Database Design

John Kim
Field Station Programs
San Diego State University

A few downloads

Exercise Data:

- NPP Data in text format (zipped)

Software:

- DB Designer
- MySQL Administrator & MySQL Query Browser Bundle

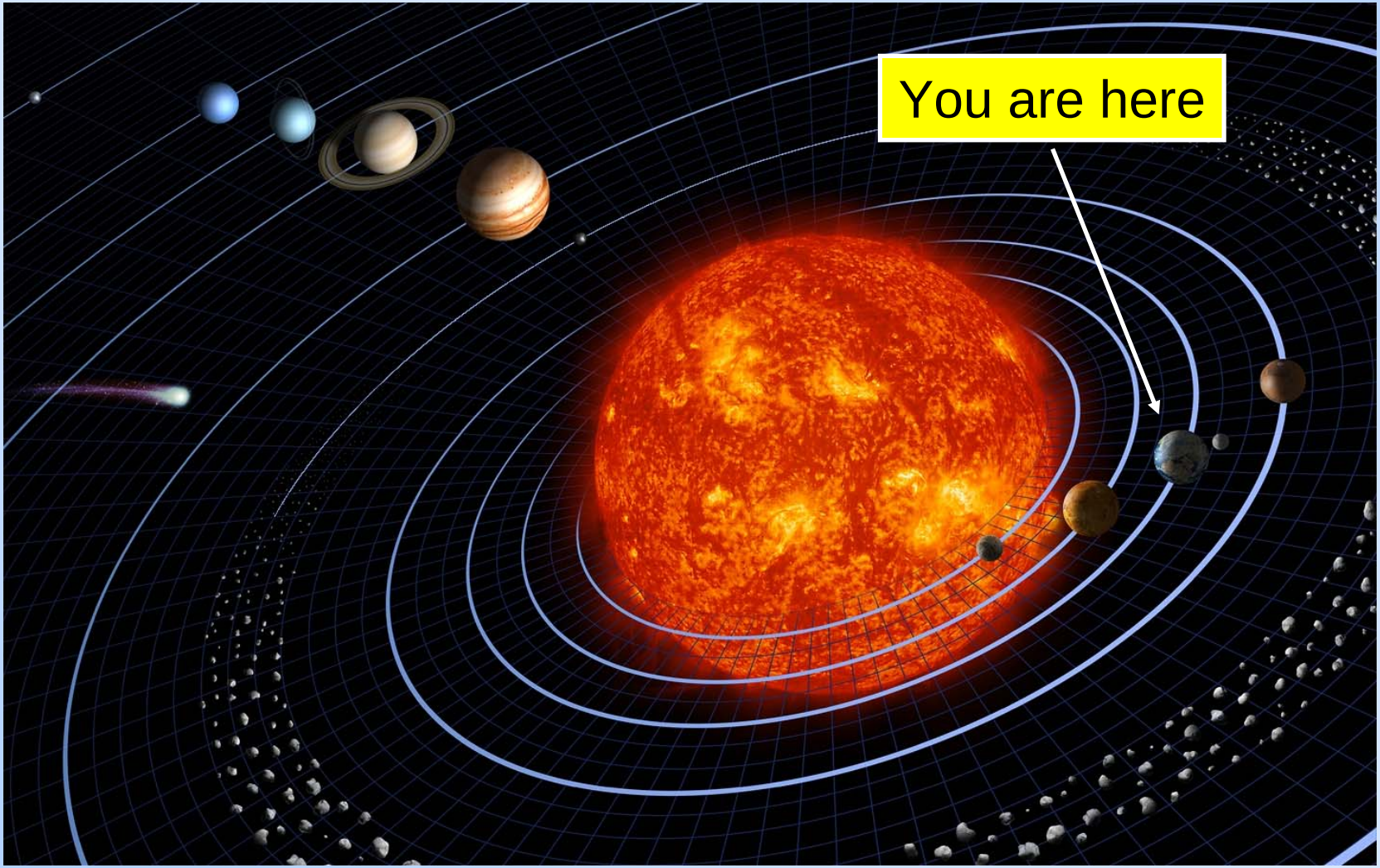
http://jkim.sdsu.edu/LaSelva/

Database Skills

- Scientific context, needs, and goals

- Database Theory
- Database software (e.g., MySQL, Oracle)
- SQL
- Operating System (Windows, Linux)
- Computer hardware
- Networking
- Webserver
- Webpage authoring
- Webpage design & graphics
- Webpage scripting languages (e.g., PHP, Perl)

Perspective



Agenda

Today

- DB Design Methods
- Design Exercise
- Implementing DB with MS Access
- QA/QC
- Intro to SQL

Tomorrow

- Intro to MySQL
- Tools for MySQL
- Data Query in MySQL

3 Basic Steps To DB Design

1) Split data into tables.

2) Determine data type of each column.

3) Identify relations among the tables

1) Split Data via Normalization

(abbreviated version)

Starting with a large table…

A) Identify a primary key for each table.
(what is a primary key?)

B) Eliminate duplicate columns.

C) Eliminate duplicate rows.

A) Identify a primary key in each table

Books

TITLE	AUTHOR 1	AUTHOR 2	PUBLISHER	ISBN	QTY.
Ecology 101	Smith, A.B.	Gordon, D.A.	Univ. Press	4873895759	4324
Ecology for Dummies	Doe, J.		Wiley & Sons	0493802020	8998
Ecology and Politics	Kim, J.B.		McGraw-Hill	7482929292	900
Ecology and Modern Cinema	Kim, J.B.		Univ. Press	2234849302	1

Personnel

id	Last	First	M.I.	Institution	Sector	Position 1	Position 2
1	Smith	Ann	A	SDSU	Academic	P.I.	Community Liaison
2	Smith	Ann	Z	Acme Inc.	Private	Administrator	Field Technician
3	Kim	John	B	SDSU	Academic	P.I.	Data Manager

B) Eliminate duplicate columns

Books

TITLE	AUTHOR 1	AUTHOR 2	PUBLISHER	ISBN	QTY.
Ecology 101	Smith, A.B.	Gordon, D.A.	Univ. Press	4873895759	4324
Ecology for Dummies	Doe, J.		Wiley & Sons	0493802020	8998
Ecology and Politics	Kim, J.B.		McGraw-Hill	7482929292	900
Ecology and Modern Cinema	Kim, J.B.		Univ. Press	2234849302	1

Primary Keys

Foreign Key

Books

TITLE	PUBLISHER	ISBN	QTY.
Ecology 101	Univ. Press	4873895759	4324
Ecology for Dummies	Wiley & Sons	0493802020	8998
Ecology and Politics	McGraw-Hill	7482929292	900
Ecology and Modern Cinema	Univ. Press	2234849302	1

Authors

Id	ISBN	Name
1	4873895759	Smith, A.B.
2	4873895759	Gordon, D.A.
3	0493802020	Doe, J.
4	7482929292	Kim, J.B.
5	2234849302	Kim, J.B.

B) Eliminate duplicate columns

id	Last	First	M.I.	Institution	Sector	Position 1	Position 2
1	Smith	Jane	A	SDSU	Academic	P.I.	Community Liaison
2	Smith	Jane	Z	Acme Inc.	Private	Administrator	Field Technician
3	Kim	John	B	SDSU	Academic	P.I.	Data Manager

personnel

id	Last	First	M.I.	Institution	Sector
1	Smith	Ann	A	SDSU	Academic
2	Smith	Ann	Z	Acme Inc.	Private
3	Kim	John	B	SDSU	Academic

positions

id	personnel_id	Position
1	1	P.I.
2	1	Community Liaison
3	2	Administrator
4	2	Field Technician
5	3	P.I.
6	3	Data Manager

Foreign Key

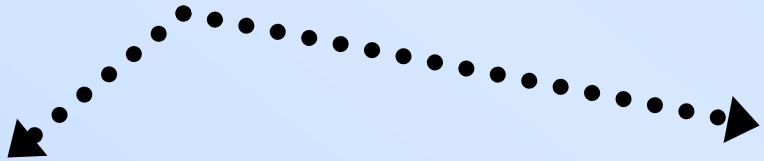
C) Eliminate duplicate rows

Books

TITLE	PUBLISHER	ISBN	QTY.
Ecology 101	Univ. Press	4873895759	4324
Ecology for Dummies	Wiley & Sons	0493802020	8998
Ecology and Politics	McGraw-Hill	7482929292	900
Ecology and Modern Cinema	Univ. Press	2234849302	1

Authors

Id	ISBN	Name
1	4873895759	Smith, A.B.
2	4873895759	Gordon, D.A.
3	0493802020	Doe, J.
4	7482929292	Kim, J.B.
5	2234849302	Kim, J.B.



Books

TITLE	PUBLISHER_id	ISBN	QTY.
Ecology 101	1	4873895759	4324
Ecology for Dummies	2	0493802020	8998
Ecology and Politics	3	7482929292	900
Ecology and Modern Cinema	1	2234849302	1

Publishers

Publisher_id	PUBLISHER
1	Univ. Press
2	Wiley & Sons
3	McGraw-Hill

Which are primary keys? Foreign keys? What about Authors?

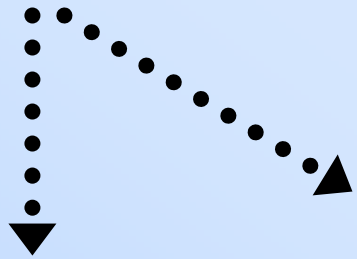
C) Eliminate duplicate rows

personnel

id	Last	First	M.I.	Institution	Sector
1	Smith	Ann	A	SDSU	Academic
2	Smith	Ann	Z	Acme Inc.	Private
3	Kim	John	B	SDSU	Academic

positions

id	personnel_id	Position
1	**0**	P.I.
2	**0**	Community Liaison
3	**1**	Administrator
4	**1**	Field Technician
5	**2**	P.I.
6	**2**	Data Manager



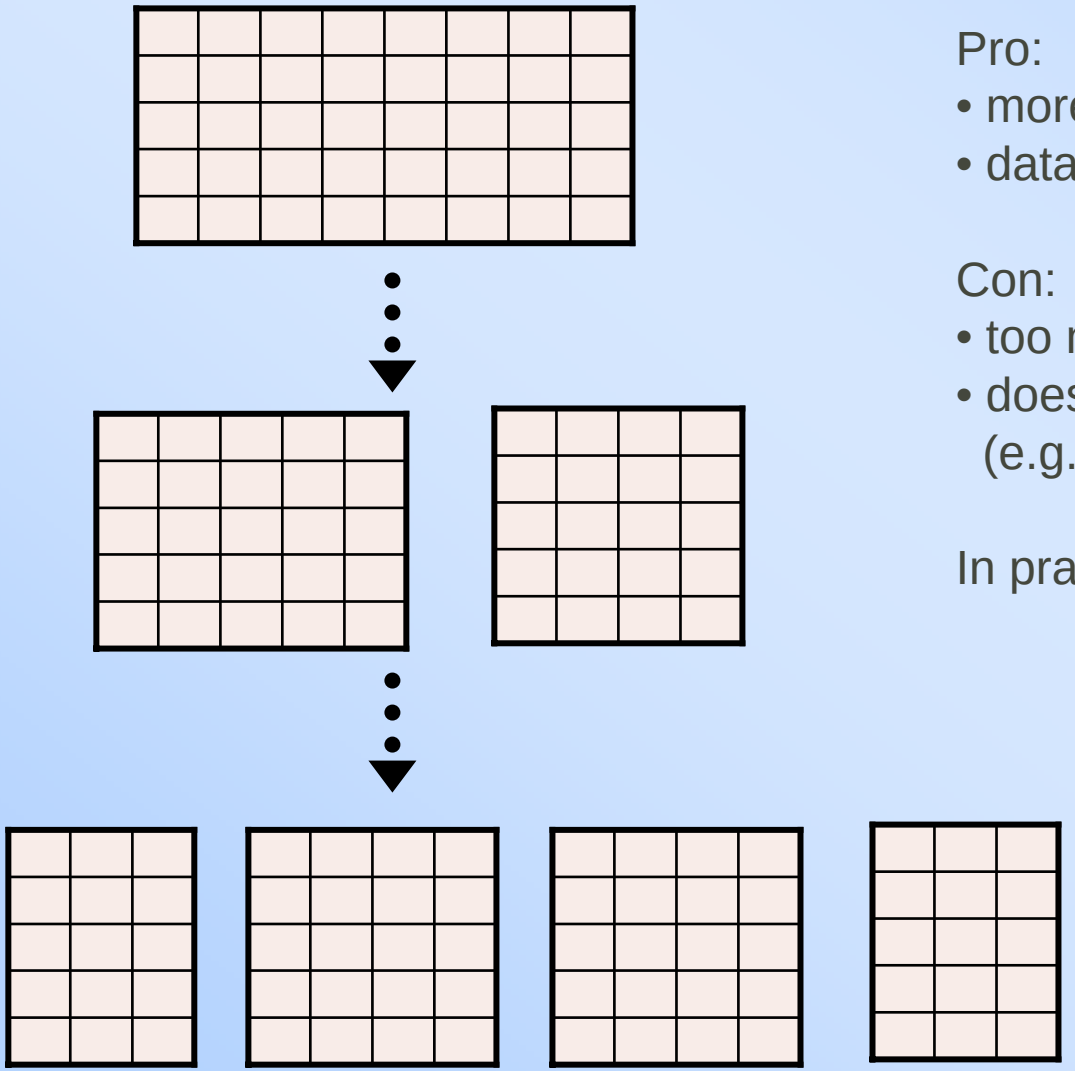
personnel

id	Last	First	M.I.	Institution_id
1	Smith	Ann	A	1
2	Smith	Ann	Z	2
3	Kim	John	B	1

institutions

Institution_id	Institution	Sector
1	SDSU	Academic
2	Acme Inc.	Private

Normalization – How far do you go?



Pro:
- more flexible relationships
- data quality

Con:
- too many tables
- doesn't reflect concept (e.g., address)

In practice: judgment call

2) Determine Data Types

Numeric types		Examples
integer	integers	2005
float	single-precision real numbers (up to 23 places)	3.1415926
double	double-precision real numbers (up to 53 places)	3.141592653589 7932384626433 8327950288419 716939937510

2) Determine Data Types

Text types		Examples
char char(n)	Characters	“pantroglodytes”
varchar(n)	variable length characters (< 255 chars)	“could not determine species”
text	a large chunk of text	
enum	enumerated values	‘north’, ‘south’, ‘east’, ‘west’
set	a set of values	‘heard’,’seen’,’captured’

2) Determine Data Types

Date & Time:		
DATETIME	‘YYYY-MM-DD HH:MM:SS'	‘2005-01-05 15:07:43'
DATE	‘YYYY-MM-DD'	'0000-00-00'
TIMESTAMP	Seconds passed since t_o	148939284758498
TIME	‘HH:MM:SS'	’15:07:43'
YEAR	YYYY	2005

More on column types:
http://dev.mysql.com/doc -> Section 11 “Column Types”

2) Determine Data Types

Books

TITLE	varchar(255)
PUBLISHER_id	integer
ISBN	varchar(10)
QTY.	integer

Authors

Id	integer
ISBN	varchar(10)
Author	varchar(255)

Publishers

PUBLISHER_id	integer
PUBLISHER	varchar(255)

Personnel

id	integer
Last	varchar(40)
First	varchar(40)
M.I.	char(1)
Institution_id	integer

institutions

Institution_id	integer
Institution	varchar(255)
Sector	enum('academic','industry','govt','NGO')

Positions

id	integer
personnel_id	integer
position	varchar

3) Identify the relations among tables.

Table X

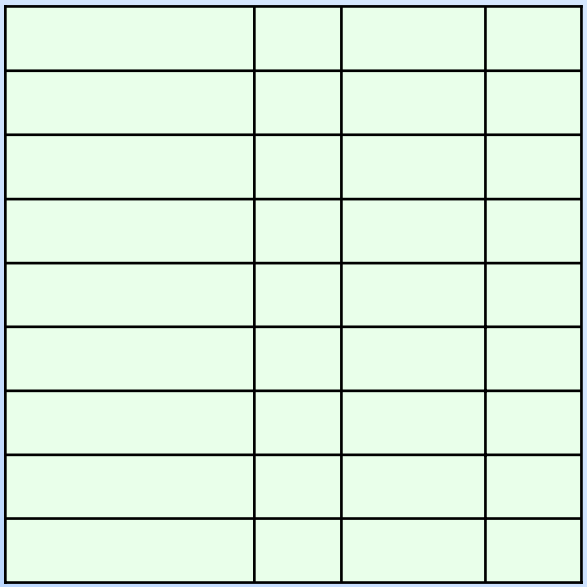
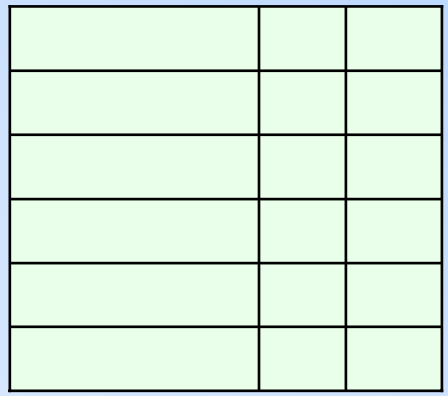


Table Y



Relations Types:

1 to 1
1 to many
many to many

many to many: problematic. May need to create a 3rd table to relate the two.

3) Identify the relations among tables.

Books

TITLE	PUBLISHER_id	ISBN	QTY.
Ecology 101	0	4873895759	4324
Ecology for Dummies	1	0493802020	8998
Ecology and Politics	2	7482929292	900
Ecology and Modern Cinema	0	2234849302	1

1 to many

1 to many

Authors

Id	ISBN	Author
0	4873895759	Smith, A.B.
1	4873895759	Gordon, D.A.
2	0493802020	Doe, J.
3	7482929292	Kim, J.B.
4	2234849302	Kim, J.B.

Publishers

PUBLISHER_id	PUBLISHER
0	Harcourt Brace
1	Wiley & Sons
2	McGraw-Hill

3) Identify the relations among tables

personnel

id	Last	First	M.I.	Institution_id
0	Smith	Ann	A	0
1	Smith	Ann	Z	1
2	Kim	John	B	0

positions

id	personnel_id	Position
0	**0**	P.I.
1	**0**	Community Liaison
2	**1**	Administrator
3	**1**	Field Technician
4	**2**	P.I.
5	**2**	Data Manager

1 to many

1 to many

1 to 1

institutions

Institution_id	Institution	Sector
0	SDSU	Academic
1	Acme Inc.	Private

contact_address

personnel_id	street	city	state
0	523 Main St.	Amherst	MA
1	1010 Sea St.	San Diego	CA
2	99 Ridge Way	Portland	ME