11 of 49

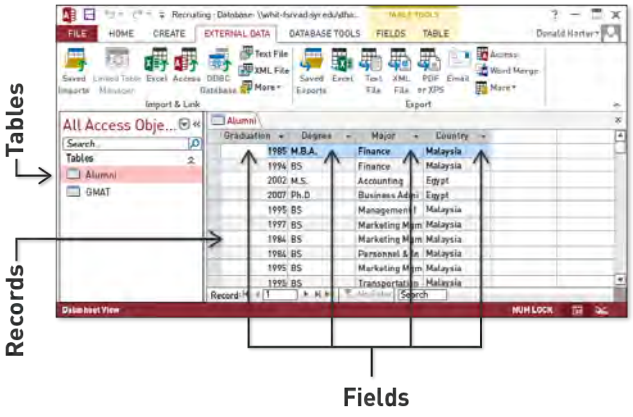
What Is a Database?

- Collecon of gr oups of data similar to an Excel workbook
- Different terminology because databases were created decades before spreadsheets

Excel	Database
Spreadsheet	Table
Column	Field
Row	Record

12 of 49

Database Example



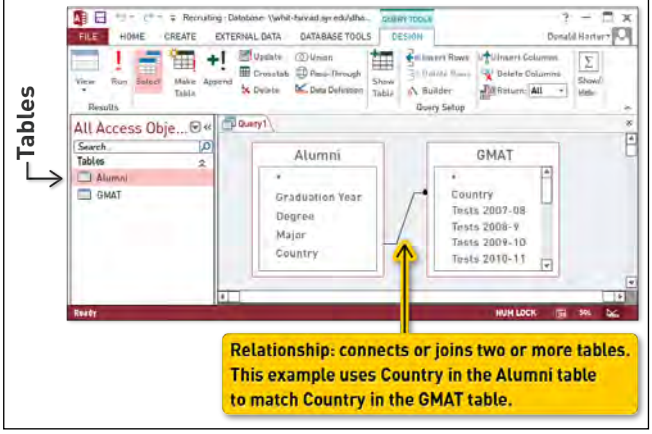
25 of 49

Differences between Databases and Spreadsheets

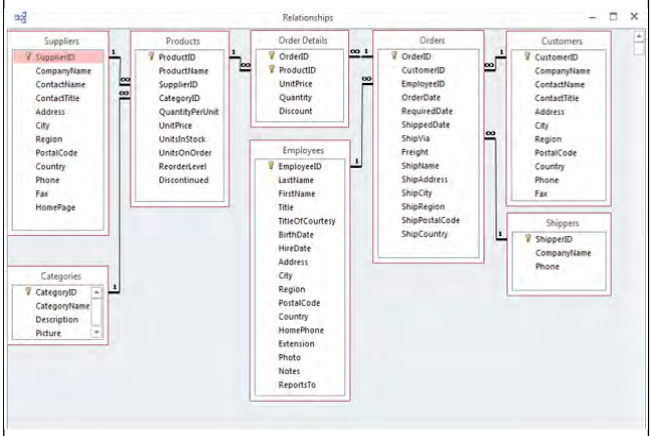
- Spreadsheet workbooks allow you to view one sheet at a me.
  - Formulas can cross spreadsheets, but difficult to combine data from mulple shee ts
- Databases allow you to create relaonship s between tables and view "joined" tables.

32 of 49

Joined Tables



33 of 49

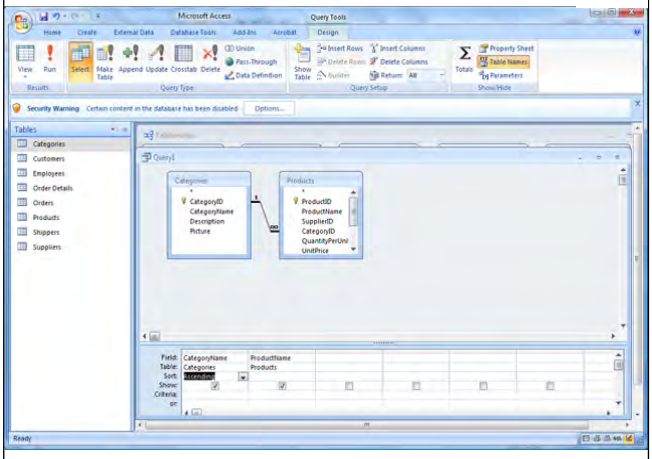


38 of 49

Database Relationships

- 1-to-1: Each row (record) in one table matches one record (row) in the second.
  - Example: Table of high school students and table of high school football players
- 1-to-many: Each record in the first table links to many records in the second.
  - Example: Couple and dependents on income tax form

39 of 49



45 of 49

Queries

- Retrieve data from database
- Can span mulple t ables, using relaonship t o join them
- Can retrieve raw data, group and sort data, perform calculaons, and filt er data
- Look like a table
- Can be saved and used in other queries (subqueries)

Hands-On Activity

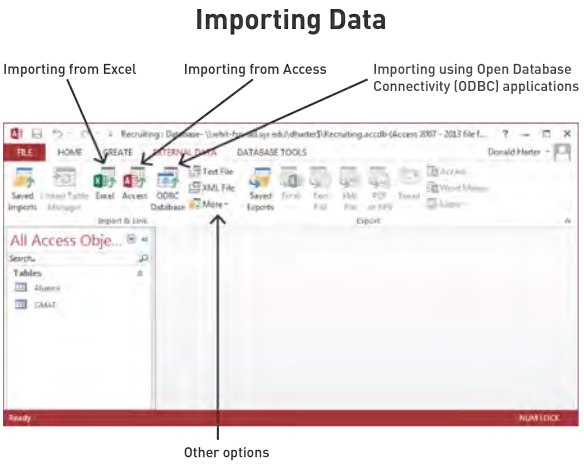
- Enter data.
- Connect data with relaonship s.
- Perform queries.

View All View Keyframes

Accessing and Importing Data

- Import from other files or systems.
  - Excel, Microsoft Access, text files, XML data, etc.
- Databases can read and import data stored in other systems.

4 of 13



5 of 13

Hands-On Activity

- Import data from various sources.
- Begin building first database.

13 of 13

View All View Keyframes

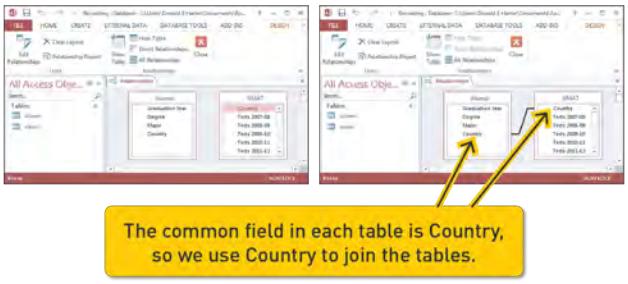
Relationships

- Determine how data is used when retrieving data from mulple t ables
- Do not typically exist when data tables are imported into database
- Must be manually idenfied/ created
- Created by finding common field(s) between two tables

5 of 13

Relationships

Before Joining After Joining



11 of 13

Hands-On Activity

Manually create joins and relaonship s in Access.

13 of 13

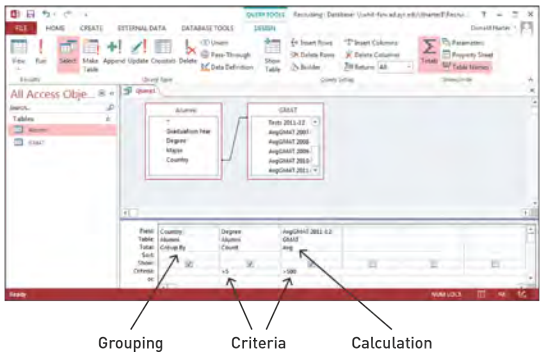
View All View Keyframes

Simple Queries

- Simple queries allow retrieval of data.
  - By idenf ying fields (columns) in tables to be retrieved
- More complex queries group data and perform calculaons.
  - E.g., number of alumni from a country or average GMAT for country
- Criteria allow filtering of data.
  - E.g., limit retrievals to countries with more than five alumni or average GMAT score more than 500

7 of 15

Simple Query Example



8 of 15

Function of a Simple Query

- Build on simple data by grouping, calculang , and using criteria to filter.
- Pracce in hands-on acvity .

15 of 15

View All View Keyframes

Dirty Data

- May be due to misspelled or mistyped informaon in da tabase
  - Difficult to correct, but new matching technology can make recommendaons
- May result when two or more parts of database use different idenfier s for same item
  - E.g., "New York City" and "NYC"
  - Can be idenfied using an op on in r elaonship joins

6 of 39

Types of Joins

- **Inner join:** only includes records from the two tables when the joining fields from each table match exactly
- **Left join:** includes all records from left table and only records from right table where there is an exact match
  - Can result in voids
- **Right join:** includes all records from right table and only records from left table where there is an exact match
- Left and right joins give broader spectrum of data than inner join.

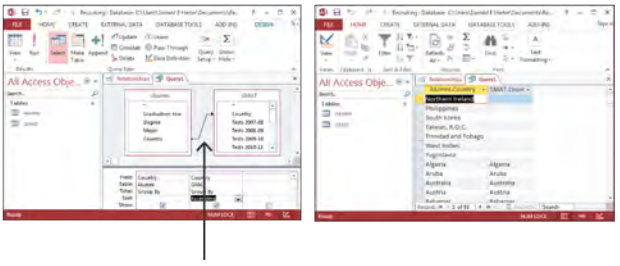
15 of 39

Identifying Relationships with Dirty Data

- Run query using left join.
  - Look for records with missing data that should have been retrieved from right side.
- Run query using right join.
  - Look for records with missing data that should have been retrieved from left side.
- Correct mismatches due to dirty data.

21 of 39

Dirty Data Example: Left Join



Left join: Direction of arrow

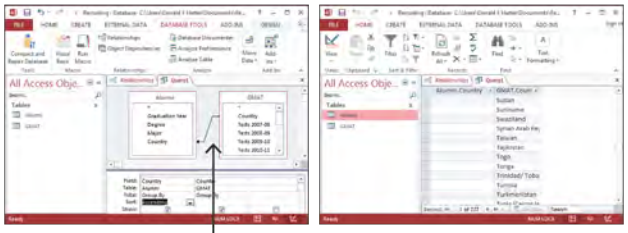
22 of 39

Using Left Join

- Selects all data from le t able and only exact matches from right table
- GMAT database uses "Taiwan," not "Taiwan, R.O.C.," so no exact match in alumni database
- Picks up data from le t able and finds what is missing from right table

31 of 39

Dirty Data Example: Right Join



Right join: Direction of arrow

32 of 39

Fixing Dirty Data

- Relaonship s can idenf y inconsistencies in data.
- Correcng da ta in tables gives a more accurate representaon in queries.

39 of 39

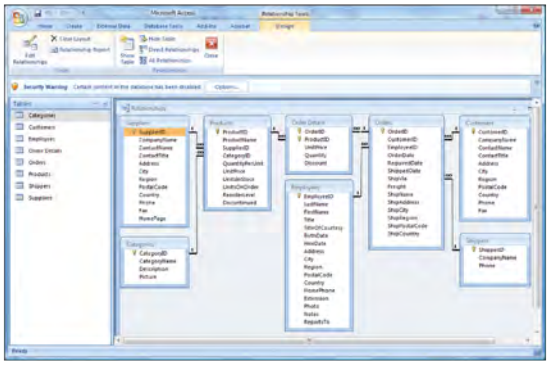
View All View Keyframes

Complex Queries

- Can retrieve data from larger numbers of tables and enr e databases
- Must have correctly defined relaonship s and dirty data corrected
- Incorporate filters to focus retrieval

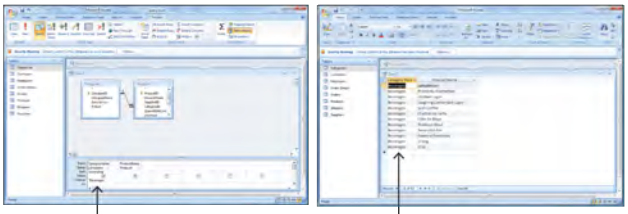
4 of 14

Complex Database



5 of 14

Complex Query with Filters



7 of 14

Using Complex Queries

- Can expand to mulple t ables, even thousands
- Need team to keep track of data and relaonship s
- Not much more difficult than simple queries

14 of 14