Introduction to database management with open source tools

Guillaume Larocque
research professional,
Quebec Center for Biodiversity Science

http://qcbs.ca/wiki/opendb

CENTRE DE LA SCIENCE DE LA BIODIVERSITÉ DU QUÉBEC
QUEBEC CENTRE FOR BIODIVERSITY SCIENCE

PostgreSQL
the worlds most advanced open source database

LibreOffice
The Document Foundation

Introduction to database management with open source tools

Guillaume Larocque
research professional,
Quebec Center for Biodiversity Science

http://qcbs.ca/wiki/opendb

CENTRE DE LA SCIENCE DE LA BIODIVERSITÉ DU QUÉBEC
QUEBEC CENTRE FOR BIODIVERSITY SCIENCE

PostgreSQL
the worlds most advanced open source database

LibreOffice
The Document Foundation

Objectives

- Understand what databases do
- Concepts of database management
- Intro to the SQL language
- Brief intro to LibreOffice Base
- Connecting PostgreSQL with other tools
- Where to find more info?

Getting started...

What is a database?

- An organized collection of data.

Options for database management

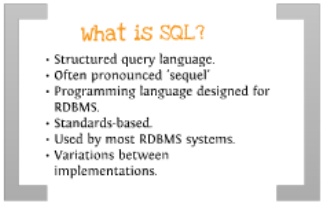
Benefits of open-Source Software

- Free today, free tomorrow.
- Collaborative development.
- Collaborative help system.
- Links more easily with other open source software.

- Disadvantages: not always good with user interface. No support from company.

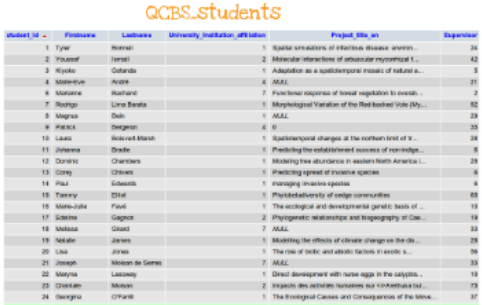
What is a database?

- An organized collection of data.



What is a relational database?

- Data is stored in formatted tables.
- Tables are linked together with "keys".



QCBS_students

student_id	Firstname	Lastname	University_institution_affiliation	Project_title_en	Supervisor
1	Tyler	[illegible]	1	Spatial simulations of infectious disease: enviro...	34
2	Youssef	Ismail	2	Molecular interactions of arbuscular mycorrhizal f...	42
3	Kyoko	[illegible]	1	Adaptation as a spatiotemporal mosaic of natural a...	5
4	Marie-Eve	André	4	NULL	21
6	Marianne	[illegible]	7	Functional response of boreal vegetation to overstr...	2
7	Rodrigo	Lima Barata	1	Morphological Variation of the Red-backed Vole (My...	52
8	Magnus	Bein	1	NULL	29
9	Patrick	Bergeron	4	0	35
10	Laura	Boisvert-Marsh	1	Spatiotemporal changes at the northern limit of tr...	28
11	Johanna	Bradie	1	Predicting the establishment success of non-indige...	6
12	Dominic	Chambers	1	Modeling tree abundance in eastern North America i...	25
13	Corey	Chivers	1	Predicting spread of invasive species	6
14	Paul	Edwards	1	managing invasive species	6
15	Tammy	Elliott	1	Phylobetadiversity of sedge communities	69
16	Marie-Julie	Favé	1	The ecological and developmental genetic basis of ...	13
17	Edeline	Gagnon	2	Phylogenetic relationships and biogeography of Cae...	19
18	Melissa	Girard	7	NULL	33
19	Natalie	James	1	Modelling the effects of climate change on the dis...	28
20	Lisa	Jones	1	The role of biotic and abiotic factors in exotic s...	56
21	Joseph	Moisan de Serres	7	NULL	33
22	Maryna	Lasoway	1	Direct development with nurse eggs in the calyptra...	13
23	Chantale	Moisan	2	Impacts des activités humaines sur <i>Arethusa bul...	73
24	Georgina	O'Farrill	1	The Ecological Causes and Consequences of the Move...	37

Universities

university_institution_affiliation_id	name	FRQNT_ID
1	McGill University	44553
2	Université de Montréal	44637
3	Université du Québec à Montréal	44586
4	Université de Sherbrooke	44611
5	Université du Québec à Rimouski	[illegible]
6	Concordia University	[illegible]
7	Université Laval	44531
8	Bishop's University	[illegible]
9	Agriculture & Agroalimentaire Canada	[illegible]
10	Service canadien des forêts	[illegible]
11	Ministère du Développement durable, de l'Environne...	[illegible]
12	Yale University	[illegible]
13	La recherche agronomique pour le développement (Ci...	0
14	Duke University	[illegible]
15	Institut de recherche pour le développement (IRD)	[illegible]
16	NA	0
17	University of British Columbia	45729
18	University of Manitoba	[illegible]
19	Université du Québec à Trois-Rivières	[illegible]

Supervisors

member_id	Firstname	Lastname	University_institution_affiliation	research_en
1	Beatrix	Beisner	3	My research program is currently centred on three ...
2	Monique	Poulin	7	Wetland management, restoration and conservati...
4	Jean-Pierre	[illegible]	3	NULL
5	Andrew	Hendry	1	Darwin suggested that evolution proceeds very slow...
6	Brian	Leung	1	Biological invasions <b>Ecology of dise...
7	Claude	Lavoie	7	The Research Laboratory on Invasive Plants (REIP)...
8	Philippe	Le Prestre	7	NULL
9	Bernadette	Pinel-Alloul	2	NULL
10	Ehab	Abouheif	1	My research program focuses on the evolution of de...
11	Bernard	Angers	2	NULL
12	Graham	Bell	1	Major Research Themes <b>Adaptive radiat...
13	Elena	Bennett	1	Research in the Bennett lab centers around questio...
15	Dominique	Berteaux	5	NULL
16	Jacques	Brisson	2	NULL
17	Jacques	Brodeur	2	<b>Plant insect interactions and biological contr...
18	Luc	Brouillet	2	NULL
19	Anne	Bruneau	2	<b>Molecular systematics of legumes and the Rosa...
20	Christopher	Buddle	1	<P>Arctic Biodiversity</P> <b>Canopy Arthropod...

QCBS_students

student_id	Firstname	Lastname	University_Institution_affiliation	Project_title_en	Supervisor
1	Tyler	Bonnell	1	Spatial simulations of infectious disease: environ...	24
2	Youssef	Ismail	2	Molecular interactions of arbuscular mycorrhizal f...	42
3	Kiyoko	Gotanda	1	Adaptation as a spatiotemporal mosaic of natural a...	5
4	Marie-Eve	André	4	*NULL*	21
6	Marianne	Bachand	7	Functional response of boreal vegetation to overab...	2
7	Rodrigo	Lima Barata	1	Morphological Variation of the Red-backed Vole (My...	52
8	Magnus	Bein	1	*NULL*	29
9	Patrick	Bergeron	4	0	35
10	Laura	Boisvert-Marsh	1	Spatiotemporal changes at the northern limit of tr...	28
11	Johanna	Bradie	1	Predicting the establishment success of non-indige...	6
12	Dominic	Chambers	1	Modeling tree abundance in eastern North America i...	28
13	Corey	Chivers	1	Predicting spread of invasive species	6
14	Paul	Edwards	1	managing invasive species	6
15	Tammy	Elliot	1	Phylobetadiversity of sedge communities	65
16	Marie-Julie	Favé	1	The ecological and developmental genetic basis of ...	10
17	Edeline	Gagnon	2	Phylogenetic relationships and biogeography of Cae...	19
18	Melissa	Girard	7	*NULL*	33
19	Natalie	James	1	Modelling the effects of climate change on the dis...	28
20	Lisa	Jones	1	The role of biotic and abiotic factors in exotic s...	56
21	Joseph	Moisan de Serres	7	*NULL*	33
22	Maryna	Lesoway	1	Direct development with nurse eggs in the calyptra...	10
23	Chantale	Moisan	2	Impacts des activités humaines sur <i>Arethusa bul...	75
24	Georgina	O'Farrill	1	The Ecological Causes and Consequences of the Move...	37

university_insitution_affiliation_id	name	FRQNT_ID
1	McGill University	44563
2	Université de Montréal	44607
3	Université du Québec à Montréal	44986
4	Université de Sherbrooke	44811
5	Université du Québec à Rimouski	44988
6	Concordia University	46092
7	Université Laval	44501
8	Bishop's University	45674
9	Agriculture & Agroalimentaire Canada	45080
10	Service canadien des forêts	45809
11	Ministère du Développement durable, de l'Environne...	48535
12	Yale University	45066
13	La recherche agronomique pour le développement (Ci...	0
14	Duke University	46724
15	Institut de recherche pour le développement (IRD)	47712
16	NA	0
17	University of British Columbia	45729
18	University of Manitoba	44933
19	Université du Québec à Trois-Rivière	44993

Supervisors

member_id	Firstname	Lastname	University_Institution_affiliation	research_en
1	Beatrix	Beisner	3	My research program is currently centred on three ...
2	Monique	Poulin	7	Wetland management, restoration and conservati...
4	Jean-Pierre	Revéret	3	*NULL*
5	Andrew	Hendry	1	Darwin suggested that evolution proceeds very slow...
6	Brian	Leung	1	Biological invasions Ecology of dise...
7	Claude	Lavoie	7	The Research Laboratory on Invasive Plants (RELIP)...
8	Philippe	Le Prestre	7	*NULL*
9	Bernadette	Pinel-Alloul	2	*NULL*
10	Ehab	Abouheif	1	My research program focuses on the evolution of de...
11	Bernard	Angers	2	*NULL*
12	Graham	Bell	1	Major Research Themes: <b>Adaptive radiat...
13	Elena	Bennett	1	Research in the Bennett lab centers around questio...
15	Dominique	Berteaux	5	*NULL*
16	Jacques	Brisson	2	*NULL*
17	Jacques	Brodeur	2	Plant-insect interactions and biological contr...
18	Luc	Brouillet	2	*NULL*
19	Anne	Bruneau	2	Molecular systematics of legumes and the Rosa ...
20	Christopher	Buddle	1	Arctic Biodiversity Canopy Arthropod...

What is a relational database?

- Data is stored in formatted tables.
- Tables are linked together with "keys".

QCBS_students

student_id	Firstname	Lastname	University_institution_affiliation	Project_title_en	Supervisor
1	Tyler	Bonnell	1	Spatial simulations of infectious disease: environ...	34
2	Youssef	Ismail	2	Molecular interactions of arbuscular mycorrhizal f...	42
3	Kyoko	Gotanda	1	Adaptation as a spatiotemporal mosaic of natural a...	5
4	Marie-Eve	André	4	NULL	21
6	Marianne	Bachand	7	Functional response of boreal vegetation to overab...	2
7	Rodrigo	Lima Barata	1	Morphological Variation of the Red-backed Vole (My...	52
8	Magnus	Bein	1	NULL	29
9	Patrick	Bergeron	4	0	35
10	Laura	Boisvert-Marsh	1	Spatiotemporal changes at the northern limit of tr...	28
11	Johanna	Bradie	1	Predicting the establishment success of non-indige...	6
12	Dominic	Chambers	1	Modeling tree abundance in eastern North America i...	25
13	Corey	Chivers	1	Predicting spread of invasive species	6
14	Paul	Edwards	1	managing invasive species	6
15	Tammy	Elliot	1	Phylobetadiversity of sedge communities	63
16	Marie-Julie	Favé	1	The ecological and developmental genetic basis of ...	13
17	Edeline	Gagnon	2	Phylogenetic relationships and biogeography of Cae...	19
18	Melissa	Girard	7	NULL	33
19	Natalie	Jones	1	Modelling the effects of climate change on the dis...	28
20	Lisa	Jones	1	The role of biotic and abiotic factors in exotic s...	56
21	Joseph	Moisan de Serres	7	NULL	33
22	Maryna	Lasowy	1	Direct development with nurse eggs in the calyptra...	13
23	Chantale	Moisan	2	Impacts des activités humaines sur <i>Aethusa bul...	73
24	Georgina	O'Farrill	1	The Ecological Causes and Consequences of the Move...	37

Universities

university_institution_affiliation_id	name	FRQNT_ID
1	McGill University	44553
2	Université de Montréal	44637
3	Université du Québec à Montréal	44586
4	Université de Sherbrooke	44611
5	Université du Québec à Rimouski	44661
6	Concordia University	39497
7	Université Laval	44531
8	Bishop's University	45674
9	Agriculture & Agroalimentaire Canada	45050
10	Service canadien des forêts	45839
11	Ministère du Développement durable, de l'Environne...	45530
12	Yale University	40696
13	La recherche agronomique pour le développement (CI...	0
14	Duke University	39735
15	Institut de recherche pour le développement (IRD)	47717
16	NA	0
17	University of British Columbia	45729
18	University of Manitoba	44833
19	Université du Québec à Trois-Rivières	44695

Supervisors

member_id	Firstname	Lastname	University_institution_affiliation	research_en
1	Beatrix	Beisner	3	My research program is currently centred on three ...
2	Monique	Poulin	7	Wetland management, restoration and conservati...
4	Jean-Pierre	Revéret	3	NULL
5	Andrew	Hendry	1	Darwin suggested that evolution proceeds very slow...
6	Brian	Leung	1	Biological invasions Ecology of dise...
7	Claude	Lavoie	7	The Research Laboratory on Invasive Plants (RELIP)...
8	Philippe	Le Prestre	7	NULL
9	Bernadette	Pinel-Alloul	2	NULL
10	Ehab	Abouheif	1	My research program focuses on the evolution of de...
11	Bernard	Angers	2	NULL
12	Graham	Bell	1	Major Research Themes Adaptive radiati...
13	Elena	Bennett	1	Research in the Bennett lab centers around questio...
15	Dominique	Berteaux	5	NULL
16	Jacques	Brisson	2	NULL
17	Jacques	Brodeur	2	Plant-insect interactions and biological contr...
18	Luc	Brouillet	2	NULL
19	Anne	Bruneau	2	Molecular systematics of legumes and the Rosa...
20	Christopher	Buddle	1	Arctic Biodiversity <b>Canopy Arthropod...

What is a DBMS?

- Database Management System.
- Software.
- Controls the creation, maintenance, and use of a database.
- Allows concurrent access by several users and applications.
- Does not necessarily come with a user interface.
- RDBMS - Relational.

What is SQL?

- Structured query language.
- Often pronounced 'sequel'
- Programming language designed for RDBMS.
- Standards-based.
- Used by most RDBMS systems.
- Variations between implementations.

Options for database management

Microsoft Excel?

- Poor at working with multiple linked tables.
- Poor at working with very large datasets.
- No multi-user access.
- No strict formatting.
- No server access.
- Performing complex queries can be difficult.
- Data and 'reports' are not separated.

Microsoft Access?

- MS Access is a front-end to a database 'server' called Jet/ACE. It can be used with other database servers such as PostgreSQL.
- Not an open platform.
- Multiple versions and document formats.
- Not frequently used on servers.
- Limits to the file sizes and concurrent users.
- Can make a good front-end for designing forms or generating reports.

Full-fledged DBMS

Open source
- MySQL/MariaDB
- PostgreSQL

Proprietary
- Oracle
- Microsoft SQL Server

MicroSoft Excel?

- Poor at working with multiple linked tables.
- Poor at working with very large datasets.
- No multi-user access.
- No strict formatting.
- No server access.
- Performing complex queries can be difficult.
- Data and 'reports' are not separated.

MicroSoft Access?

- MS Access is a front-end to a database 'server' called Jet/ACE. It can be used with other database servers such as PostgreSQL.
- Not an open platform.
- Multiple versions and document formats.
- Not frequently used on servers.
- Limits to the file sizes and concurrent users.
- Can make a good front-end for designing forms or generating reports.

Full-fledged DBMS

Open source

- MySQL/MariaDB
- PostgreSQL

Proprietary

- Oracle
- Microsoft SQL Server

Benefits of open-Source Software

- Free today, free tomorrow.
- Collaborative development.
- Collaborative help system.
- Links more easily with other open source software.

- Disadvantages: not always good with user interface. No support from company.

MySQL/MariaDB and PostgreSQL

- Open source.
- Extremely powerful.
- Fast.
- Can handle very large datasets.
- Good development.
- Used by major companies (Facebook, Twitter, Google, etc.).
- SQL standard.
- MySQL - developed mostly by Oracle.
- Postgres and MariaDB - community development.
- Database servers with no backends.

When Should I use a DBMS?

- Standardization of data is important
- Very large datasets
- Multiple users, multiple platforms
- Access on a server
- Interaction with other software or computer tools

When I Should not use a DBMS?

- Small datasets that don't need standardization
- Single users, simple needs
- Lack of technical expertise/time

Examples of DBMS use

- QCBS student/prof database and website.
- Des nids chez vous website
- The internet (e.g. http://quebio.ca/bam)
- Collaborative databases (GBIF, GENBANK, etc.)

Examples of DBMS use

- QCBS student/prof database and website.
- Des nids chez vous website
- The internet (e.g. http://quebio.ca/bam)
- Collaborative databases (GBIF, GENBANK, etc.)

Database design principles

Basic principles

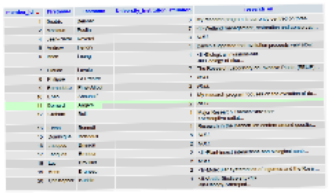
- Standardization
- Coherence
- Expandability
- Flexibility

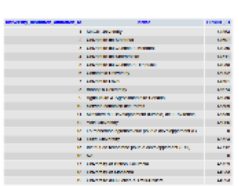
Steps of database design

- Determine entities and tables
- Relationships/cardinality
- Establish primary keys
- Establish data types

Determine entities and tables

- One table per type of entity
- Short but meaningful table and column names. No spaces (use _), no strange characters (accented, ?&%$@'(",)





student_id	Firstname	Lastname	University_Institution_affiliation	Project_title_en	Supervisor
1	Tyler	Bonnell	1	Spatial simulations of infectious disease: environ...	24
2	Youssef	Ismail	2	Molecular interactions of arbuscular mycorrhizal f...	42
3	Kiyoko	Gotanda	1	Adaptation as a spatiotemporal mosaic of natural a...	5
4	Marie-Eve	André	4	*NULL*	21
6	Marianne	Bachand	7	Functional response of boreal vegetation to overab...	2
7	Rodrigo	Lima Barata	1	Morphological Variation of the Red-backed Vole (My...	52
8	Magnus	Bein	1	*NULL*	29
9	Patrick	Bergeron	4	0	35
10	Laura	Boisvert-Marsh	1	Spatiotemporal changes at the northern limit of tr...	28
11	Johanna	Bradie	1	Predicting the establishment success of non-indige...	6
12	Dominic	Chambers	1	Modeling tree abundance in eastern North America i...	28
13	Corey	Chivers	1	Predicting spread of invasive species	6
14	Paul	Edwards	1	managing invasive species	6
15	Tammy	Elliot	1	Phylobetadiversity of sedge communities	65
16	Marie-Julie	Favé	1	The ecological and developmental genetic basis of ...	10
17	Edeline	Gagnon	2	Phylogenetic relationships and biogeography of Cae...	19
18	Melissa	Girard	7	*NULL*	33
19	Natalie	James	1	Modelling the effects of climate change on the dis...	28
20	Lisa	Jones	1	The role of biotic and abiotic factors in exotic s...	56
21	Joseph	Moisan de Serres	7	*NULL*	33
22	Maryna	Lesoway	1	Direct development with nurse eggs in the calyptra...	10
23	Chantale	Moisan	2	Impacts des activités humaines sur <i>Arethusa bul...	75
24	Georgina	O'Farrill	1	The Ecological Causes and Consequences of the Move...	37

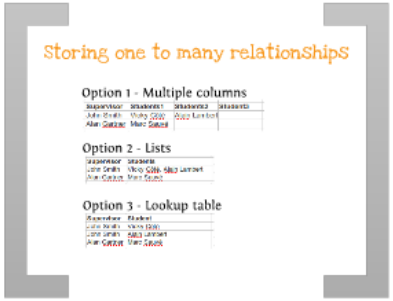
member_id	Firstname	Lastname	University_Institution_affiliation	research_en
1	Beatrix	Beisner	3	My research program is currently centred on three ...
2	Monique	Poulin	7	Wetland management, restoration and conservati...
4	Jean-Pierre	Revéret	3	*NULL*
5	Andrew	Hendry	1	Darwin suggested that evolution proceeds very slow...
6	Brian	Leung	1	Biological invasions Ecology of dise...
7	Claude	Lavoie	7	The Research Laboratory on Invasive Plants (RELIP)...
8	Philippe	Le Prestre	7	*NULL*
9	Bernadette	Pinel-Alloul	2	*NULL*
10	Ehab	Abouheif	1	My research program focuses on the evolution of de...
11	Bernard	Angers	2	*NULL*
12	Graham	Bell	1	Major Research Themes: <b>Adaptive radiat...
13	Elena	Bennett	1	Research in the Bennett lab centers around questio...
15	Dominique	Berteaux	5	*NULL*
16	Jacques	Brisson	2	*NULL*
17	Jacques	Brodeur	2	Plant-insect interactions and biological contr...
18	Luc	Brouillet	2	*NULL*
19	Anne	Bruneau	2	Molecular systematics of legumes and the Rosa ...
20	Christopher	Buddle	1	Arctic Biodiversity Canopy Arthropod...

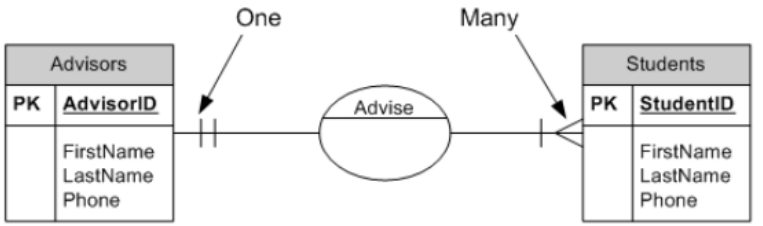
university_insitution_affiliation_id	name	FRQNT_ID
1	McGill University	44563
2	Université de Montréal	44607
3	Université du Québec à Montréal	44986
4	Université de Sherbrooke	44811
5	Université du Québec à Rimouski	44988
6	Concordia University	46092
7	Université Laval	44501
8	Bishop's University	45674
9	Agriculture & Agroalimentaire Canada	45080
10	Service canadien des forêts	45809
11	Ministère du Développement durable, de l'Environne...	48535
12	Yale University	45066
13	La recherche agronomique pour le développement (Ci...	0
14	Duke University	46724
15	Institut de recherche pour le développement (IRD)	47712
16	NA	0
17	University of British Columbia	45729
18	University of Manitoba	44933
19	Université du Québec à Trois-Rivière	44993

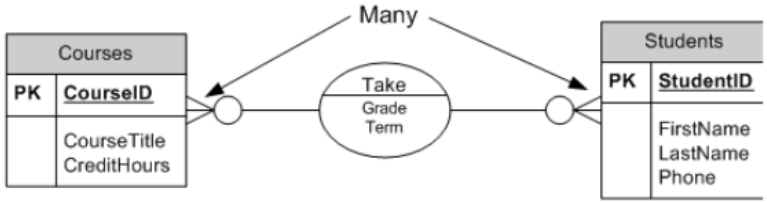
keyword_id	keyword
1	biodiversity
2	ecology
3	ecosystem
4	ecosystem services
5	population ecology
6	genetics
7	ecosystem functioning
8	forest ecology
9	phylogenetics
10	population genetics
11	geographic information systems
12	generalized linear models
13	limnology
14	forestry
15	bayesian analysis
16	spatial statistics

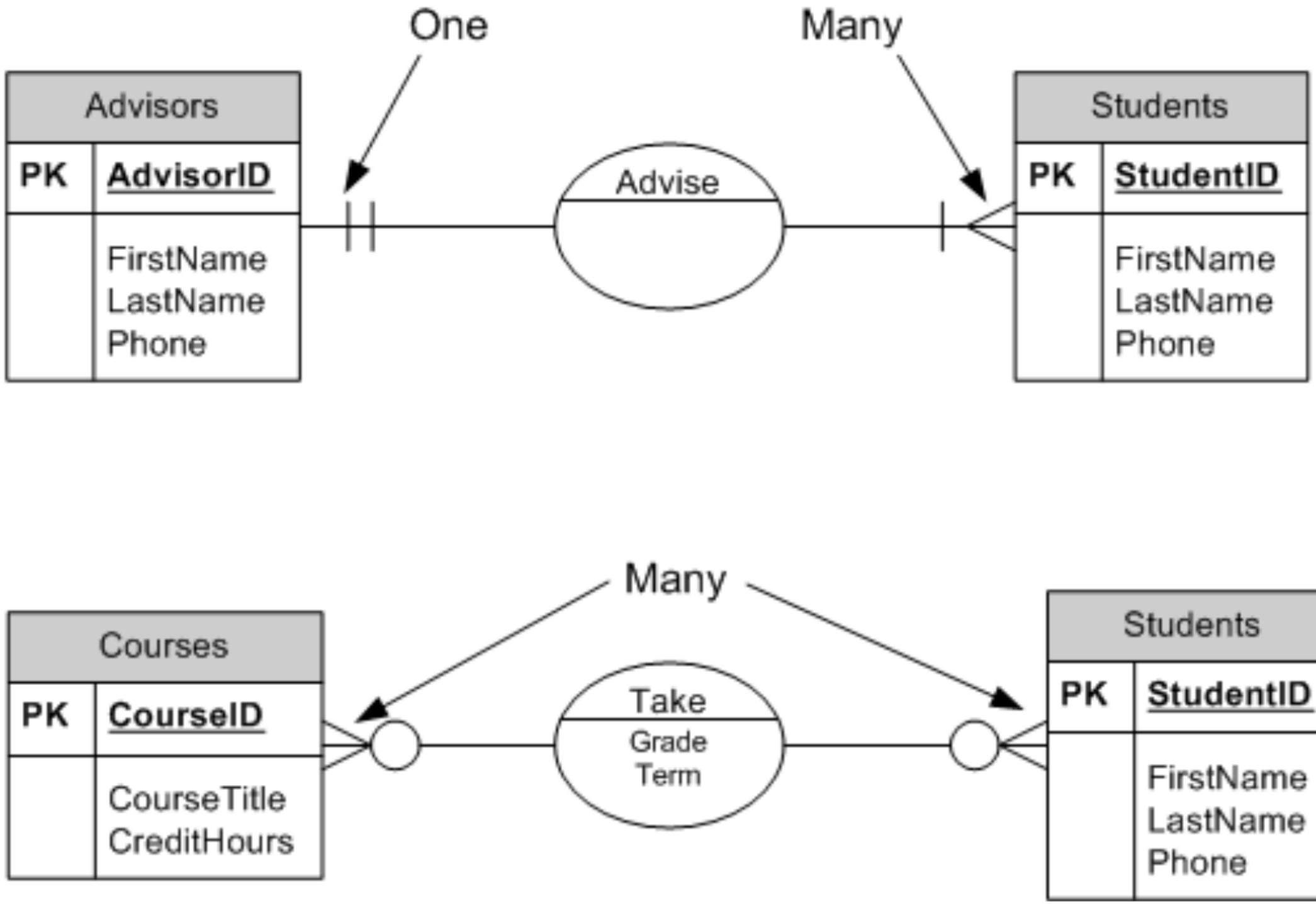
Establish relationships/ cardinality

- One to one
- One to many
- Many to many








One
Many
Advisors
PK
AdvisorID
FirstName
LastName
Phone
Advise
Students
PK
StudentID
FirstName
LastName
Phone
Many
Courses
PK
CourseID
CourseTitle
CreditHours
Take
Grade
Term
Students
PK
StudentID
FirstName
LastName
Phone

Storing one to many relationships

Option 1 - Multiple columns

Supervisor	Students1	Students2	Student3
John Smith	Vicky Côté	Alain Lambert	
Alan Gartner	Marc Sauvé		

Option 2 - Lists

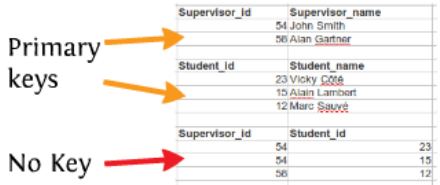
Supervisor	Students
John Smith	Vicky Côté, Alain Lambert
Alan Gartner	Marc Sauvé

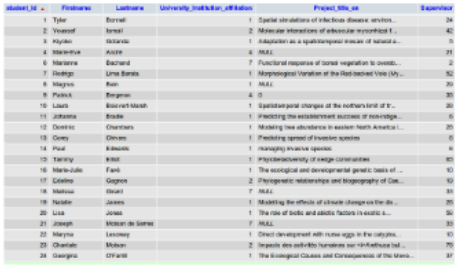
Option 3 - Lookup table

Supervisor	Student
John Smith	Vicky Côté
John Smith	Alain Lambert
Alan Gartner	Marc Sauvé

Establish primary keys

- Mechanism used to relate tables together
- One entity = one unique key
- Unique columns, combination of columns or logical primary key





primary ke

Primary keys

No Key

Supervisor_id	Supervisor_name
54	John Smith
58	Alan Gartner
Student_id	**Student_name**
23	Vicky Côté
15	Alain Lambert
12	Marc Sauvé
Supervisor_id	**Student_id**
54	23
54	15
58	12

student_id	Firstname	Lastname	University_Institution_affiliation	Project_title_en	Supervisor
1	Tyler	Bonnell	1	Spatial simulations of infectious disease: environ...	24
2	Youssef	Ismail	2	Molecular interactions of arbuscular mycorrhizal f...	42
3	Kiyoko	Gotanda	1	Adaptation as a spatiotemporal mosaic of natural a...	5
4	Marie-Eve	André	4	*NULL*	21
6	Marianne	Bachand	7	Functional response of boreal vegetation to overab...	2
7	Rodrigo	Lima Barata	1	Morphological Variation of the Red-backed Vole (My...	52
8	Magnus	Bein	1	*NULL*	29
9	Patrick	Bergeron	4	0	35
10	Laura	Boisvert-Marsh	1	Spatiotemporal changes at the northern limit of tr...	28
11	Johanna	Bradie	1	Predicting the establishment success of non-indige...	6
12	Dominic	Chambers	1	Modeling tree abundance in eastern North America i...	28
13	Corey	Chivers	1	Predicting spread of invasive species	6
14	Paul	Edwards	1	managing invasive species	6
15	Tammy	Elliot	1	Phylobetadiversity of sedge communities	65
16	Marie-Julie	Favé	1	The ecological and developmental genetic basis of ...	10
17	Edeline	Gagnon	2	Phylogenetic relationships and biogeography of Cae...	19
18	Melissa	Girard	7	*NULL*	33
19	Natalie	James	1	Modelling the effects of climate change on the dis...	28
20	Lisa	Jones	1	The role of biotic and abiotic factors in exotic s...	56
21	Joseph	Moisan de Serres	7	*NULL*	33
22	Maryna	Lesoway	1	Direct development with nurse eggs in the calyptra...	10
23	Chantale	Moisan	2	Impacts des activités humaines sur <i>Arethusa bul...	75
24	Georgina	O'Farrill	1	The Ecological Causes and Consequences of the Move...	37

Establish data types

- Has to allow all possible current and future values.

String types

CHAR( ) A fixed section from 0 to 255 characters long.
VARCHAR( ) A variable section from 0 to 255 characters long.
TINYTEXT A string with a maximum length of 255 characters.
TEXT A string with a maximum length of 65535 characters.
BLOB A string with a maximum length of 65535 characters.
MEDIUMTEXT A string with a maximum length of 16777215 characters.
MEDIUMBLOB A string with a maximum length of 16777215 characters.
LONGTEXT A string with a maximum length of 4294967295 characters.
LONGBLOB A string with a maximum length of 4294967295 characters.

Numeric types

TINYINT( ) -128 to 127 normal 0 to 255 UNSIGNED.
SMALLINT( ) -32768 to 32767 normal 0 to 65535 UNSIGNED.
MEDIUMINT( ) -8388608 to 8388607 normal 0 to 16777215 UNSIGNED.
INT( ) -2147483648 to 2147483647 normal 0 to 4294967295 UNSIGNED.
BIGINT( ) -9223372036854775808 to 9223372036854775807 normal 0 to 18446744073709551615 UNSIGNED.
FLOAT A small number with a floating decimal point.
DOUBLE(,) A large number with a floating decimal point.
DECIMAL(,) A DOUBLE stored as a string , allowing for a fixed decimal point.

Date-time

DATE YYYY-MM-DD.
DATETIME YYYY-MM-DD HH:MM:SS.
TIMESTAMP YYYYMMDDHHMMSS.
TIME HH:MM:SS.

NULL

Used for unknown attributes

String types

CHAR( ) A fixed section from 0 to 255 characters long.
VARCHAR( ) A variable section from 0 to 255 characters long.
TINYTEXT A string with a maximum length of 255 characters.
TEXT A string with a maximum length of 65535 characters.
BLOB A string with a maximum length of 65535 characters.
MEDIUMTEXT A string with a maximum length of 16777215 characters.
MEDIUMBLOB A string with a maximum length of 16777215 characters.
LONGTEXT A string with a maximum length of 4294967295 characters.
LONGBLOB A string with a maximum length of 4294967295 characters.

Date-time

DATE YYYY-MM-DD.

t values.

Numeric types

TINYINT( ) -128 to 127 normal 0 to 255 UNSIGNED.
SMALLINT( ) -32768 to 32767 normal 0 to 65535 UNSIGNED.
MEDIUMINT( ) -8388608 to 8388607 normal 0 to 16777215 UNSIGNED.
INT( ) -2147483648 to 2147483647 normal 0 to 4294967295 UNSIGNED.
BIGINT( ) -9223372036854775808 to 9223372036854775807 normal
0 to 18446744073709551615 UNSIGNED.
FLOAT A small number with a floating decimal point.
DOUBLE(,) A large number with a floating decimal point.
DECIMAL(,) A DOUBLE stored as a string , allowing for a fixed decimal point.

NULL

Used for unknown attributes

Date-time

DATE YYYY-MM-DD.

DATETIME YYYY-MM-DD HH:MM:SS.

TIMESTAMP YYYYMMDDHHMMSS.

TIME HH:MM:SS.

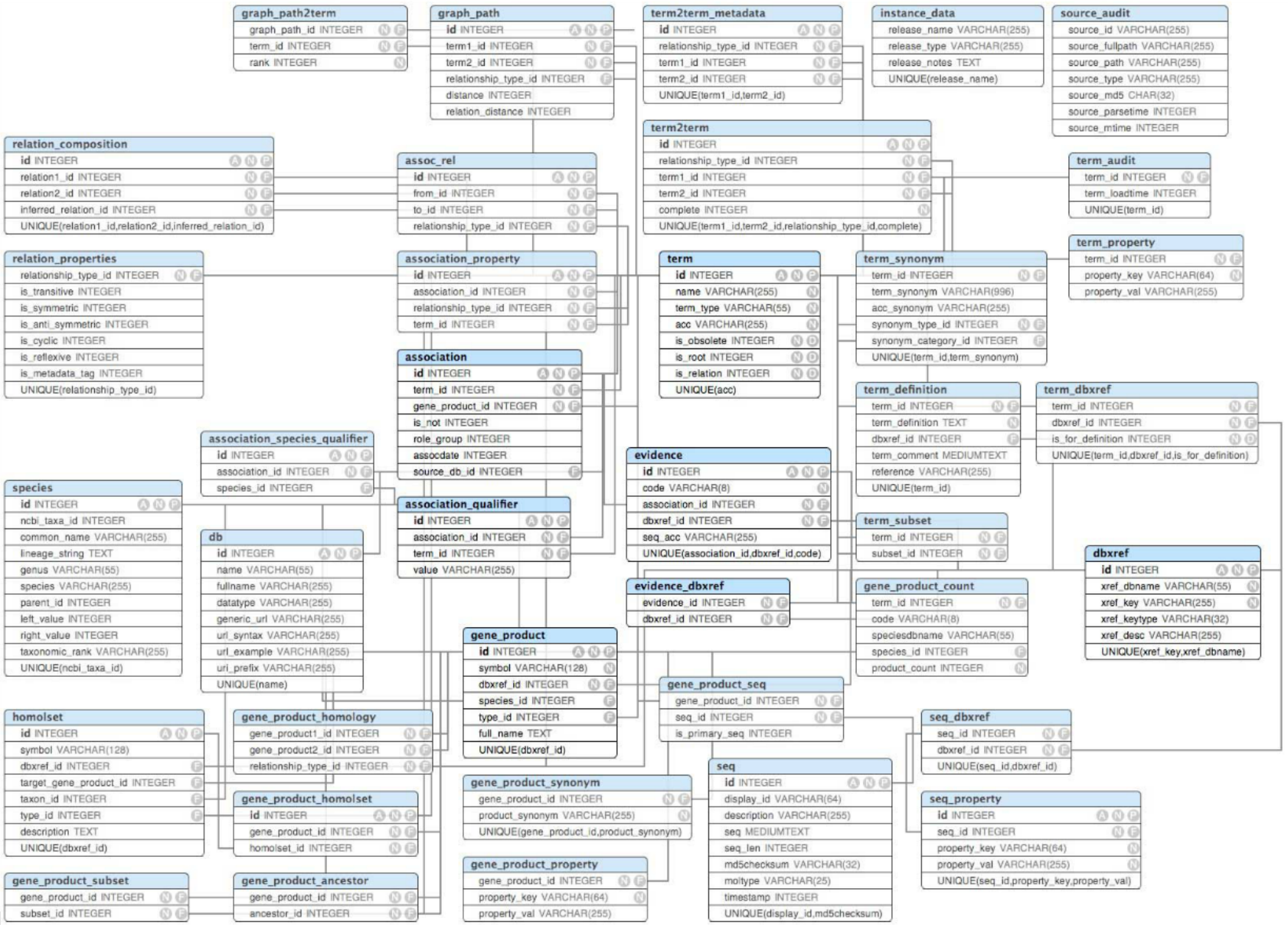
NULL

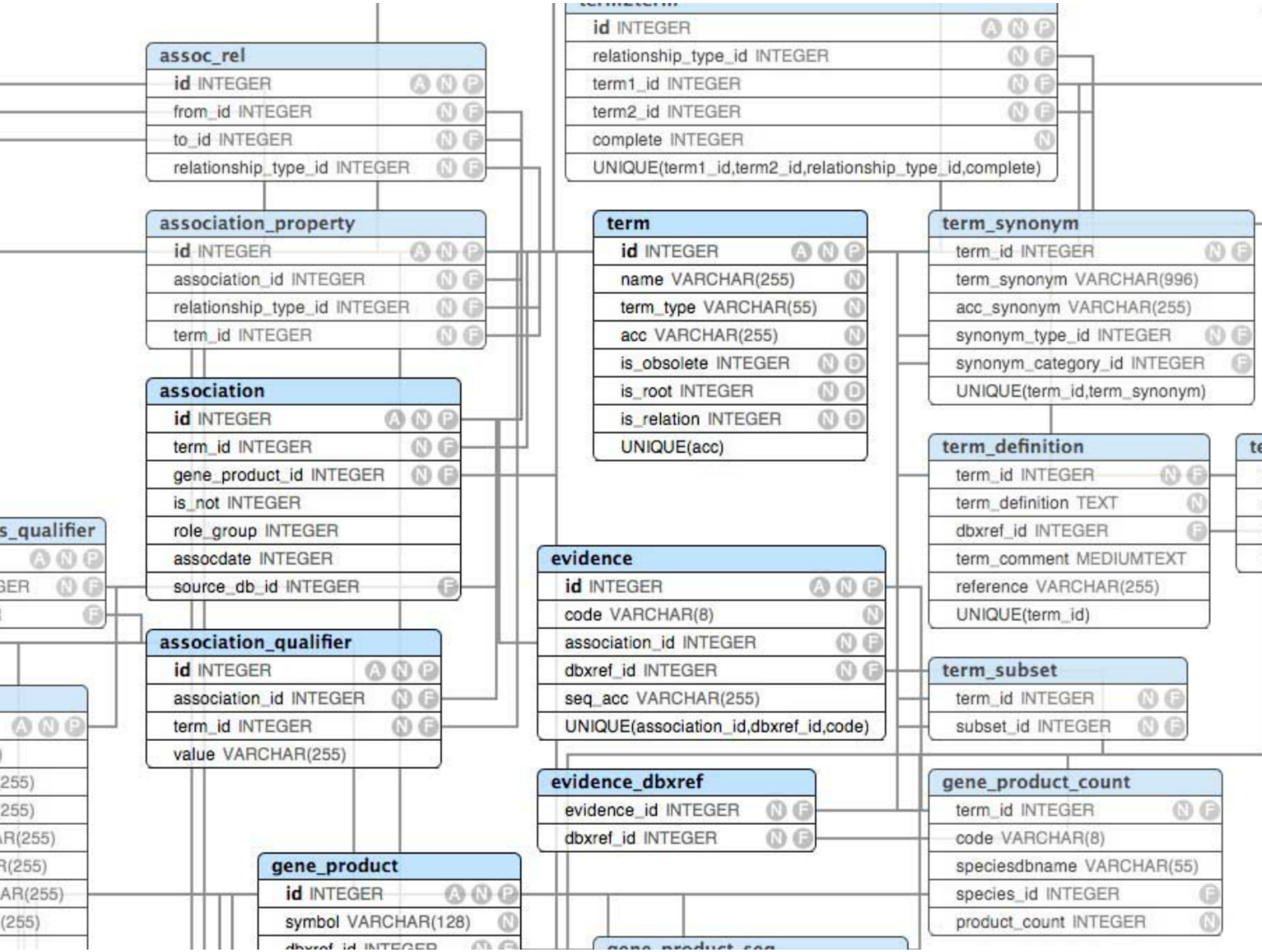
Used for unknown attributes

Rules of thumb

- One attribute per column. Cannot be subdivided.
- Unique way to identify each row.
- No repeating info.
- Simple but meaningful table and column names. Avoid words such as 'name', 'text','count','long', etc.

Entity-relationship diagram




assoc_rel
id INTEGER
from_id INTEGER
to_id INTEGER
relationship_type_id INTEGER
id INTEGER
relationship_type_id INTEGER
term1_id INTEGER
term2_id INTEGER
complete INTEGER
UNIQUE(term1_id,term2_id,relationship_type_id,complete)
association_property
id INTEGER
association_id INTEGER
relationship_type_id INTEGER
term_id INTEGER
term
id INTEGER
name VARCHAR(255)
term_type VARCHAR(55)
acc VARCHAR(255)
is_obsolete INTEGER
is_root INTEGER
is_relation INTEGER
UNIQUE(acc)
term_synonym
term_id INTEGER
term_synonym VARCHAR(996)
acc_synonym VARCHAR(255)
synonym_type_id INTEGER
synonym_category_id INTEGER
UNIQUE(term_id,term_synonym)
association
id INTEGER
term_id INTEGER
gene_product_id INTEGER
is_not INTEGER
role_group INTEGER
assocdate INTEGER
source_db_id INTEGER
term_definition
term_id INTEGER
term_definition TEXT
dbxref_id INTEGER
term_comment MEDIUMTEXT
reference VARCHAR(255)
UNIQUE(term_id)
evidence
id INTEGER
code VARCHAR(8)
association_id INTEGER
dbxref_id INTEGER
seq_acc VARCHAR(255)
UNIQUE(association_id,dbxref_id,code)
association_qualifier
id INTEGER
association_id INTEGER
term_id INTEGER
value VARCHAR(255)
term_subset
term_id INTEGER
subset_id INTEGER
evidence_dbxref
evidence_id INTEGER
dbxref_id INTEGER
gene_product_count
term_id INTEGER
code VARCHAR(8)
speciesdbname VARCHAR(55)
species_id INTEGER
product_count INTEGER
gene_product
id INTEGER
symbol VARCHAR(128)

Basics of the PostgreSQL language

User interface

- Command line client
- LibreOffice BASE/ MS Access
- PgAdmin III
- Phppgadmin

Entering data...

- Create forms (Access, LibreOffice, PHP)
- Edit directly in user interface (e.g. Libreoffice Base)
- INSERT command

Importing data...

- From LibreOffice/Access (csv,xls,ods)
- From the PostgreSQL command line (CREATE TABLE, COPY...)

Basic Commands

Basics

- Table, column names separated by commas, not commands
- Strings are separated by ''
- Names can be enclosed with "" when spaces or special characters are present or for reserved words.
- Not case-sensitive. But always use good case.

USer interface

- Command line client
- LibreOffice BASE/ MS Access
- PgAdmin III
- Phppgadmin

Basics

- Table, column names separated by commas, not commands
- Strings are separated by ''
- Names can be enclosed with "" when spaces or special characters are present or for reserved words.
- Not case-sensitive. But always use good case.

Entering data...

- Create forms (Access, LibreOffice, PHP)
- Edit directly in user interface (e.g. Libreoffice Base)
- INSERT command

Importing data...

- From Libreoffice/Access (csv,xls,ods)
- From the PostgreSQL command line (CREATE TABLE, COPY...)

Basic commands

\l list tables in database
\d+ describe table and columns
CREATE create database or table
INSERT insert line(s) into table
SELECT make queries
UPDATE modify column
DELETE delete lines
ALTER add columns or modify format
DROP delete table or database

Operators

SELECT columns FROM tables WHERE conditions
JOIN ... GROUP BY columns HAVING condition
ORDER BY columns

Basic commands

\l list tables in database
\d+ describe table and columns
CREATE create database or table
INSERT insert line(s) into table
SELECT make queries
UPDATE modify column
DELETE delete lines
ALTER add columns or modify format
DROP delete table or database

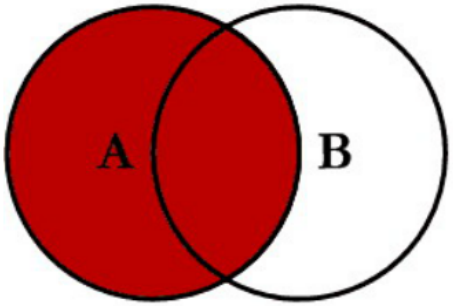
Operators

DELETE delete lines
ALTER add columns or modify format
DROP delete table or database

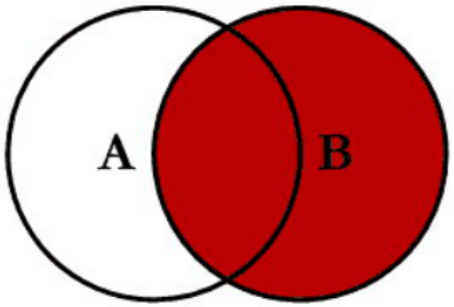
Operators

SELECT columns FROM tables WHERE conditions
JOIN ... GROUP BY columns HAVING condition
ORDER BY columns

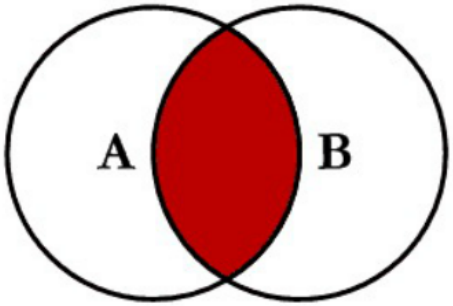
SQL JOINS



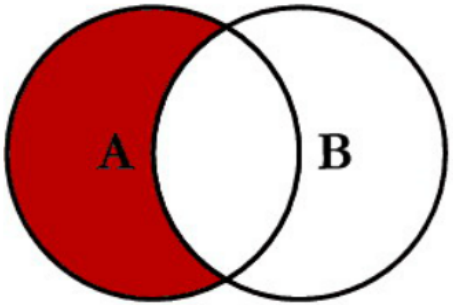

```
SELECT <select_list>
FROM TableA A
LEFT JOIN TableB B
ON A.Key = B.Key
```



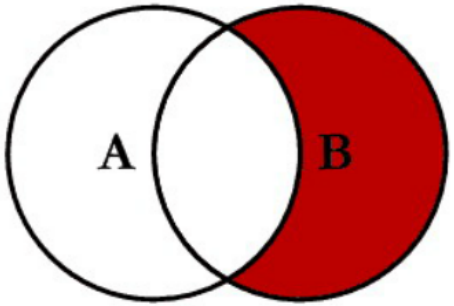

```
SELECT <select_list>
FROM TableA A
RIGHT JOIN TableB B
ON A.Key = B.Key
```



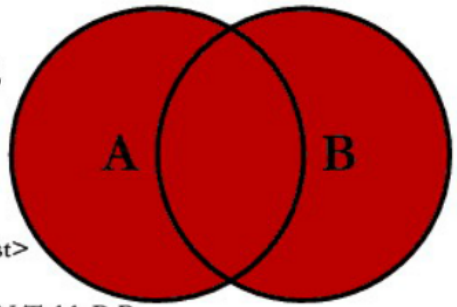

```
SELECT <select_list>
FROM TableA A
INNER JOIN TableB B
ON A.Key = B.Key
```



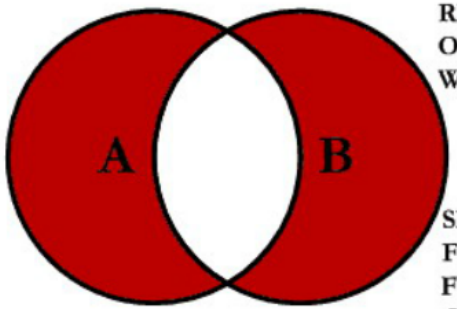

```
SELECT <select_list>
FROM TableA A
LEFT JOIN TableB B
ON A.Key = B.Key
WHERE B.Key IS NULL
```



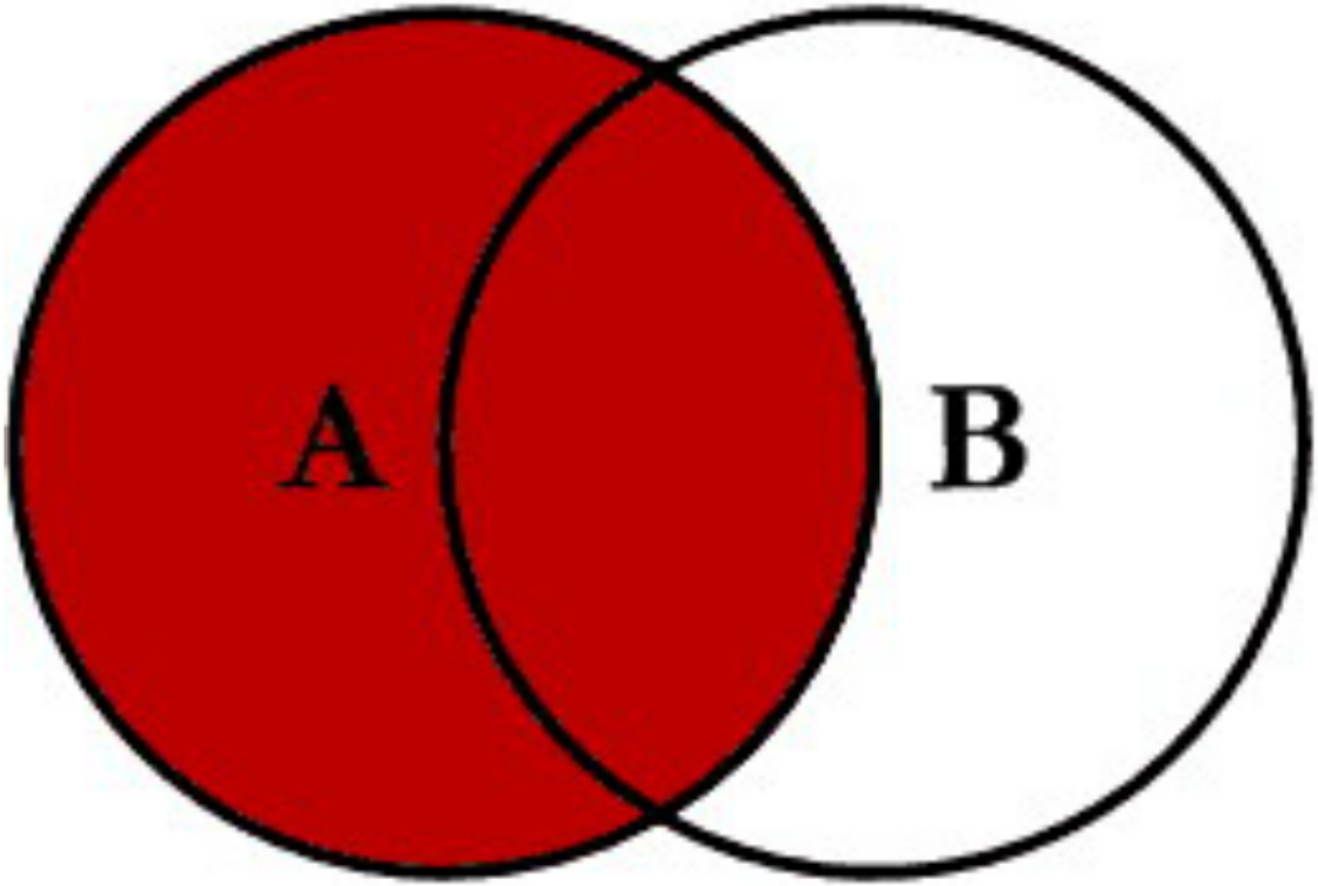

```
SELECT <select_list>
FROM TableA A
RIGHT JOIN TableB B
ON A.Key = B.Key
WHERE A.Key IS NULL
```



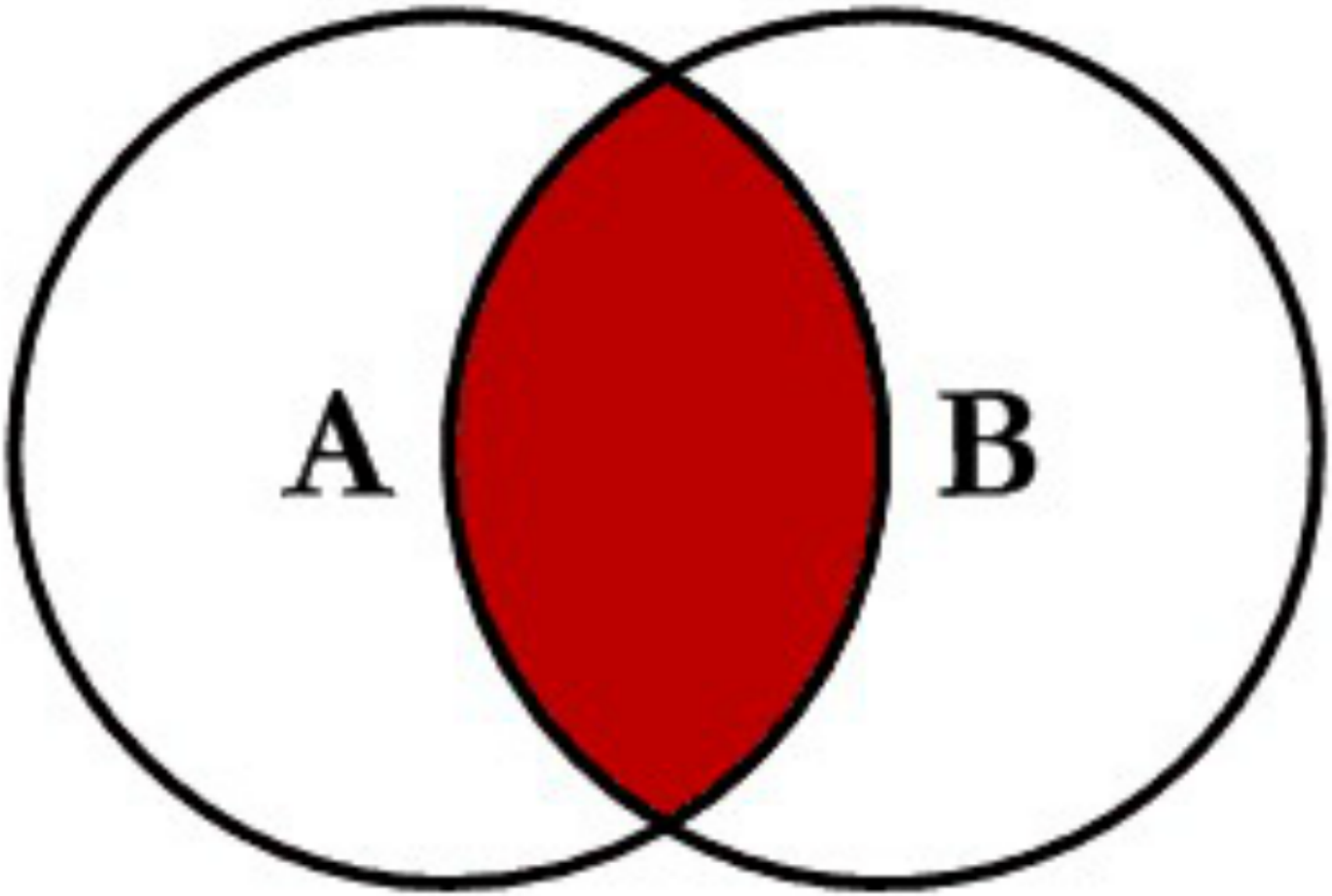

```
SELECT <select_list>
FROM TableA A
FULL OUTER JOIN TableB B
ON A.Key = B.Key
```




```
SELECT <select_list>
FROM TableA A
FULL OUTER JOIN TableB B
ON A.Key = B.Key
WHERE A.Key IS NULL
OR B.Key IS NULL
```

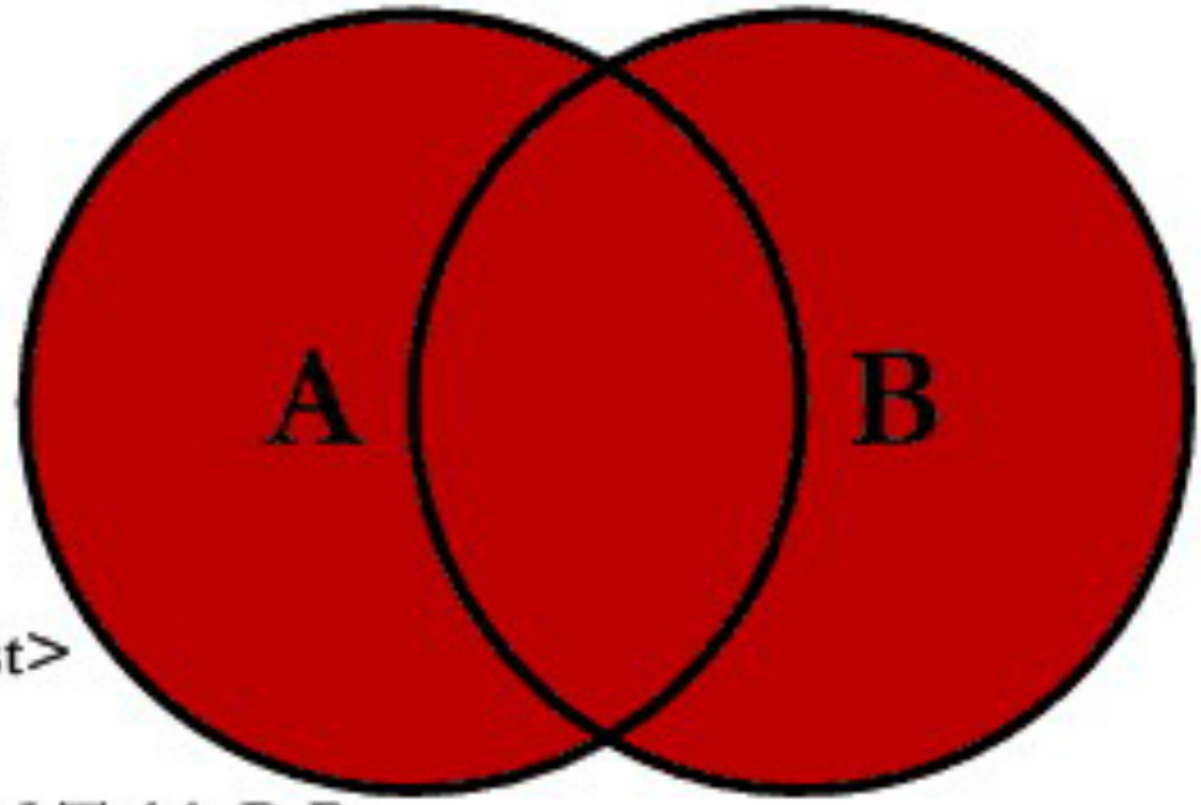



```
SELECT <select_list>
FROM TableA A
LEFT JOIN TableB B
ON A.Key = B.Key
```




```
SELECT <select_list>
FROM TableA A
INNER JOIN TableB B
ON A.Key = B.Key
```

```
SELECT <select_list>
FROM TableA A
LEFT JOIN TableB B
ON A.Key = B.Key
WHERE B.Key IS NULL
```




```
SELECT <select_list>
FROM TableA A
FULL OUTER JOIN TableB B
ON A.Key = B.Key
```

Introduction to database management with open source tools

Guillaume Larocque
research professional,
Quebec Center for Biodiversity Science

http://qcbs.ca/wiki/opendb

CENTRE DE LA SCIENCE DE LA BIODIVERSITÉ DU QUÉBEC
QUEBEC CENTRE FOR BIODIVERSITY SCIENCE

PostgreSQL
the worlds most advanced open source database

LibreOffice
The Document Foundation