The single entity

I want to be alone

Greta Garbo

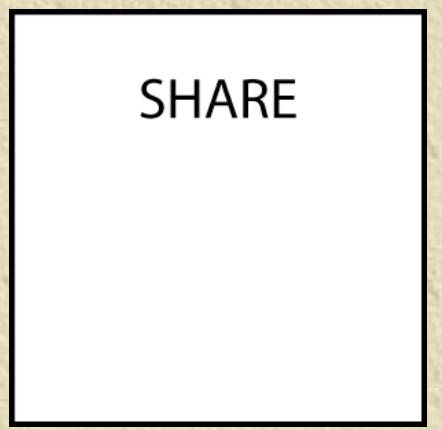
Modeling reality

- A database must mirror the real world if it is to answer questions about the real world
- Data modeling is a design technique for capturing reality

Reality matters

An entity

- Some thing in the environment
- Represented by a rectangle
- An instance is a particular occurrence of an entity



Attributes

- An attribute is a discrete data element that describes an entity
- Attribute names must be unique within a data model
- Attribute names must be meaningful

Identifiers

- Every instance of an entity must be uniquely identified
- An identifier can be an attribute or collection of attributes
- An identifier can be created if there is no obvious attribute
- A leading asterisk denotes an identifier

Global legal entity identifier(LEI)

- No global standard for identifying legal entities
- Lehman Brothers collapse in 2008
  - 209 registered subsidiaries, legal entities, in 21 countries
  - Party to more than 900,000 derivatives contracts
  - Creditors were unable to assess their exposure
- Transitive nature of many investments (i.e., A owes B, B owes C, and C owes D)
- LEI is in the process of global adoption

Exercise

- Design a data model for recording details of Olympic cities
- See http://en.wikipedia.org/wiki/List_of_Olympic_Games_host_cities

Rules for creating a table

- Each entity becomes a table
- The entity name becomes the table name
- Each attribute becomes a column
- The identifier becomes the primary key

Defining a table

```
CREATE TABLE share (
    shrcode       CHAR(3),
    shrfirm       VARCHAR(20)NOT NULL,
    shrprice      DECIMAL(6,2),
    shrqty        DECIMAL(8),
    shrdiv        DECIMAL(5,2),
    shrpe         DECIMAL(2),
      PRIMARY KEY(shrcode));
```

Are the data types selected a good choice?

Allowable data types

SQL standard

Numeric	integer	A 31-bit signed binary value
	smallint	A 15-bit signed binary value
	float(*p*)	A scientific format number of *p* binary digits precision
	decimal(*p*,*q*)	A packed decimal number of *p* digits total length; *q* decimal places to the right of the decimal point may be specified
String	char(*n*)	A fixed length character string of *n* characters
	varchar(*n*)	A variable length character string up to *n* characters
	text	A variable-length character string of up to 65,535 characters
Date/time	date	Date in the form *yyyymmdd*
	time	Time in the form *hhmmss*
	timestamp	A combination of date and time to the nearest microsecond
	time with time zone	Same as time, with the addition of an offset from UTC of the specified time
	timestamp with time zone	Same as timestamp, with the addition of an offset from UTC of the specified time

The share table

share					
shrcode	shrfirm	shrprice	shrqty	shrdiv	shrpe
FC	Freedonia Copper	27.50	10529	1.84	16
PT	Patagonian Tea	55.25	12635	2.50	10
AR	Abyssinian Ruby	31.82	22010	1.32	13
SLG	Sri Lankan Gold	50.37	32868	2.68	16
ILZ	Indian Lead & Zinc	37.75	6390	3.00	12
BE	Burmese Elephant	0.07	154713	0.01	3
BS	Bolivian Sheep	12.75	231678	1.78	11
NG	Nigerian Geese	35.00	12323	1.68	10
CS	Canadian Sugar	52.78	4716	2.50	15
ROF	Royal Ostrich Farms	33.75	1234923	3.00	6

Inserting rows

```
INSERT INTO share
   (shrcode,shrfirm,shrprice,shrqty,shrdiv,shrpe)
   VALUES ('FC','Freedonia Copper',27.5,10529,1.84,16);
```

Or

```
INSERT INTO share
   VALUES ('FC','Freedonia Copper',27.5,10529,1.84,16);
```

Importing from a text file

```
LOAD DATA LOCAL INFILE
'/Users/rtw/desktop/share.txt'  INTO TABLE SHARE
FIELDS TERMINATED BY ','
ENCLOSED BY "'"
LINES TERMINATED BY '\r’

FC,'Freedonia Copper',27.5,10529,1.84,16
PT,'Patagonian Tea',55.25,12635,2.5,10
AR,'Abyssinian Ruby',31.82,22010,1.32,13
SLG,'Sri Lankan Gold',50.37,32868,2.68,16
ILZ,'Indian Lead & Zinc',37.75,6390,3,12
BE,'Burmese Elephant',0.07,154713,0.01,3
BS,'Bolivian Sheep',12.75,231678,1.78,11
NG,'Nigerian Geese',35,12323,1.68,10
CS,'Canadian Sugar',52.78,4716,2.5,15
ROF,'Royal Ostrich Farms',33.75,1234923,3,6
```

Querying a table

List all data in the share table.

```
SELECT * FROM share;
```

shrcode	shrfirm	shrprice	shrqty	shrdiv	shrpe
FC	Freedonia Copper	27.50	10529	1.84	16
PT	Patagonian Tea	55.25	12635	2.50	10
AR	Abyssinian Ruby	31.82	22010	1.32	13
SLG	Sri Lankan Gold	50.37	32868	2.68	16
ILZ	Indian Lead & Zinc	37.75	6390	3.00	12
BE	Burmese Elephant	0.07	154713	0.01	3
BS	Bolivian Sheep	12.75	231678	1.78	11
NG	Nigerian Geese	35.00	12323	1.68	10
CS	Canadian Sugar	52.78	4716	2.50	15
ROF	Royal Ostrich Farms	33.75	1234923	3.00	6

Project

- Choosing columns
- A vertical slice

share

shrcode	shrfirm	shrprice	shrqty	shrdiv	shrpe
FC	Freedonia Copper	27.50	10529	1.84	16
PT	Patagonian Tea	55.25	12635	2.50	10
AR	Abyssinian Ruby	31.82	22010	1.32	13
SLG	Sri Lankan Gold	50.37	32868	2.68	16
ILZ	Indian Lead & Zinc	37.75	6390	3.00	12
BE	Burmese Elephant	0.07	154713	0.01	3
BS	Bolivian Sheep	12.75	231678	1.78	11
NG	Nigerian Geese	35.00	12323	1.68	10
CS	Canadian Sugar	52.78	4716	2.50	15
ROF	Royal Ostrich Farms	33.75	1234923	3.00	6

Project

Report a firm's name and price-earnings ratio.

```
SELECT shrfirm, shrpe FROM share;
```

shrfirm	shrpe
Freedonia Copper	16
Patagonian Tea	10
Abyssinian Ruby	13
Sri Lankan Gold	16
Indian Lead & Zinc	12
Burmese Elephant	3
Bolivian Sheep	11
Nigerian Geese	10
Canadian Sugar	15
Royal Ostrich Farms	6

Restrict

- Choosing rows
- A horizontal slice

share

shrcode	shrfirm	shrprice	shrqty	shrdiv	shrpe
FC	Freedonia Copper	27.50	10529	1.84	16
PT	Patagonian Tea	55.25	12635	2.50	10
AR	Abyssinian Ruby	31.82	22010	1.32	13
SLG	Sri Lankan Gold	50.37	32868	2.68	16
ILZ	Indian Lead & Zinc	37.75	6390	3.00	12
BE	Burmese Elephant	0.07	154713	0.01	3
BS	Bolivian Sheep	12.75	231678	1.78	11
NG	Nigerian Geese	35.00	12323	1.68	10
CS	Canadian Sugar	52.78	4716	2.50	15
ROF	Royal Ostrich Farms	33.75	1234923	3.00	6

Restrict

Get all firms with a price-earnings ratio less than 12.

```
SELECT * FROM share WHERE shrpe < 12;
```

shrcode	shrfirm	shrprice	shrqty	shrdiv	shrpe
PT	Patagonian Tea	55.25	12635	2.50	10
BE	Burmese Elephant	0.07	154713	0.01	3
BS	Bolivian Sheep	12.75	231678	1.78	11
NG	Nigerian Geese	35.00	12323	1.68	10
ROF	Royal Ostrich Farms	33.75	1234923	3.00	6

Project and restrict combo

Choosing rows and columns

List the firm's name, price, quantity, and dividend where share holding is at least 100,000.

```
SELECT shrfirm, shrprice, shrqty, shrdiv
   FROM share WHERE shrqty >= 100000;
```

shrfirm	shrprice	shrqty	shrdiv
Burmese Elephant	0.07	154713	0.01
Bolivian Sheep	12.75	231678	1.78
Royal Ostrich Farms	33.75	1234923	3.00

Exercise

- Report the name and price of those shares where the share price is greater than 10

Primary key retrieval

- A query using the primary key returns at most one row

Report firms whose code is AR.

```
SELECT * FROM share WHERE shrcode = 'AR';
```

shrcode	shrfirm	shrprice	shrqty	shrdiv	shrpe
AR	Abyssinian Ruby	31.82	22010	1.32	13

Primary key retrieval

- A query not using the primary key can return more than one row

Report firms with a dividend of 2.50.

```
SELECT * FROM share WHERE shrdiv = 2.5;
```

shrcode	shrfirm	shrprice	shrqty	shrdiv	shrpe
PT	Patagonian Tea	55.25	12635	2.50	10
CS	Canadian Sugar	52.78	4716	2.50	15

IN

- Used with a list of values

Report data on firms with codes of FC, AR, or SLG.

```
SELECT * FROM share WHERE shrcode IN
('FC','AR','SLG');
```

or

```
SELECT * FROM share WHERE shrcode = 'FC' OR
  shrcode = 'AR' OR shrcode = 'SLG';
```

shrcode	shrfirm	shrprice	shrqty	shrdiv	shrpe
FC	Freedonia Copper	27.50	10529	1.84	16
AR	Abyssinian Ruby	31.82	22010	1.32	13
SLG	Sri Lankan Gold	50.37	32868	2.68	16

NOT IN

Not in a list of values

Report all firms other than those with the code CS or PT.

```
SELECT * FROM share WHERE shrcode NOT IN ('CS', 'PT');
```

is equivalent to:

```
SELECT * FROM share WHERE shrcode <> 'CS' AND shrcode <>
```

shrcode	shrfirm	shrprice	shrqty	shrdiv	shrpe
AR	Abyssinian Ruby	31.82	22010	1.32	13
SLG	Sri Lankan Gold	50.37	32868	2.68	16
ILZ	Indian Lead & Zinc	37.75	6390	3.00	12
BE	Burmese Elephant	0.07	154713	0.01	3
BS	Bolivian Sheep	12.75	231678	1.78	11
NG	Nigerian Geese	35.00	12323	1.68	10
ROF	Royal Ostrich Farms	33.75	1234923	3.00	6

Ordering output

- Ordering columns
  - Columns are reported in the order specified in the SQL command
- Ordering rows
  - Rows are ordered using the `ORDER BY` clause

Ordering columns

```
SELECT shrcode, shrfirm FROM share WHERE shrpe = 10;
```

shrcode	shrfirm
PT	Patagonian Tea
NG	Nigerian Geese

```
SELECT shrfirm, shrcode FROM share WHERE shrpe = 10;
```

shrfirm	shrcode
Patagonian Tea	PT
Nigerian Geese	NG

Ordering rows

List all firms where PE is at least 12, and order the report in descending PE. Where PE ratios are identical, list firms in alphabetical order.

```
SELECT * FROM share WHERE shrpe >= 12
    ORDER BY shrpe DESC, shrfirm;
```

shrcode	shrfirm	shrprice	shrqty	shrdiv	shrpe
FC	Freedonia Copper	27.50	10529	1.84	16
SLG	Sri Lankan Gold	50.37	32868	2.68	16
CS	Canadian Sugar	52.78	4716	2.50	15
AR	Abyssinian Ruby	31.82	22010	1.32	13
ILZ	Indian Lead & Zinc	37.75	6390	3.00	12

Calculating

Get firm name, price, quantity, and firm yield.

```
SELECT shrfirm, shrprice, shrqty,
   shrdiv/shrprice*100 AS yield FROM
   share;
```

shrfirm	shrprice	shrqty	yield
Freedonia Copper	27.50	10,529	6.69
Patagonian Tea	55.25	12,635	4.52
Abyssinian Ruby	31.82	22,010	4.15
Sri Lankan Gold	50.37	32,868	5.32
Indian Lead & Zinc	37.75	6,390	7.95
Burmese Elephant	0.07	154,713	14.29
Bolivian Sheep	12.75	231,678	13.96
Nigerian Geese	35.00	12,323	4.80
Canadian Sugar	52.78	4,716	4.74
Royal Ostrich Farms	33.75	1,234,923	8.89

Exercise

- Calculate the total dividends earned by each share. Report the name of the firm and the payment sorted from highest to lowest payment.

Built-in functions

- COUNT, AVG, SUM, MIN, and MAX

Find the average dividend.

```
SELECT AVG(shrdiv) AS avgdiv FROM
share;
```

avgdiv
2.03

What is the average yield for the portfolio?

```
SELECT AVG(shrdiv/shrprice*100) AS
avgyield FROM share;
```

avgyield
7.53

COUNT

- `COUNT(*)` counts all rows
- `COUNT(columname)` counts rows with non null values for `columname`

Subqueries

- A query within a query

  Report all firms with a PE ratio greater than the average for the portfolio.

```
SELECT shrfirm, shrpe FROM share WHERE
shrpe >(SELECT AVG(shrpe)FROM share);
```

shrfirm	shrpe
Freedonia Copper	16
Abyssinian Ruby	13
Sri Lankan Gold	16
Indian Lead & Zinc	12
Canadian Sugar	15

Regular expression

- A concise and flexible method for string searching
- Commands are handled by a regular expression processor
- Supported by many programming languages

Regular expression

Search for a string

List all firms containing 'Ruby' in their name.

```
SELECT shrfirm FROM share
  WHERE shrfirm REGEXP 'Ruby';
```

shrfirm
Abyssinian Ruby

Regular expression

- Search for alternative strings
- [a|b] finds 'a' or 'b'
- | is the alternation symbol
- *List the firms containing gold or zinc in their name.*

```
SELECT * FROM share
  WHERE shrfirm
     REGEXP 'gold|zinc|Gold|Zinc';
```

Regular expression

- Search for a beginning string
- ^ means at the start of the string
- *List the firms whose name begins with Sri.*

```
SELECT * FROM share
  WHERE shrfirm REGEXP '^Sri';
```

Regular expression

- Search for a ending string
- $ means at the end of the string
- *List the firms whose name ends in Geese.*

```
SELECT shrfirm
 FROM share
  WHERE shrfirm REGEXP 'Geese$';
```

Exercise

- List names of shares whose name contains sheep or geese

DISTINCT

- Eliminating duplicate rows

Find the number of different PE ratios.

```
SELECT COUNT(DISTINCT shrpe)AS 'Different PEs'
    FROM share;
```

Different PEs
8

DISTINCT column-name is not implemented by all relational systems

DISTINCT

- Eliminating duplicate rows when reporting

  Report the different values of the PE ratio.

```
SELECT DISTINCT shrpe FROM share;
```

shrpe
3
6
10
11
12
13
15
16

DELETE - deleting rows

Erase the data for Burmese Elephant. All the shares have been sold.

```
DELETE FROM share
   WHERE shrfirm = 'Burmese Elephant';
```

UPDATE - changing rows

Change the share price of FC to 31.50.

```
UPDATE share
    SET shrprice = 31.50
    WHERE shrcode = 'FC';
```

UPDATE - changing rows

Increase the total number of shares for Nigerian Geese by 10% because of the recent bonus issue.

```
UPDATE share
    SET shrqty = shrqty*1.1
    WHERE shrfirm = 'Nigerian Geese';
```

Quotes

- Three kinds of quotes
  - Single ' (must be straight not curly)
  - Double " (must be straight not curly)
  - Back ` (left of 1 key)
- In MySQL, the first two are equivalent and can be used interchangeably

```
SELECT `person first` FROM person WHERE
`person last` = "O'Hara";
```

Summary

- Introduced
  - Entity
  - Attribute
  - Identifier
  - SQL
    - `CREATE`
    - `INSERT`
    - `SELECT`
    - `DELETE`
    - `UPDATE`