



Database Part II

Database Design

John Kim
Field Station Programs
San Diego State University





Database Design

- Format for this section:
 - 1) I present design strategies
 - 2) You design a db in groups
 - 3) We discuss your designs
 - 4) I show how to design in DB Designer
 - 5) You design in DB Designer





Database Design

3 main tasks:

Task 1: Split up the data into tables.

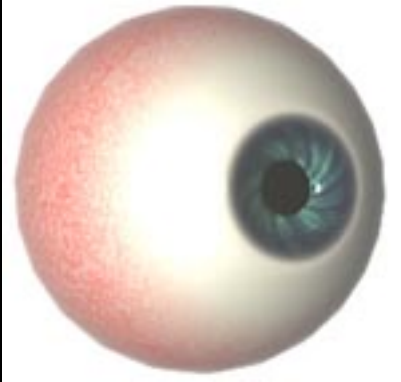
Task 2: Designate the type of data each column should hold.

Task 3: Identify relations among the tables



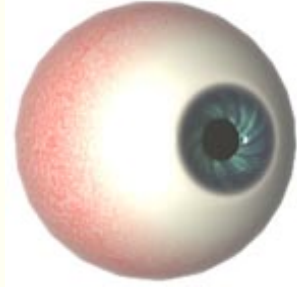


Task 1 – Split up data into tables ...there are 3 ways.





Task 1 - Split up data by “Eyeballing”



... or the quasi-“object-oriented” approach:
model tables after objects.

- Create tables that represent tangible things, e.g.,
 - Books
 - Organisms
 - Instruments

- Create tables that represent conceptual objects, e.g.,
 - Projects
 - Observations
 - Locations





Eyeballing – Book Example.

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



Books

TITLE	?	?
Ecology 101		
Ecology for Dummies		
Ecology and Politics		
Ecology and Modern Cinema		

Authors

AUTHOR	?	?
Smith, A.B.		
Doe, J.		
Kim, J.B.		
Gordon, D.A.		

Publishers

PUBLISHERS	?	?
Univ. Press		
Wiley & Sons		
McGraw-Hill		





Eyeballing – Personnel Example.

Personnel

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



Personnel

Last	First	M.I.	?	?
Smith	Ann	A		
Smith	Ann	Z		
Kim	John	B		

Institutions

Institution	Sector	?	?
SDSU	Academic		
Acme Inc.	Private		
SDSU	Academic		

Positions

Position	?	?
P.I.		
Administrator		
Community Liaison		
Field Technician		
Data Manager		





Task 1 - Split up data by “Normalizing”



▣ Abbreviated Normalization Steps:

Step 1: Identify a **primary key** for each table.

Step 2: Eliminate duplicate columns.

Step 3: Eliminate duplicate rows.





Normalization Step 1: Identify a **primary key** for every 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
0	Smith	Ann	A	SDSU	Academic	P.I.	Community Liaison
1	Smith	Ann	Z	Acme Inc.	Private	Administrator	Field Technician
2	Kim	John	B	SDSU	Academic	P.I.	Data Manager





Normalization Step 2: Eliminate duplicate columns (book example)

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



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

Foreign Key





Normalization Step 2: Eliminate duplicate columns (personnel example)

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



personnel

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

positions  Foreign Key

id	person_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





Normalization Step 3: Eliminate duplicate rows (book example)

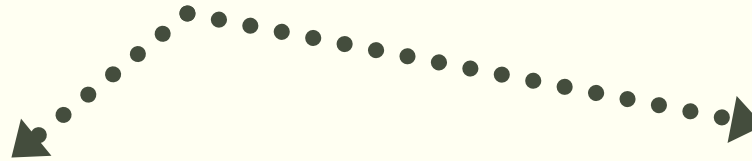
Books

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	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.

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

Publishers

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





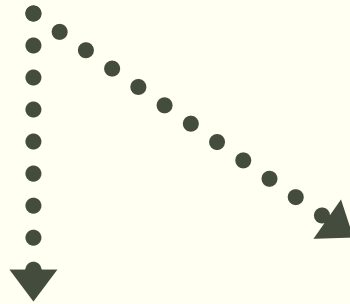
Normalization Step 3: Eliminate duplicate rows (personnel example)

personnel

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

positions

id	person_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



personnel

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

institutions

Institution_id	Institution	Sector
0	SDSU	Academic
1	Acme Inc.	Private

Task 1 - Split up data - the Zen way.

- The truth is attained by directly knowing.



- Attachment to worldly arguments lead only to suffering.



Task 2: Designate the type of data that each column should hold.

Available 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)	3.141592653 58979323846 26433832795 02884197169 39937510





Task 2: Designate the type of data that each column should hold.

Available data types (continued) –

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’



Task 2: Designate the type of data that each column should hold.

Available data types (continued) –

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_0	148939284758498
TIME	'HH:MM:SS'	'15:07:43'
YEAR	YYYY	2005

More on column types:

<http://www.mysql.com/documentation> -> Section 11 "Column Types"





Task 2: Designate the type of data that each column should hold.

Books

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

Authors

Id	integer
ISBN	varchar(10)
Author	varchar

Publishers

PUBLISHER_id	integer
PUBLISHER	varchar

Personnel

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

institutions

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

Positions

id	integer
person_id	integer
position	varchar





Task 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

1:1

1:n

m:n

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





Exercise: Design tables from NPP data

DATE	SITE	WEB	PLOT	QD	SPECIES	OBS	COVER	HEIGHT	COUNT	PHEN	COMMENTS
2/3/1999	FPC	1	E	1	ERPU8	1	0.5	4	13	V	NA
2/3/1999	FPC	1	E	1	ERPU8	2	0.1	2	16	V	NA
2/3/1999	FPC	1	E	1	GUSA2	1	0.01	4	2	V	NA
2/3/1999	FPC	1	E	1	GUSA2	2	0.1	5	1	V	NA
2/3/1999	FPC	1	E	1	GUSA2	3	0.5	12	1	V	NA
2/3/1999	FPC	1	E	1	LEFE	1	0.25	5	1	V	NA
2/3/1999	FPC	1	E	2	LATR2	1	7	36	2	V	NA
2/3/1999	FPC	1	E	2	LATR2	2	2	32	3	V	NA
2/3/1999	FPC	1	E	2	LATR2	3	8	61	1	V	NA
2/3/1999	FPC	1	E	2	LATR2	4	3	45	1	V	NA
2/3/1999	FPC	1	E	2	LATR2	5	2	24	1	V	NA
2/3/1999	FPC	1	E	2	ERPU8	1	0.25	3	3	V	NA
2/3/1999	FPC	1	E	2	ERPU8	2	0.05	2	11	V	NA
2/3/1999	FPC	1	E	2	LEFE	1	0.1	4	3	V	NA
2/3/1999	FPC	1	E	2	LEFE	2	0.25	5	8	V	NA
2/3/1999	FPC	1	E	2	GUSA2	1	0.01	5	1	V	NA
2/3/1999	FPC	1	E	3	LATR2	1	3	39	1	V	NA
2/3/1999	FPC	1	E	3	LATR2	2	2	21	2	V	NA
2/3/1999	FPC	1	E	3	ERPU8	1	0.25	4	23	V	NA





Exercise: Go for it...

Form into 4 groups

20 minutes





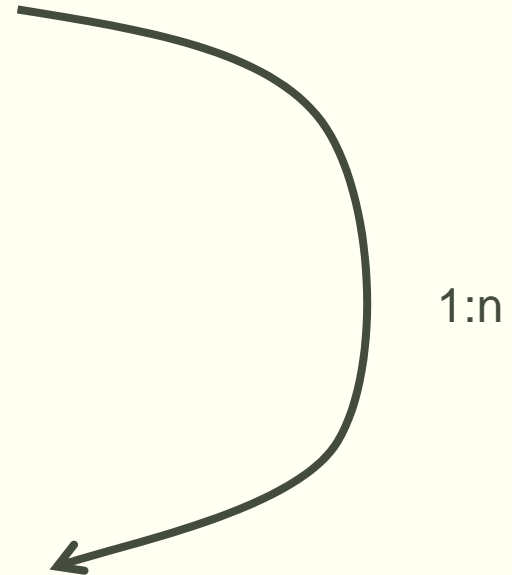
Exercise Results – What I would do (version 1)

Location

Site	enum('B','C','G','P')
Web	integer(1)
Plot	enum('N','S','E','W')
Quad	integer(2)
Location_id (PK)	integer

Observation

Date	date
Species	char(6)
Observation	integer
Cover	float
Height	float
Count	integer
Phen	enum('V','F','NA')
Comments	text
Location_id (FK)	integer
Observation_id (PK)	integer





Exercise Results – What I would do (alternate version)

Locations

Site
Web
Plot
Quad
Location_id (PK)

Visits

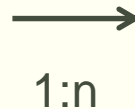
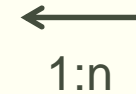
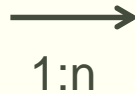
Date
Personnel
Visit_id (PK)

Observations

Visit_id (FK)
Species_id (FK)
Observation
Cover
Height
Count
Phen
Comments
Location_id (FK)
Observation_id (PK)

Species

Species
Scientific Name
Common Name
Species_id (PK)





Exercise: Design Using DB Designer

- ❑ Separate design tool to work w/ MySQL databases
- ❑ GUI-based, draw, drag & drop tools for designing tables, relations, and queries.
- ❑ Runs on your Windows machine
- ❑ Can connect to a database engine to create or modify databases.





Exercise: Design Using DB Designer – Create tables

DBDesigner 4 - [DB Model | NPP model]

File Edit Display Database Plugins Options Windows Help

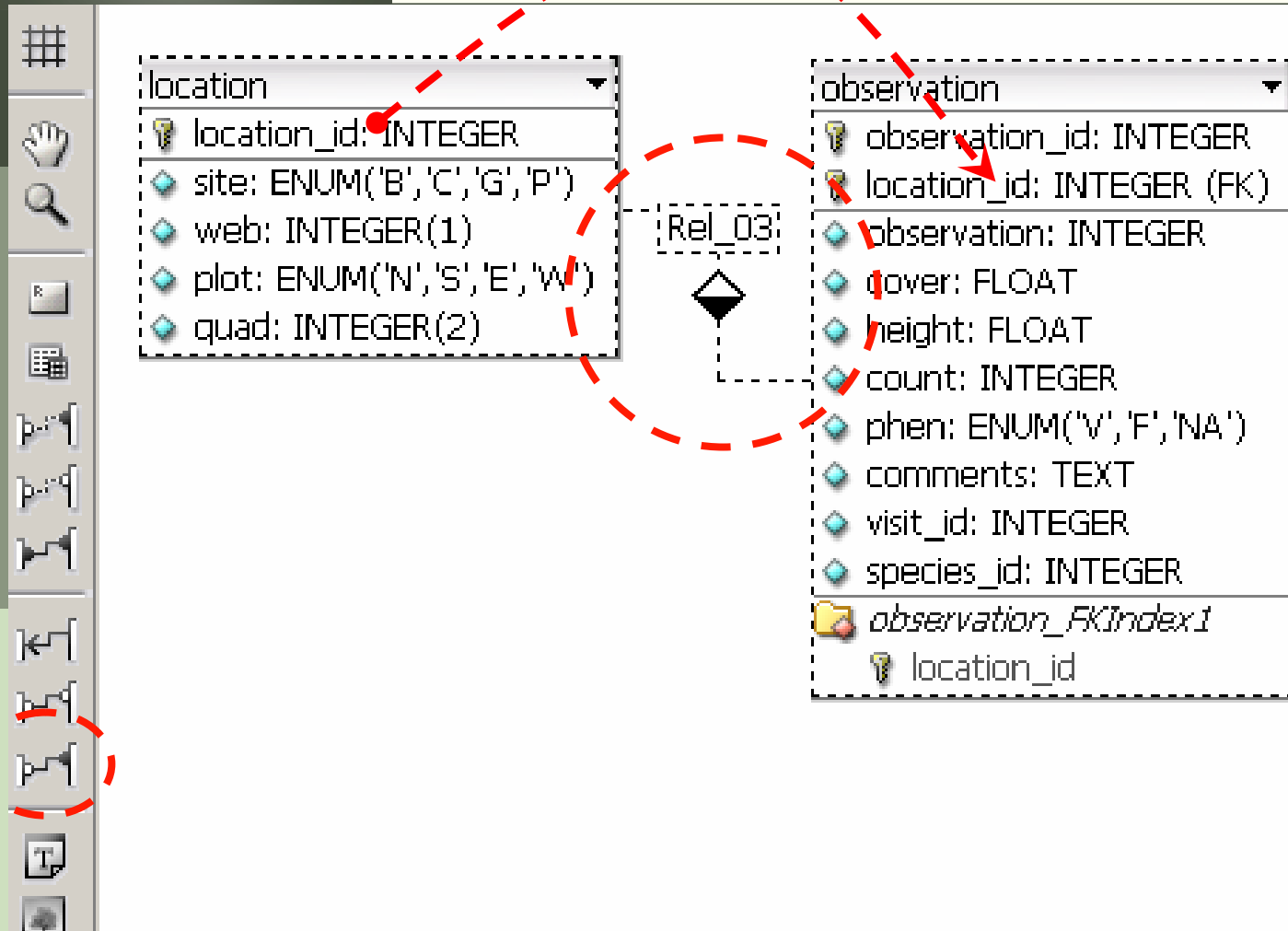
Table Editor

Table Name: Table_01 Table Prefix: Default (no prefix) Table Type: MYISAM (Standard)

Column Name	DataType	NN	AI	Flags
location_id	INTEGER	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> UNSIGNED <input type="checkbox"/> ZEROFILL
site	ENUM('B','C','G','P')	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
web	INTEGER(1)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/> UNSIGNED <input type="checkbox"/> ZEROFILL
plot	ENUM('N','S','E','W')	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
quad	INTEGER(2)	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/> UNSIGNED <input type="checkbox"/> ZEROFILL



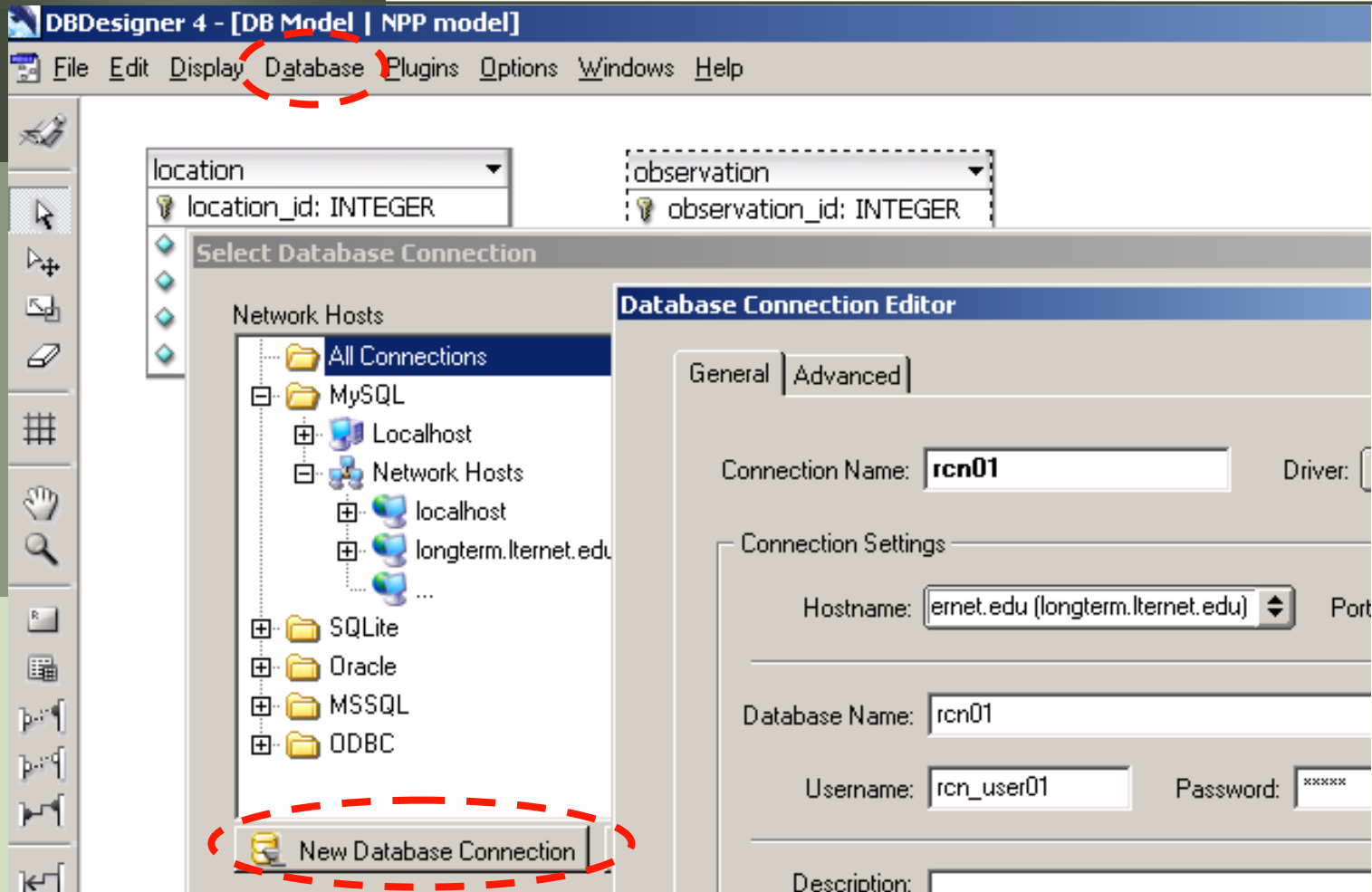
Exercise: Design Using DB Designer – Relate tables





Exercise: Design Using DB Designer – Synchronizing (Implementing)

Top Menu -> Database -> Connect to Database





Exercise: Design Using DB Designer

Go for it...

Locations

Site
Web
Plot
Quad
Location_id (PK)

Observations

Visit_id (FK)
Species_id (FK)
Observation
Cover
Height
Count
Phen
Comments
Location_id (FK)
Observation_id (PK)

Species

Species
Scientific Name
Common Name
Species_id (PK)

Visits

Date
Personnel
Visit_id (PK)

→
1:n

←
1:n

→
1:n

Hostname: longterm.lternet.edu
login: rcn_user01
pwd: rcn01
database: rcn01

