

15.561
Information Technology Essentials

Session 6

Relational Databases

Outline

- **What is a database?**
- **What is a database management system?**
- **An Introduction to Microsoft Access**
 - *How to create a database*
 - *How to retrieve data from a database*
 - *How to build a nice Graphical User Interface on top of a database*

Why are we learning this?

- Databases are perhaps the single most important class of corporate applications
- Databases are surprisingly powerful data modeling and analysis tools in situations where spreadsheets fall short
 - Students who plan to work in management consulting will soon find this out
- MS Access is a great example of how easy it is to build powerful applications without the need of a background in technology

What is a database

- **Boring answer**
 - A structured collection of data
 - Example: A telephone directory

- **Insightful answer**
 - A data-centered mirror of an organization's business processes

 - Structure of data reflects organizational processes
 - Content of data reflects organization's history

Example: Northwind Traders

Representing the Real World as Data

What Data Are Businesses Interested In?

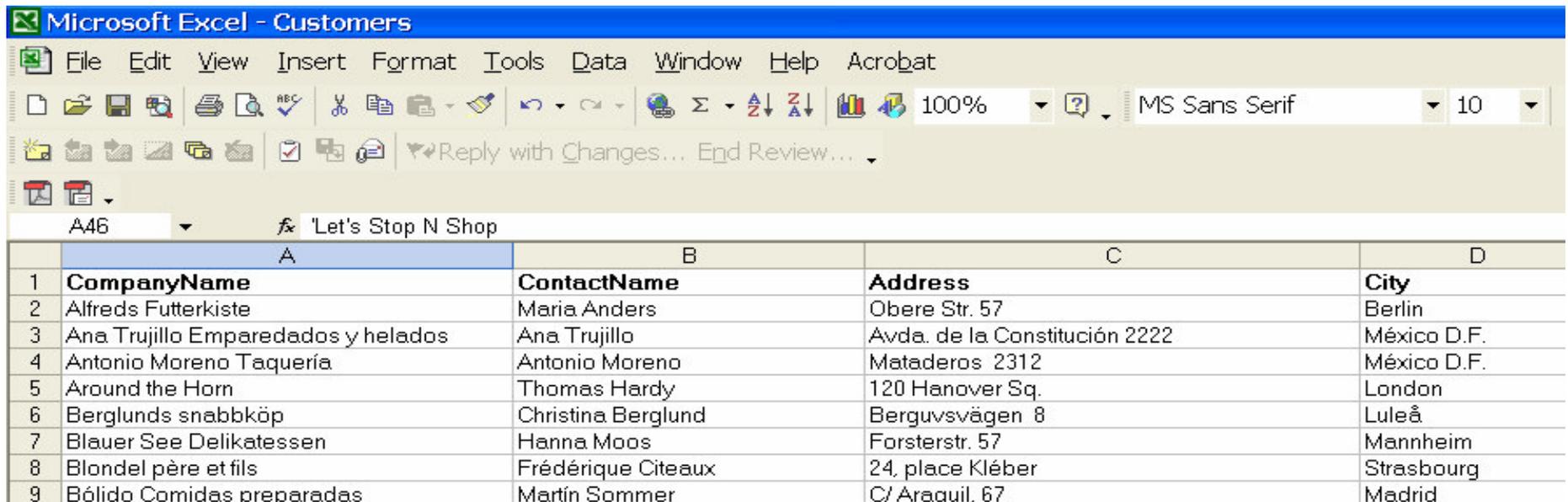
- **Entity**
 - a person, place, thing, or event on which we maintain information
 - Examples: Employees, Customers, Products, Warehouses
- **Attribute**
 - characteristic or quality of particular entity
 - Examples: Employee's SSN, Customer's Address, Product's Unit Price
- **Relationships Among Entities**
 - Examples:
 - » Customer - *Orders* - Product(s)
 - » Order - *Serviced by* - Employee

From Spreadsheets to Databases

- Spreadsheets are great for keeping track of data for one type of entities
 - Participants of a conference
 - Students of a class
 - Customers of a company
 -

What is the basic spreadsheet “data model”?

- *Each row stores data about one entity*
- *Each column stores data about an attribute*
- *Each cell stores data about an attribute of an entity*



The screenshot shows a Microsoft Excel spreadsheet titled "Customers". The spreadsheet has four columns: CompanyName, ContactName, Address, and City. The data is as follows:

	A	B	C	D
1	CompanyName	ContactName	Address	City
2	Alfreds Futterkiste	Maria Anders	Obere Str. 57	Berlin
3	Ana Trujillo Emparedados y helados	Ana Trujillo	Avda. de la Constitución 2222	México D.F.
4	Antonio Moreno Taquería	Antonio Moreno	Mataderos 2312	México D.F.
5	Around the Horn	Thomas Hardy	120 Hanover Sq.	London
6	Berglunds snabbköp	Christina Berglund	Berguvsvägen 8	Luleå
7	Blauer See Delikatessen	Hanna Moos	Forsterstr. 57	Mannheim
8	Blondel père et fils	Frédérique Citeaux	24, place Kléber	Strasbourg
9	Bólido Comidas preparadas	Martín Sommer	C/ Araquil, 67	Madrid

Spreadsheet limitations

- Things get complicated when we want to keep track of several inter-related entities
- For example:
 - Customers
 - Products
 - Orders
- Let's try it!

Spreadsheets are awkward for storing relationships

- Main difficulty is that an “Order” is essentially a *relationship* between one Customer and one or more Products

Storage of information is not even half the story

- The reason we build databases is in order to easily retrieve information to answer questions that support managerial decision-making

- For example:

Who are our top 10 customers based on their total order value in the year 2002?

- Can you do this using a spreadsheet?

Enter Relational Databases

- **A relational DB supports storage of data as a set of inter-related tables**
 - Each table stores data about a set of Entities
 - Each table row is a record about one such Entity
 - Each record column is a field specifying an attribute of this Entity
 - Each record has a field that acts as a unique identifier of an entity
 - Relationships among entities are specified by referring to this unique identifier from other tables

Customer Unique Id

Customers : Table

Customer ID	Company Name	Contact Name	Contact Title	Address	City	Region	Postal Code	Country
ALFKI	Alfreds Futterkiste	Maria Anders	Sales Representative	Obere Str. 57	Berlin		12209	Germany
ANATR	Ana Trujillo Emparedados y helados	Ana Trujillo	Owner	Avda. de la Constitución 2222	México D.F.		05021	Mexico
ANTON	Antonio Moreno Taquería			Mataderos 2312	México D.F.		05023	Mexico
AROUT	Around the Horn		Representative	120 Hanover Sq.	London		WA1 1DP	UK
BERGS	Berglunds snabbköp		Administrator	Berguvsvägen 8	Luleå		S-958 22	Sweden
BLAUS	Blauer See Delikatessen	Hanna Moos	Sales Representative	Forsterstr. 57	Mannheim		68306	Germany

Product Unique Id

Products : Table

Product ID	Product Name	Supplier	Category	Quantity Per Unit	Unit Price	Units In Stock	Units On Order	Reorder
1	Chai	Exotic Liquids	Beverages	10 boxes x 20 bags	\$18.00	39	0	
2	Chang	Exotic Liquids	Beverages	24 - 12 oz bottles	\$19.00	17	40	
3	Aniseed Syrup	Exotic Liquids	Condiments	12 - 550 ml bottles	\$10.00	13	70	
4	Chef Anton's Cajun Seasoning	New Orleans Cajun Delights	Condiments	48 - 6 oz jars	\$22.00	53	0	
5	Chef Anton's Gumbo Mix	New Orleans Cajun Delights	Condiments	36 boxes	\$21.35	0	0	
6	Grandma's Boysenberry Spread	Grandma Kelly's Homestead	Condiments	12 - 8 oz jars	\$25.00	120	0	
7	Uncle Bob's Organic Dried Pears		Produce	12 - 1 lb pkgs.	\$30.00	15	0	
8	Northwoods Cranberry Sauce		Condiments	12 - 12 oz jars	\$40.00	6	0	
9	Mishi Kobe Niku		Meat/Poultry	18 - 500 g pkgs.	\$97.00	29	0	
10	Ikura	Tokyo Traders	Seafood	12 - 200 ml jars	\$31.00	31	0	
11	Queso Cabotijos		Dairy Products	4 - 100 g pkgs.	\$21.00	20	0	
12	Queso Pasa		Dairy Products	4 - 100 g pkgs.	\$21.00	20	0	

Order Unique Id

Orders : Table

Order ID	Customer	Employee	Order Date	Required Date	Shipped Date	Ship Via	Freight	Ship
10248	VINET	Buchanan, Steven	04-Aug-94	01-Sep-94	16-Aug-94	Federal Shipping	\$32.38	Vins et alcools
10249	TOMSP	Suyama, Michael	05-Aug-94	16-Sep-94	10-Aug-94	Speedy Express	\$11.61	Toms Spezialita
10250	HANAR	Peacock, Margaret	08-Aug-94	05-Sep-94	12-Aug-94	United Package	\$65.83	Hanari Carnes
10251	VICTE	Leverling, Janet	08-Aug-94	05-Sep-94	15-Aug-94	Speedy Express	\$41.34	Victuailles en s
10252	SUPRD		08-Aug-94	06-Sep-94	11-Aug-94	United Package	\$51.30	Suprêmes délic
10253	HANAR		08-Aug-94	24-Aug-94	16-Aug-94	United Package	\$58.17	Hanari Carnes
10254	CHOPS		08-Aug-94	08-Sep-94	23-Aug-94	United Package	\$22.98	Chop-suey Chin
10255	RICSU	Dodsworth, Anne	12-Aug-94	09-Sep-94	15-Aug-94	Federal Shipping	\$148.33	Richter Superm
10256	WELLI	Leverling, Janet	15-Aug-94	12-Sep-94	17-Aug-94	United Package	\$13.97	Wellington Impo
10257	HILAA	Peacock, Margaret	16-Aug-94	13-Sep-94	22-Aug-94	Federal Shipping	\$81.91	HILARIÓN-Abas
10258	ERNSH	Davolio, Nancy	17-Aug-94	14-Sep-94	23-Aug-94	Speedy Express	\$140.51	Ernst Handel
10259	CENTC	Peacock, Margaret	18-Aug-94	15-Sep-94	25-Aug-94	Federal Shipping	\$3.25	Centro comerci
10260	OTTIK	Peacock, Margaret	19-Aug-94	16-Sep-94	29-Aug-94	Speedy Express	\$55.09	Ottlies Käselac
10261	QUEDE	Peacock, Margaret	19-Aug-94	16-Sep-94	30-Aug-94	United Package	\$3.05	Que Delicia
10262	RATTC	Callahan, Laura	22-Aug-94	19-Sep-94	25-Aug-94	Federal Shipping	\$48.29	Rattlesnake Ca
10263	ERNSH	Dodsworth, Anne	23-Aug-94	20-Sep-94	31-Aug-94	Federal Shipping	\$146.06	Ernst Handel
10264	FOLKO	Suyama, Michael	24-Aug-94	21-Sep-94	23-Sep-94	Federal Shipping	\$3.67	Folk och få HB
10265	BLONP	Fuller, Andrew	25-Aug-94	22-Sep-94	12-Sep-94	Speedy Express	\$55.28	Blondel père et
10266	WARTH	Leverling, Janet	26-Aug-94	07-Oct-94	31-Aug-94	Federal Shipping	\$25.73	Wartian Herkku
10267	FRANK	Peacock, Margaret	29-Aug-94	26-Sep-94	06-Sep-94	Speedy Express	\$208.58	Frankenversand
10268	GROSR	Callahan, Laura	30-Aug-94	27-Sep-94	02-Sep-94	Federal Shipping	\$66.29	GROSELLA-Re
10269	WHITC	Buchanan, Steven	31-Aug-94	14-Sep-94	09-Sep-94	Speedy Express	\$4.56	White Clover M
10270	WARTH	Davolio, Nancy	01-Sep-94	29-Sep-94	02-Sep-94	Speedy Express	\$136.54	Wartian Herkku

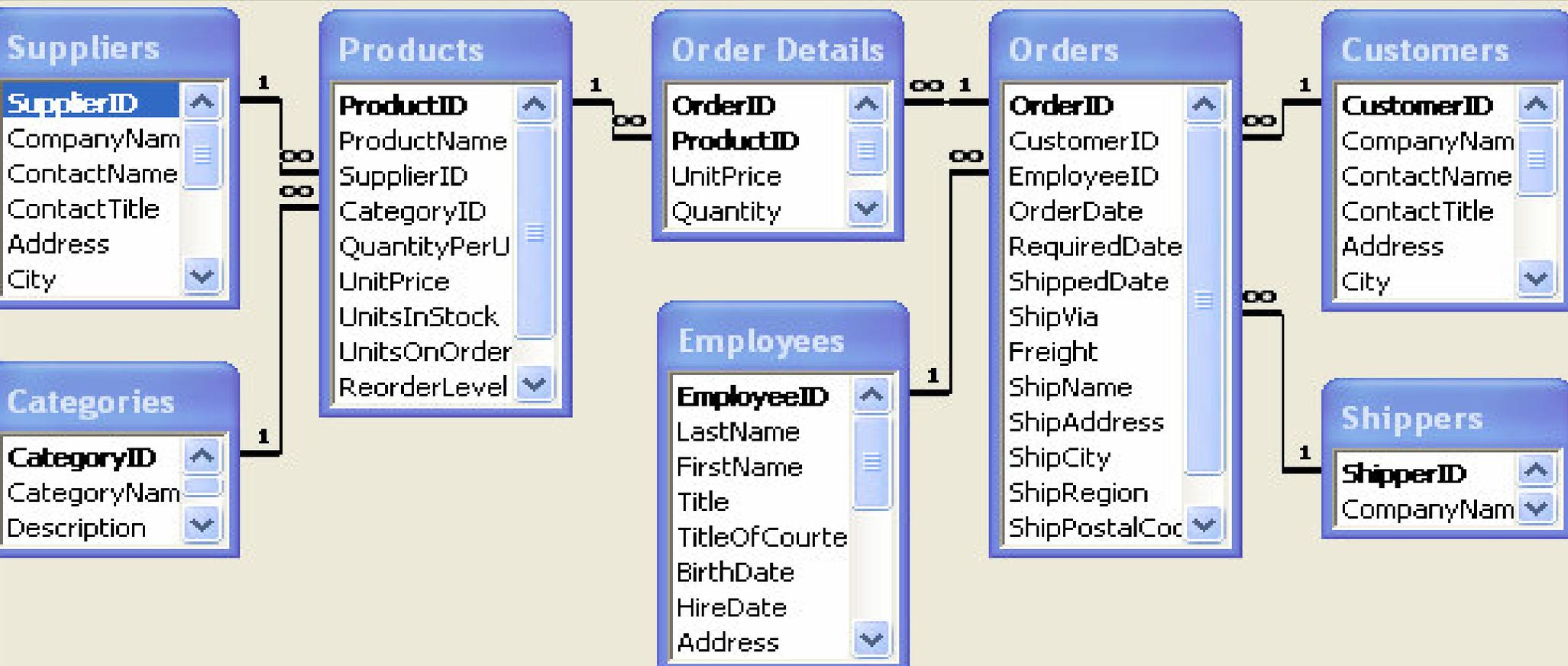
Reference to a Customer

Relational Database Management Systems (DBMS)

- Allows the creation of relational databases
- Supports specialized languages for easy retrieval of data from a set of inter-related tables
- Supports easy construction of a Graphical User Interface on top of the database
- Allows very large table sizes
- Provides security, fault tolerance, multi-user support, etc.

Microsoft Access - [Relationships]

File Edit View Relationships Tools Window Help Show Me



SQL – Structured Query Language

- Every statement yields a table of values as output
 - Sometimes there's only one row in the table!

select	columns and/or expressions
from	tables
where	conditions on the rows
group by	group rows together
order by	order the rows

Display an Entire Table

```
SELECT *  
FROM Employees;
```

Choose Columns

- Choosing a subset of columns is sometimes called a "project" operation
- Display first and last name of all employees

```
SELECT FirstName, LastName  
FROM Employees;
```

- Display company name and contact name for all customers

Choose Rows

- **Find US Employees**

```
SELECT FirstName, LastName  
FROM Employees  
WHERE Country = "USA";
```

- **Find employees hired after Jan. 1, 1993**

Compute Columns

- **Find total inventory value of each product**

```
SELECT ProductName,  
       UnitPrice*UnitsInStock AS TotalValue  
FROM Products;
```

- **Nice names for output columns**
 - Name following computed column (e.g., TotalValue) will be used to name output column
- **Find total price for each line item in “Order Details” table**

Sorting

- **Can sort output by contents of a column**
 - sort in ascending or descending order
 - sort by more than one column (second one breaks ties)
- **Sort products by total inventory value**

```
SELECT ProductName,  
       UnitPrice*UnitsInStock AS TotalValue  
FROM Products  
ORDER BY TotalValue DESC;
```
- **What are our 10 most expensive products?**

Aggregates

- **Can make calculations on entire columns**

- sum, avg, max, min, count

- **What is the total value of a given customer order**

```
SELECT OrderID, Sum([UnitPrice]*[Quantity]*(1-[Discount]))  
        AS Subtotal  
FROM [Order Details]  
WHERE OrderID=11001;
```

- returns a table with just one row!

- **What is average unit price of our products?**

Grouping and Aggregates

- **Each different value for the GROUP BY fields defines a new group**
 - One row of output is produced for each group
 - Several rows of input table may belong to same group. They are aggregated using aggregation operator.

- **Compute total value of all orders**

```
SELECT OrderID,  
       Sum([UnitPrice]*[Quantity]*(1-[Discount]))  
       AS Subtotal  
FROM [Order Details]  
GROUP BY OrderID;
```

- **Create a table that shows how many line items are in each order**

Joins

- **Combine rows from one table with rows from another**
- **Usually join on some common column**
 - Don't combine rows unless their value in the common column is the same
 - WHERE clause says the common column must be same in each table
- **Produce a list of all products and their categories**

```
SELECT Products.ProductName,  
       Categories.CategoryName  
FROM Categories, Products  
WHERE  
       Categories.CategoryID = Products.CategoryID;
```


SQL Summary

select	columns and/or expressions
from	tables
where	conditions on the rows
group by	group rows together
order by	order the rows