount of information that may be contained on a label Then left-click on the **Next** button.

7. On the wizard screen, select the fields you want to include on the label by double-clicking on each field. Between fields, insert the proper punctuation. When you have selected the appropriate fields for a line on the label, press the **Enter** key to move to the next line. Repeat this process for all information on the label, then left-click on the **Next** button.

8. Choose the field on which to sort the label output.

9. Left-click on the **Next** button, then on the **Finish** button

10. The finished label report will appear as illustrated in **figure 013.** Note that there are spacing errors in the report.

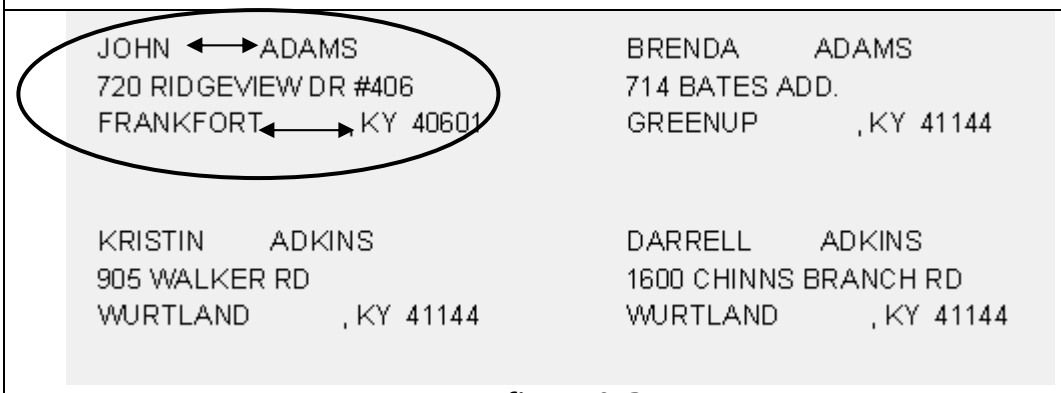


figure 013

11. To correct for the spacing errors, switch to **Design View**

12. Left-click the object for the first line of the label to select. Then display the **Properties** window. Choose the **Data** tab from the properties window and then left-click to select the **Control Source** section. While holding the "**SHIFT"** key, press the **"F2"** key to view the ZOOM window

13. Edit the code in the Zoom window so that it appears as outlined below (note that changes are boldfaced.
=Trim([NAME_FIRST]**)**& " "&**Trim(**[NAME_LAST])

14. Select and Edit the final line in the label in the same manner, replacing the code with the code below

=Trim([CITY_NAME]**)** & ", " &**Trim(**[STATE_CODE])& " "
&**Trim(**[ZIP_CODE])

The label report is now ready to be printed.

IX. Mail Merge

You can use the results of an Access query to merge with a Microsoft Word document to create individualized form letters.

Creating a Mail Merge

1. Open Microsoft Word

2. Trace from the **Tools** menu to **Mail Merge.** A dialogue box will appear as illustrated in **figure 014**

3. Click with the left-mouse button to select **Create** and activate the drop-down list of document types. Choose **Form Letters…** from the list, then click with the left-mouse button to select **Active Document**. This will select the active document as the document to be merged with your data.

4. Now choose the data by left-clicking on **Get Data** dropdown list under the **Data Source** heading. From the drop-down list, select **Open Data Source…**

5. A **Open Data Source** dialogue box will appear (similar to a Open File box). Be sure to change the **Files of type** to MS Access Database. Then navigate to select the Access Database linked to the Data Warehouse. Select the query you would like to merge after clicking on the **Queries** tab. Make sure that the **Link to Query** box is checked, then left-click on **OK**. After a brief wait, choose **Edit Main Document.**

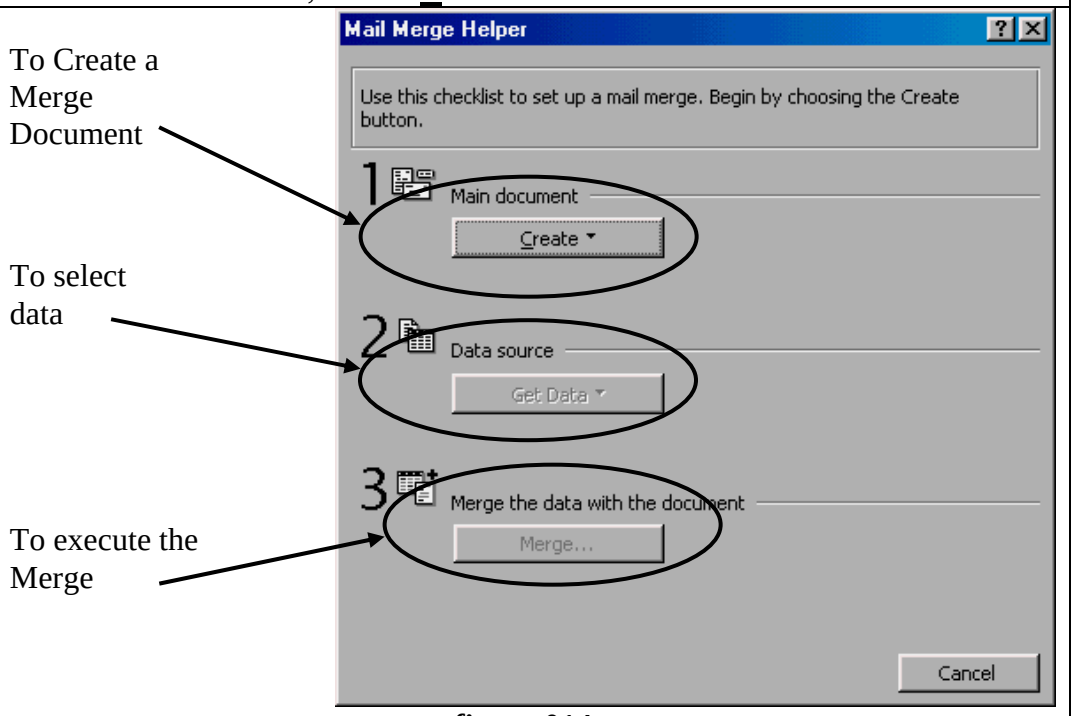


figure 014

6. Note that the Word Tool Bar has changed, giving you the ability to select data fields from the **Insert Merged Field** drop-down list as highlighted in **figure 015**

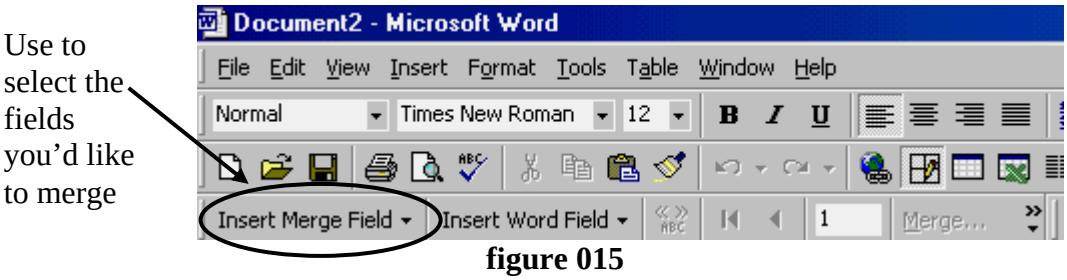


figure 015

7. Select the appropriate fields from the **Insert Merged Field** drop-down list to insert them in your document. Be sure to include appropriate punctuation, such as a space between first and last names, or a comma between City and State. Use the Enter key to move to a new line.

8. Type in the body of the letter as you would a normal Word document. When finished, trace to **Mail Merge** from the **Tools** menu. Then select **Merge**. From the resulting dialogue box, accept the default options and left-click **Merge** to execute the mail merge.

X. Exporting E-mail Addresses

You can query the data warehouse to find e-mail addresses for students, then export those e-mails to a text file. The text file can in turn be cut and pasted into the TO or BCC fields in your e-mail client.

To Export E-mail Addresses

1. Open the already created query **"AcctingEmail"** in design view. Note that this query extracts currently enrolled Accounting majors.

2. To extract e-mail addresses and include a comma to separate addresses, create a new column in the query with the following information:

   email: [STUD_EMAIL_ADDR] & ","

3. Close and save the query.

4. We now need to create a new query to extract only the email column for our output. From the query tab, double-click with the left-mouse button on "**Create a Query in Design View".**

5. From the Show Table dialogue box, choose the "**Queries**" tab. Select the **AcctingEmail** query and left-click **Add**. Now close the Show Table dialogue box by left-clicking **Close**.

6. Scroll through the field list to view the newly created field "**email**". Double-click with the left-mouse button to select that field as the only field in the new query. Close the query and save it under the name "**emailexport**"

7. "Right-click" on the "**email**" field and select **Properties** from the short-cut menu. Type "<" in the format section, located in the **General** tab. This will force the formatting of the e-mail addresses to be in lower-case.

8. From the Access database window. Click on the "**emailexport**" query with the right-mouse button. From the menu, select **Export**. A file save dialogue box will appear as illustrated in **figure 016.**

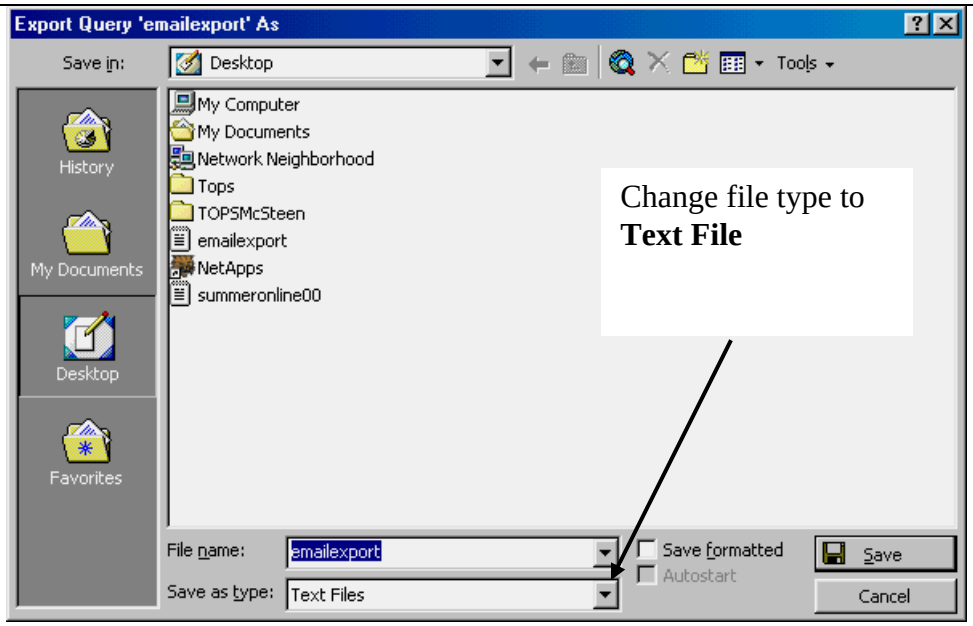


figure 016

9. Choose the filename and location. Be sure to change the **Save as type** selection to **Text Files** using the drop-down selection list. Then left-click on **Save**. Access will open the **Export Wizard** as illustrated in **figure 017**

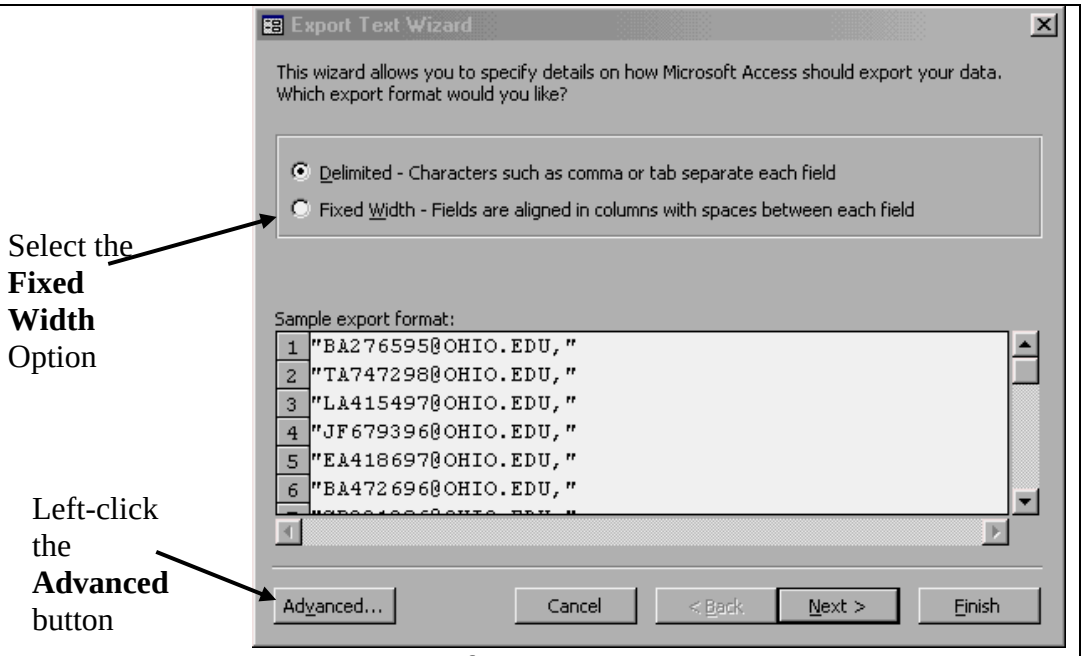


figure 017

10. As illustrated in **figure 017** above, select the **Fixed Width** option and then left-click on the **Advanced** button. The **Export Specification** dialogue box will appear as illustrated in **figure 018.**

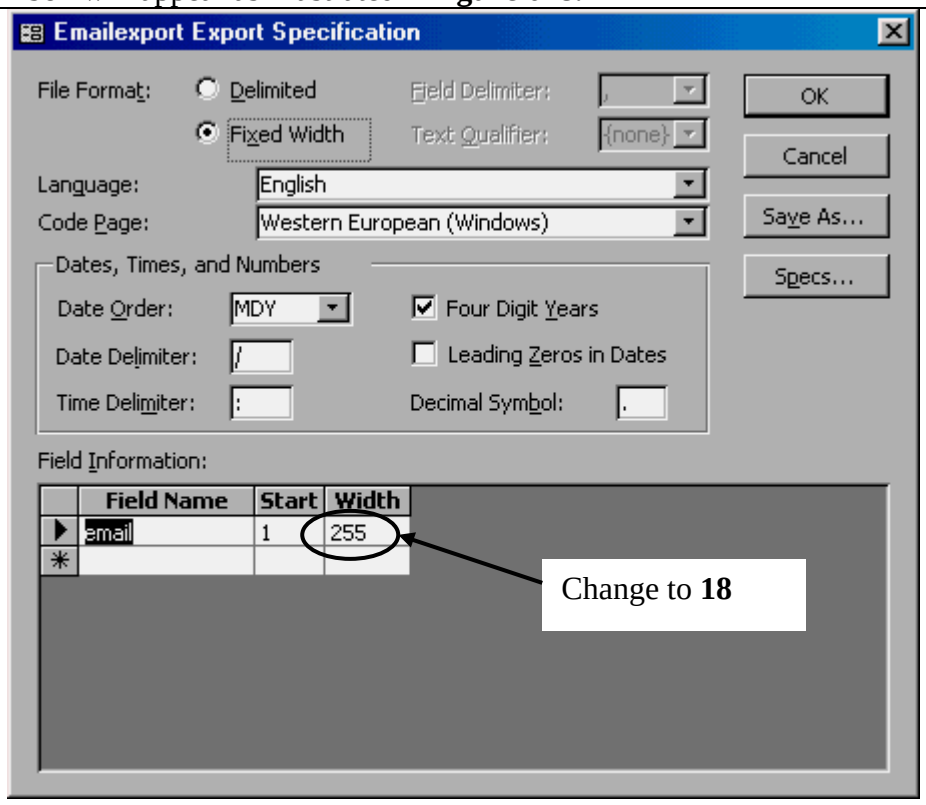


figure 018

11. Change the width of the email field to **18**. This will prevent extra spaces from being exported into your resulting text file. Now left-click on the **OK** button.

12. After you return (automatically) to the **Export Text Wizard** left-click the **Next** button twice in turn, then the **Finish** button. The query will be exported into a text file. You will receive a confirmation box when the export is finished.

13. Open the text file with the **NotePad** or similar application. From the **Edit** menu, left-click on **Select All**. Then select **Copy** from the **Edit** menu. (Alternatively, you can copy the records by pressing the **Control** key in combination with the **C** key). Close the **NotePad** application.

14. Open your e-mail client. Paste the copied records into the **TO BCC** fields of a new email message. You can do this by choosing **Paste** from the **Edit** menu or pressing the **Control** and **V** keys in combination.

X. Charting Query Results

Microsoft Access contains a charting utility, known as the Chart Wizard. This tool can be used to display query results in a graphical format.

To Chart Query Results

1. From the Access Database window. Choose the **Reports** section. Left-click on the **New** icon at the top of the window.

2. As illustrated in **figure 019** below, select Chart Wizard. Choose the COB majors from the drop-down list of tables and queries

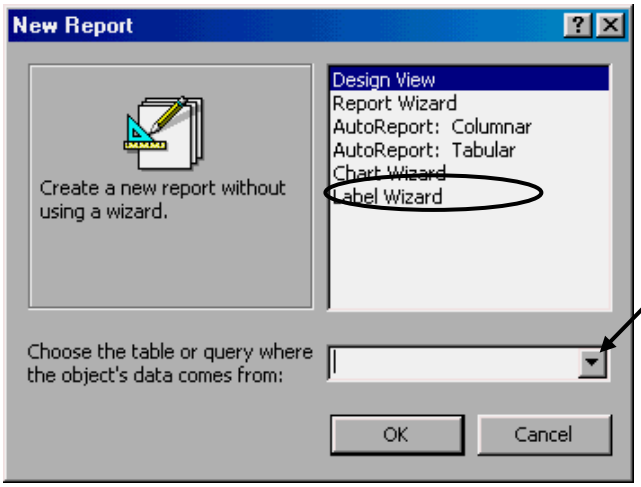


figure 019

3. The **Chart Wizard** will launch. From the first screen in the **Chart Wizard**, select the fields to be included in the chart by double-clicking with the left-mouse button. Alternatively, you can select the fields then press the right arrow button on the wizard. For our example, select **PROGRAM_NAME** and **NAME_FULL**. Left-click the **Next** button.

4. The next screen in the chart wizard allows us to select the chart type. For our example, select the Column chart located in the first column of the first row of char examples. Then left-click the **Next** button.

5. As illustrated in **figure 020** left-click and hold the left-mouse button on icon representing the **NAME_FULL** field and drag to the empty **Data** cell. Note that the **PROGRAM_NAME** field is already assigned to the X axis. Note that the **NAME_FULL** data changes to "CountofNAME_FULL"

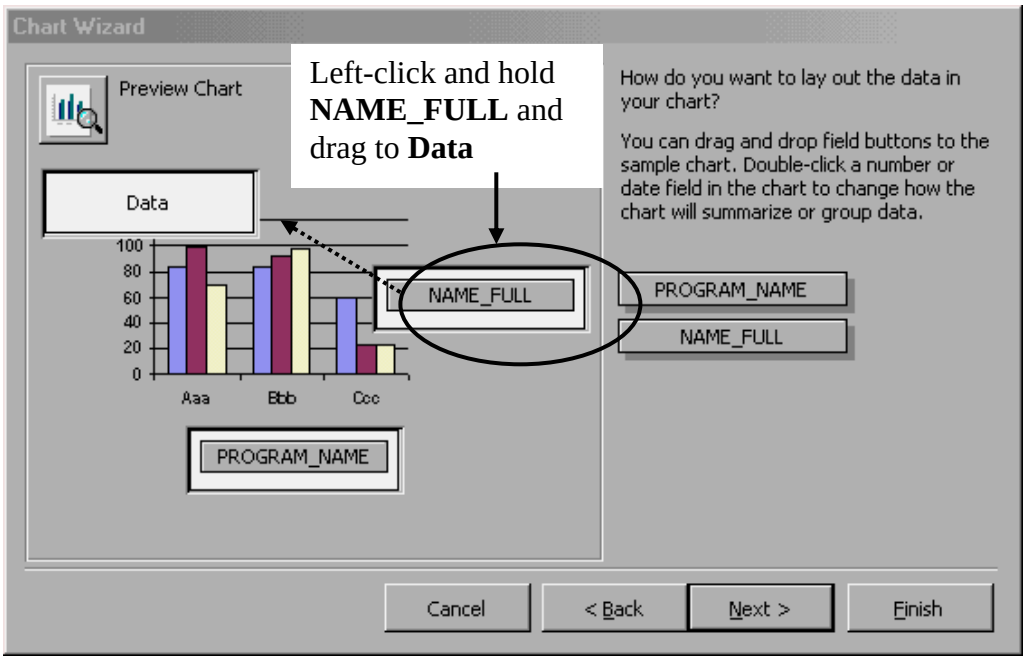


figure 020

6. Left-click on the **Next** button.

7. Select a name for the Chart and left-click on **Finish**. Access will process the chart and display the results, as displayed in **figure 021** below.

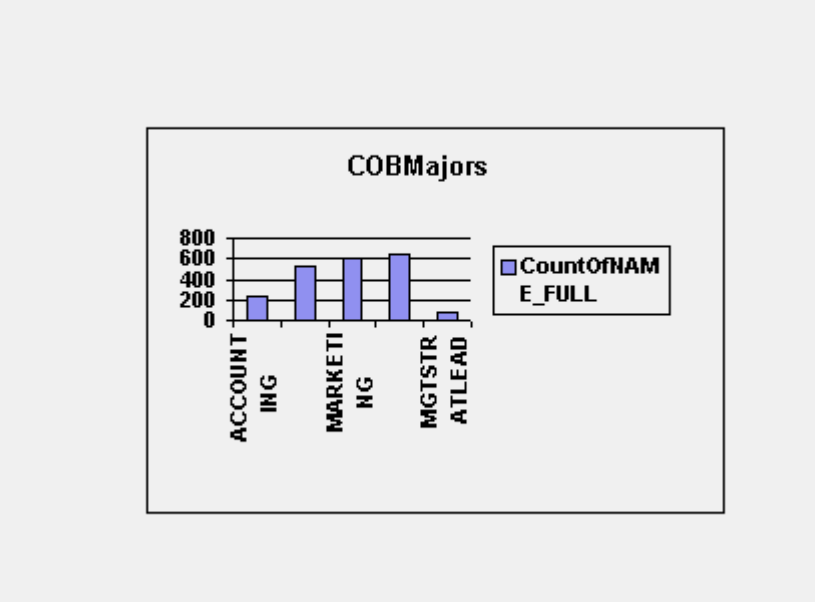


figure 021

Modifying Chart Elements.

As you will note from the finished chart, Access may not format charts in size and format that is ideal. You may need to refine chart elements in the report's design view.

To Modify The Existing Chart

1. From the Print Preview mode of the chart, switch to Design View.

2. Double-click with the left-mouse button to select the entire chart in the Detail section of the chart's design view. The view in design view will appear as below in **figure 022**

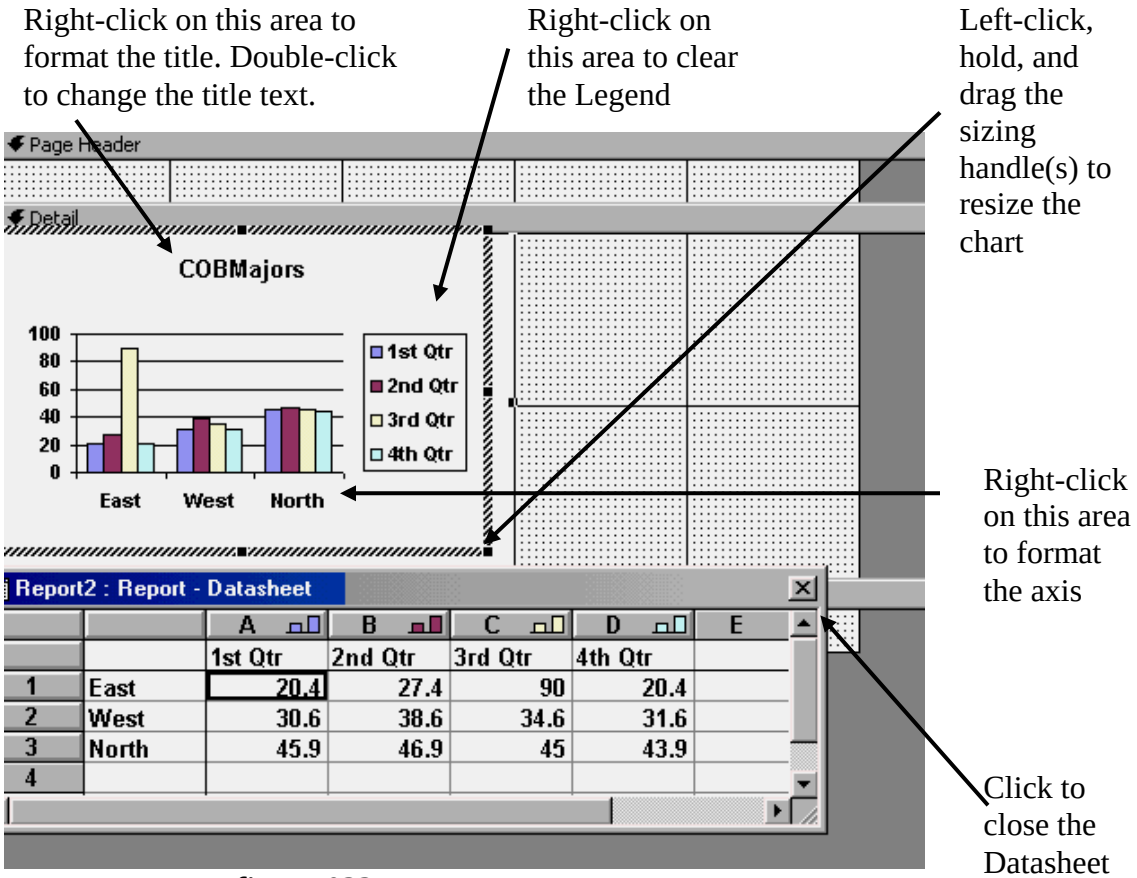


figure 022

3. Left-click on the **X** icon to close the Report's Datasheet

4. Use the sizing handle on the lower-left corner of the chart to resize the chart. This will not only change the size of the finished chart, but should also allow more room so that the X axis titles display completely.

5. Right-click on the legend area. Select **Clear** from the menu to remove the legend.

6. Double-click on the chart title to switch to editing mode. This will allow you to change the text that appears as the title of the chart. Note, this will not change the Object Name as it appears in the database window.

7. Right-click on the chart X axis. Choose **Format Axis…** from the menu. Now you can change the Font, Alignment, and other attributes of the axis.

8. When finished editing, switch to Print Preview view to view your results.