

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



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



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".



ID	NAME	AGE
1	John	25
2	Jane	30
3	Bob	22

ID	NAME	AGE
1	John	25
2	Jane	30
3	Bob	22

ID	NAME	AGE
1	John	25
2	Jane	30
3	Bob	22

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.

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 in	Supervisor
1	Tyler	Reynold	1	Spittle: Simulations of infectious disease systems...	24
2	Thibault	Jean-L	2	Molecular interactions of adenoviral mycoplasma f...	42
3	Kivler	Clara	1	Adaptation as a spatiotemporal result of natural s...	5
4	Mahdavi	Amir	4	AMAL	21
4	Mahdavi	Hassan	7	Functional responses of insect vegetation to insects...	2
7	Mejias	Lina Beatriz	1	Microbiological Variation of the Red-bellied Vole (S...	32
8	Meyers	Don	1	AMAL	29
9	Witacka	Bergeron	4	W	35
10	Laine	Richard Mark	1	Questioning changes at the northern limit of t...	24
11	Johnson	Brenda	1	Phenology for establishment success of nonindige...	6
12	Cornejo	Charmelo	1	Modeling tree abundance in eastern North America...	28
13	Gray	Chris	1	Phenology spread of invasive species	4
14	Paul	Edward	1	managing invasive species	4
18	Taney	Chad	1	Phylogenetic diversity of edge communities	63
18	Maria-Julia	Felix	1	The ecological and developmental genetic basis of...	19
19	Schmitt	Gilbert	2	Phylogenetic relationships and biogeography of Co...	18
19	Melissa	Gilbert	7	AMAL	23
19	Nabali	James	1	Modeling the effects of climate change on the di...	28
20	Uma	John	1	The role of biotic and abiotic factors in exotic s...	86
20	Joseph	Robert de Saint	7	AMAL	33
22	Melissa	Lacey	1	Oral development with house eggs in the scypha...	19
23	Christine	Melissa	2	Impacto des atividades humanas nos rios da bacia h...	29
24	Georgina	Ortiz	1	The Ecological Causes and Consequences of the M...	37

Universities

University Institution affiliation id	name	PROJECT_ID
1	NOVA University	4422
2	Université de Montréal	4427
3	Université de Québec en Maritimes	4426
4	Université de Sherbrooke	4411
5	Université de la Gaspésie (Université)	4444
6	Université Laval	4447
7	Université Laval	4521
8	Stamps University	4521
9	Agriculture & Agri-Food Canada	4520
10	Science Canada (des Instituts)	4522
11	Ministère du Développement durable, de l'Énergie... (des Instituts)	4523
12	York University	4596
13	Université d'agriculture pour le développement (CA)	0
14	Université de Sherbrooke	3005
15	Université de Moncton pour le développement (CA)	3707
16	UQ	0
17	University of British Columbia	4524
18	University of Manitoba	4425
19	Université de Québec à Trois-Rivières	4423

Supervisors

member id	Firstname	Lastname	University Institution affiliation	research in
1	Benoit	Benoit	2	My second program is currently under review...
2	Monica	Paul	7	in Web-based management, natural resource conserv...
4	Jean-Marc	Reynold	3	NAL
5	Andrew	Hardy	1	Danah suggested that inclusion proceed very slow...
6	Wesley	Leung	1	in Ecological restoration in Ecology of d...
7	Clara	Laurin	7	The Research Laboratory on Invasive Plants (RLIP)...
8	William	Le-Py	7	NAL
8	Demetrio	Franklin	2	NAL
10	John	Johnson	1	My second program focuses on the evolution of di...
11	Benoit	Anges	2	NAL
13	Griffin	Neil	1	major research themes in Adaptive net...
13	Clara	Benoit	1	Research in the Forest: its carries around quarr...
13	Christian	Berthelet	5	NAL
15	Nicolas	Benoit	2	NAL
17	Joseph	Benoit	7	in Web-based management, natural resource conserv...
18	Ugo	Benoit	2	NAL
18	Anne	Benoit	2	in Molecular systematics of legume and the flora...
20	Christopher	Benoit	1	in Web-based management, natural resource conserv...

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: 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 in	Supervisor
1	Tyler	Reynold	1	Spittle: Simulations of infectious disease systems...	24
2	Thibault	Jamal	2	Molecular interactions of adenoviral mycoplasma f...	42
3	Kylee	Colombo	1	Adaptation as a spatiotemporal result of natural s...	5
4	Melanie	Avant	4	AMAL	21
4	Melanie	Harrison	7	Functional responses of insect vegetation to insects...	2
7	Marissa	Lima Bevilacqua	1	Microbiological Variation of the Red-bellied Vole (P...	32
8	Meghan	Doh	1	AMAL	29
9	Malika	Bejerman	4	W	35
10	Liam	McConnell-Mark	1	Questionnaire changes at the northern limit of t...	24
11	Shanna	Brady	1	Phenology for establishment success of nonindige...	6
12	Danielle	Chambers	1	Modeling tree abundance in eastern North America...	28
13	Greg	Chivers	1	Phenology spread of invasive species	4
14	Paul	Edwards	1	managing invasive species	4
18	Tammy	Choi	1	Phylogenetic diversity of edge communities	63
18	Maria-Julia	Favre	1	The ecological and developmental genetic basis of...	19
19	Scottie	Gagnon	2	Phylogenetic relationships and biogeography of Co...	19
19	Melissa	Gillett	7	AMAL	23
19	Natalie	James	1	Modeling the effects of climate change on the di...	28
20	Uma	Johns	1	The role of biotic and abiotic factors in exotic s...	86
20	Joseph	Holmes de Souza	7	AMAL	33
22	Melissa	Lacey	1	Oral development with house eggs in the scypha...	19
23	Christine	Molloy	2	Impacto des atividades humanas nos rios da bacia h...	29
24	Georgina	O'Hara	1	The Ecological Causes and Consequences of the M...	37

Universities

University Institution affiliation id	name	PROJECT_ID
1	NOVA University	4422
2	Université de Montréal	4427
3	Université de Québec à Montréal	4426
4	Université de Sherbrooke	4411
5	Université de Guelph	4444
6	University of Toronto	4447
7	University of York	4521
8	Simon Fraser University	4523
9	Agriculture & Agri-Food Canada	4520
10	Science Canada/Desjardins	4522
11	Ministry of Development, Studies, and Technol...	4629
12	York University	4696
13	Université d'agriculture pour le développement (U...	0
14	Université de Sherbrooke	3005
15	Université de Moncton pour le développement (U...	3707
16	UQ	0
17	University of British Columbia	4524
18	University of Manitoba	4425
19	Université de Québec à Trois-Rivières	4423

Supervisors

member id	Firstname	Lastname	University Institution affiliation	research in
1	David	Besser	2	My second program is currently under review...
2	Monica	Paulin	7	in Web-based management, natural resource conserv...
4	Jean-Marc	Reynold	3	NAL
5	Andrew	Hardy	1	David suggested that evolution proceeds very slow...
6	Wesley	Leung	1	in Ecological invasions in Ecology of d...
7	Clare	Lewis	7	The Research Laboratory on Invasive Plants (RLIP)...
8	William	Le-Pyost	7	NAL
8	Demetris	Frankel	2	NAL
10	John	Johnson	1	My second program focuses on the evolution of d...
11	Bernad	Anger	2	NAL
12	Griffin	Reid	1	major research themes in Adaptive net...
13	Clare	Bennett	1	Research in the Desert: its centres around quads...
13	Christophe	Berthouex	5	NAL
15	Nicolas	Brechet	2	NAL
17	Joseph	Brisset	7	in Water input interactions and nitrogen cover...
18	Luca	Brudini	2	NAL
18	Anne	Bruneau	2	in Molecular systematics of legume and the flow...
20	Christopher	Budde	1	in Arctic Biodiversity in Diversity Assessment

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.



PostgreSQL



MariaDB

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
- Relationship cardinality
- Establish primary key
- Establish data types

Rules of thumb

- One attribute per column. Create for
- "atomic"
- Unique key to identify each row
- No repeating tags
- Define and normalized table and
- Define name, field width, field no.
- Name: table, column, group, etc.

Database design

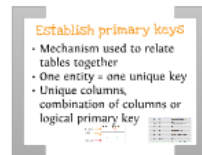
Database design is the process of creating a logical data model for a database. It involves defining the structure of the database, including tables, columns, and relationships. The design process typically starts with a conceptual data model, which is then translated into a logical data model. The logical data model is then implemented as a physical database. The design process is iterative and involves refining the design as more information is gathered.

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, ?&%\$@'(",)

Author ID	Author	Category	Category ID	Book ID	Book
1	John	Science	1	1	1. The Structure of Matter
2	John	Science	1	2	2. The Structure of Matter
3	John	Science	1	3	3. The Structure of Matter
4	John	Science	1	4	4. The Structure of Matter
5	John	Science	1	5	5. The Structure of Matter
6	John	Science	1	6	6. The Structure of Matter
7	John	Science	1	7	7. The Structure of Matter
8	John	Science	1	8	8. The Structure of Matter
9	John	Science	1	9	9. The Structure of Matter
10	John	Science	1	10	10. The Structure of Matter
11	John	Science	1	11	11. The Structure of Matter
12	John	Science	1	12	12. The Structure of Matter
13	John	Science	1	13	13. The Structure of Matter
14	John	Science	1	14	14. The Structure of Matter
15	John	Science	1	15	15. The Structure of Matter
16	John	Science	1	16	16. The Structure of Matter
17	John	Science	1	17	17. The Structure of Matter
18	John	Science	1	18	18. The Structure of Matter
19	John	Science	1	19	19. The Structure of Matter
20	John	Science	1	20	20. The Structure of Matter

Author ID	Author	Category	Category ID	Book ID	Book
1	John	Science	1	1	1. The Structure of Matter
2	John	Science	1	2	2. The Structure of Matter
3	John	Science	1	3	3. The Structure of Matter
4	John	Science	1	4	4. The Structure of Matter
5	John	Science	1	5	5. The Structure of Matter
6	John	Science	1	6	6. The Structure of Matter
7	John	Science	1	7	7. The Structure of Matter
8	John	Science	1	8	8. The Structure of Matter
9	John	Science	1	9	9. The Structure of Matter
10	John	Science	1	10	10. The Structure of Matter
11	John	Science	1	11	11. The Structure of Matter
12	John	Science	1	12	12. The Structure of Matter
13	John	Science	1	13	13. The Structure of Matter
14	John	Science	1	14	14. The Structure of Matter
15	John	Science	1	15	15. The Structure of Matter
16	John	Science	1	16	16. The Structure of Matter
17	John	Science	1	17	17. The Structure of Matter
18	John	Science	1	18	18. The Structure of Matter
19	John	Science	1	19	19. The Structure of Matter
20	John	Science	1	20	20. The Structure of Matter

Author ID	Author	Category	Category ID	Book ID	Book
1	John	Science	1	1	1. The Structure of Matter
2	John	Science	1	2	2. The Structure of Matter
3	John	Science	1	3	3. The Structure of Matter
4	John	Science	1	4	4. The Structure of Matter
5	John	Science	1	5	5. The Structure of Matter
6	John	Science	1	6	6. The Structure of Matter
7	John	Science	1	7	7. The Structure of Matter
8	John	Science	1	8	8. The Structure of Matter
9	John	Science	1	9	9. The Structure of Matter
10	John	Science	1	10	10. The Structure of Matter
11	John	Science	1	11	11. The Structure of Matter
12	John	Science	1	12	12. The Structure of Matter
13	John	Science	1	13	13. The Structure of Matter
14	John	Science	1	14	14. The Structure of Matter
15	John	Science	1	15	15. The Structure of Matter
16	John	Science	1	16	16. The Structure of Matter
17	John	Science	1	17	17. The Structure of Matter
18	John	Science	1	18	18. The Structure of Matter
19	John	Science	1	19	19. The Structure of Matter
20	John	Science	1	20	20. The Structure of Matter

Keyword ID	Keyword
1	biochemistry
2	ecology
3	ecophysiology
4	ecophysiology
5	ecophysiology
6	ecophysiology
7	ecophysiology
8	ecophysiology
9	ecophysiology
10	ecophysiology
11	ecophysiology
12	ecophysiology
13	ecophysiology
14	ecophysiology
15	ecophysiology
16	ecophysiology

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

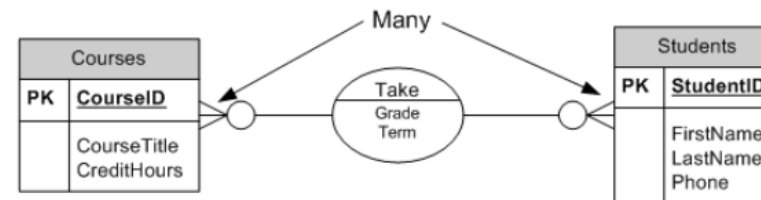
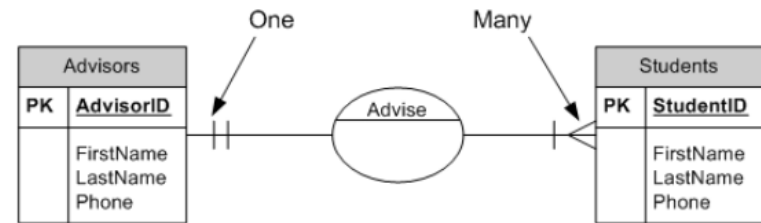
member_id ▲	Firstname	Lastname	University_Institution_affiliation	research_en
1	Beatrix	Beisner	3	My research program is currently centred on three ...
2	Monique	Poulin	7	Wetland management, restoration and conservati...
4	Jean-Pierre	Revéret	3	NULL
5	Andrew	Hendry	1	Darwin suggested that evolution proceeds 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: 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



Storing one to many relationships

Option 1 - Multiple columns

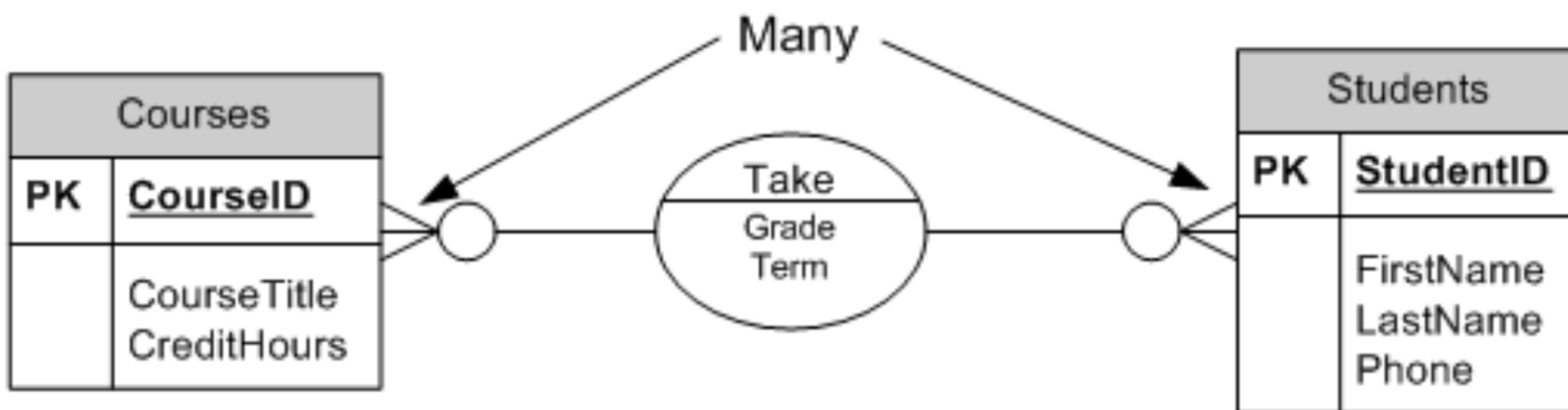
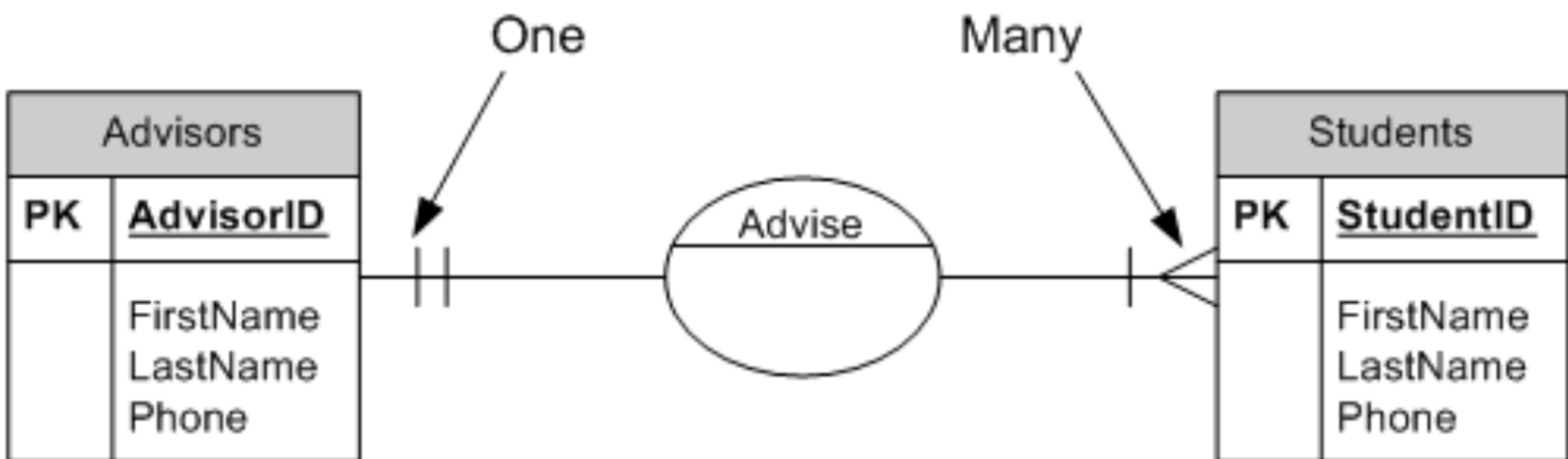
AdvisorID	StudentID	AdvisorID	StudentID
John Smith	John Smith	John Smith	John Smith
John Smith	John Smith	John Smith	John Smith

Option 2 - Lists

AdvisorID	StudentID
John Smith	John Smith
John Smith	John Smith

Option 3 - Lookup table

AdvisorID	StudentID
John Smith	John Smith
John Smith	John Smith



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

Supervisor_id	Supervisor_name
54	John Smith
56	Alan Gartner

Student_id	Student_name
23	Vicky Cook
15	Alain Lambert
12	Marc Sawyer

Supervisor_id	Student_id
54	23
54	15
56	12

Primary keys

No Key

Project ID	Firstname	Lastname	University Institution Affiliation	Project Title	Supervisor
1	Tyler	Barnett		Optical simulations of electron disease services	54
2	Yusef	Shariq		Molecular interactions of electron micrographs	52
3	Kyoko	Okazaki		Adaptation as a spatiotemporal process of natural systems	5
4	Shelley	Asch	A.M.U.		21
5	Heidi	Stachow		Functional response of forest vegetation to insects	2
7	George	Lima Barros		Morphological Variation of the Red-backed Vireo (Vireo ...)	52
8	Magnus	Rui	A.M.U.		20
9	Patrick	Diagnon	A.O.		30
10	Luis	Balcarraz-Masch		Spatiotemporal changes of the southern limit of ...	20
11	Johanna	Priddy		Investigating the establishment success of introduced ...	5
12	Dennis	Chapman		Modeling the abundance of western North American ...	30
13	Gery	Olivero		Modeling spread of invasive species	5
14	Paul	Kawachi		Humidity invasive species	4
15	Tiffany	Shel		Impacts of diversity of edge communities	30
16	Maria-Julia	Faria		The ecological and developmental genetic basis of ...	10
17	Esther	Gagnon		Phylogenetic relationships and biogeography of ...	10
18	Isabelle	Chouf	F.M.U.		18
19	Nikola	Jones		Modeling the effects of climate change on the ...	20
20	Lisa	Jones		The role of biotic and abiotic factors in water ...	30
21	Joseph	Isabelle de Sarrac		F.M.U.	18
22	Mayno	Leaveney		Direct development with ruse eggs in the cichlids	10
23	Chantal	Moham		Deposits des arthropodes humides sur les surfaces ...	10
24	Georgina	Chiriac		The Biological Causes and Consequences of the ...	18

Primary Key

Primary
keys



Supervisor_id	Supervisor_name
54	John Smith
58	Alan <u>Gartner</u>

Student_id	Student_name
23	Vicky <u>Côté</u>
15	<u>Alain</u> Lambert
12	Marc <u>Sauvé</u>

Supervisor_id	Student_id
54	23
54	15
58	12

No Key



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.
LOBLOB A string with a maximum length of 4294967295 characters.

Date-time

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

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

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.

LOBLOB A string with a maximum length of 4294967295 characters.

Date-time

DATE YYYY-MM-DD.

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.

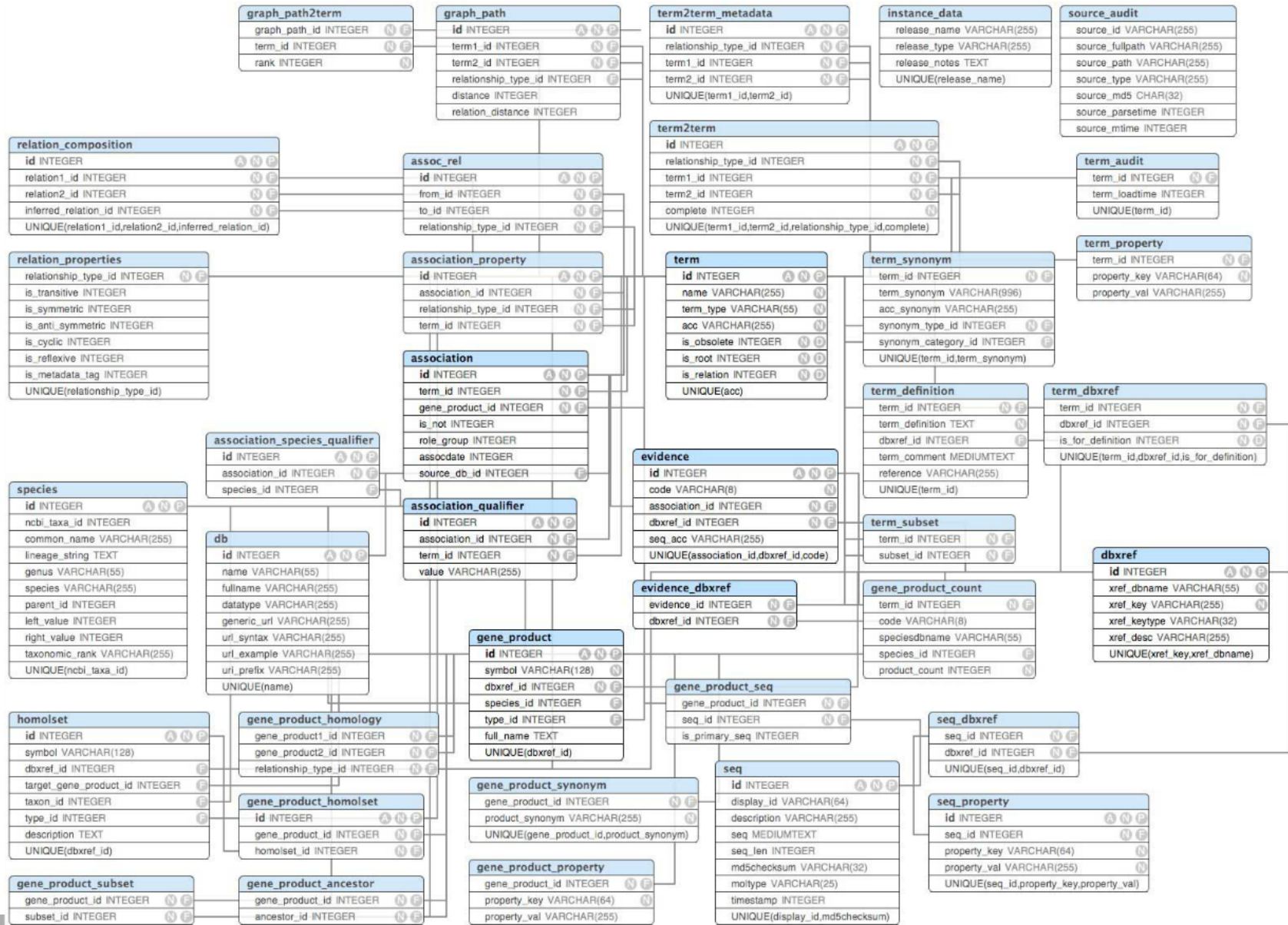
NULL

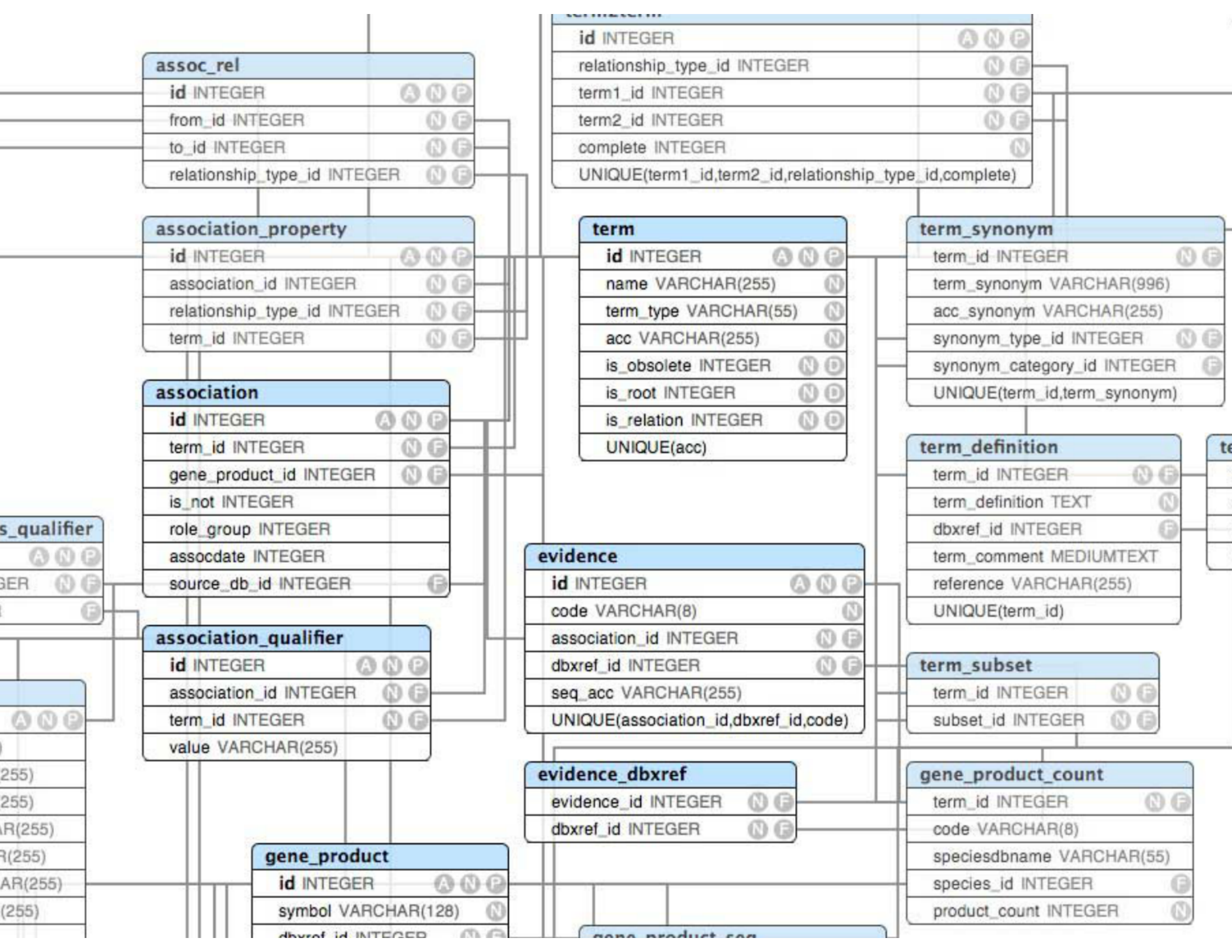
Used for unknown attributes

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





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/ods)
- From the PostgreSQL command line (CREATE TABLE COPY...)

Basic commands

```
CREATE TABLE table (column1 datatype, column2 datatype, column3 datatype);
INSERT INTO table (column1, column2, column3) VALUES (value1, value2, value3);
SELECT * FROM table;
UPDATE table SET column1 = value1, column2 = value2;
DELETE FROM table WHERE condition;
```

Operators

```
SELECT column1 FROM table WHERE condition;
JOIN: GROUP BY column1 HAVING condition;
CREATE BY Column;
```

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

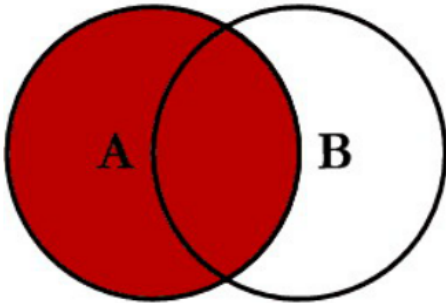
DROP delete table or database

Operators

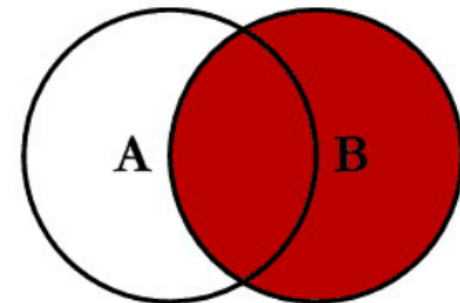
SELECT columns **FROM** tables **WHERE** conditions
JOIN ... GROUP BY columns **HAVING** condition
ORDER BY columns



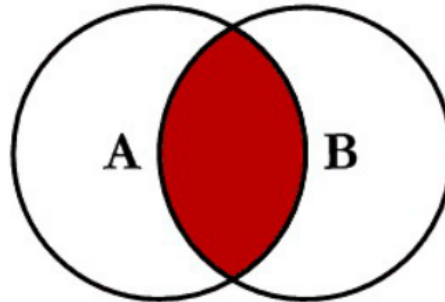
SQL JOINS



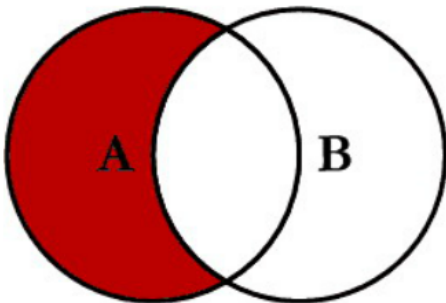
```
SELECT <select_list>  
FROM TableA A  
LEFT JOIN TableB B  
ON A.Key = B.Key
```



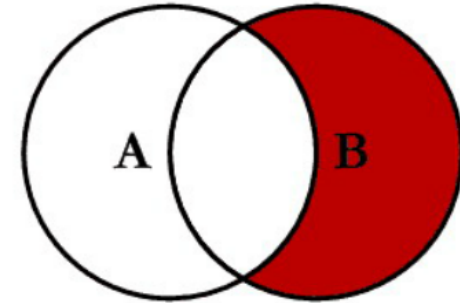
```
SELECT <select_list>  
FROM TableA A  
RIGHT JOIN TableB B  
ON A.Key = B.Key
```



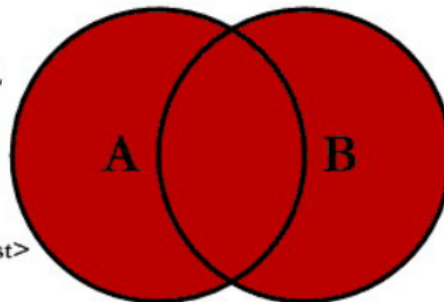
```
SELECT <select_list>  
FROM TableA A  
INNER JOIN TableB B  
ON A.Key = B.Key
```



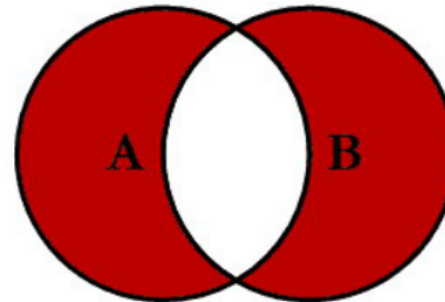
```
SELECT <select_list>  
FROM TableA A  
LEFT JOIN TableB B  
ON A.Key = B.Key  
WHERE B.Key IS NULL
```



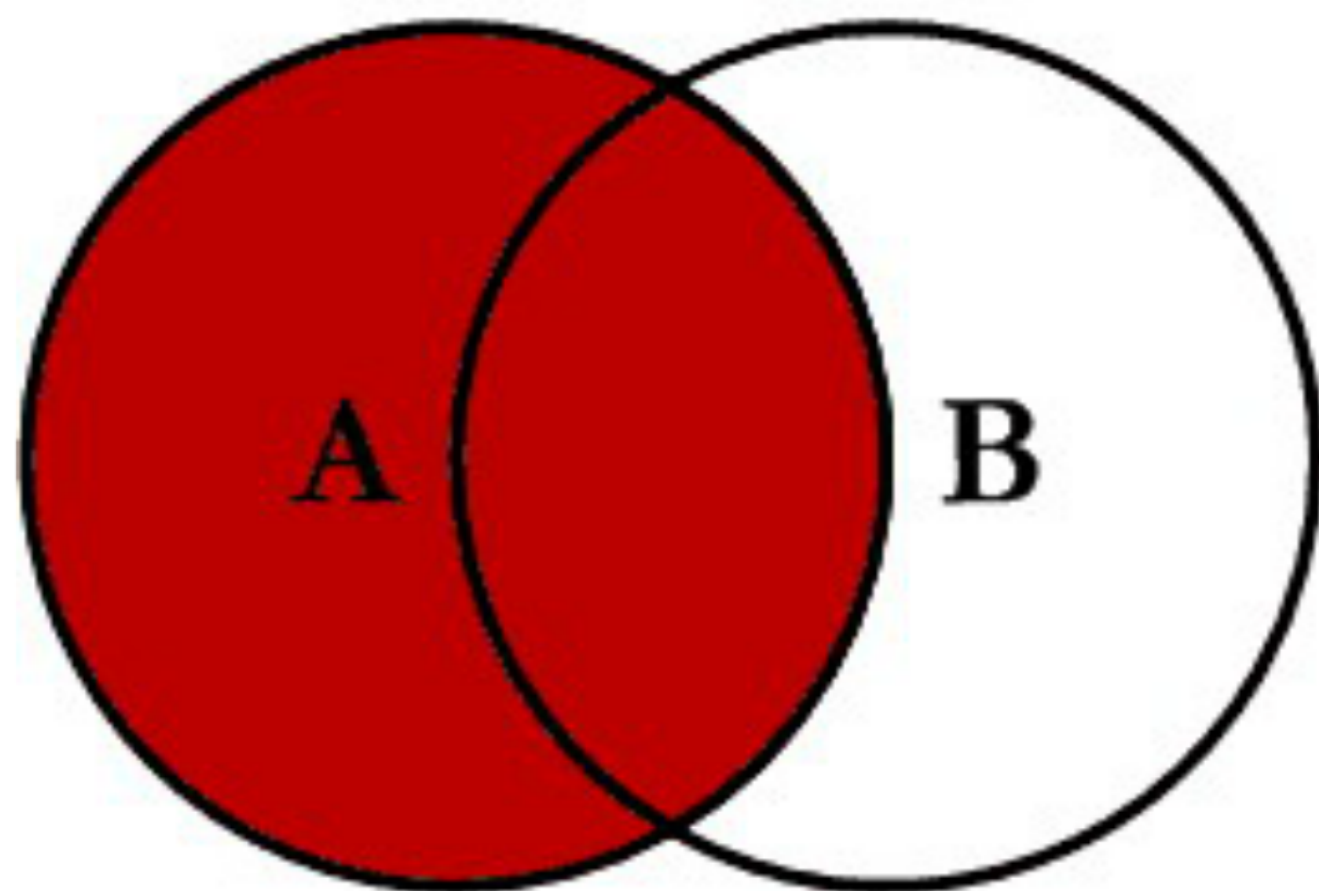
```
SELECT <select_list>  
FROM TableA A  
RIGHT JOIN TableB B  
ON A.Key = B.Key  
WHERE A.Key IS NULL
```



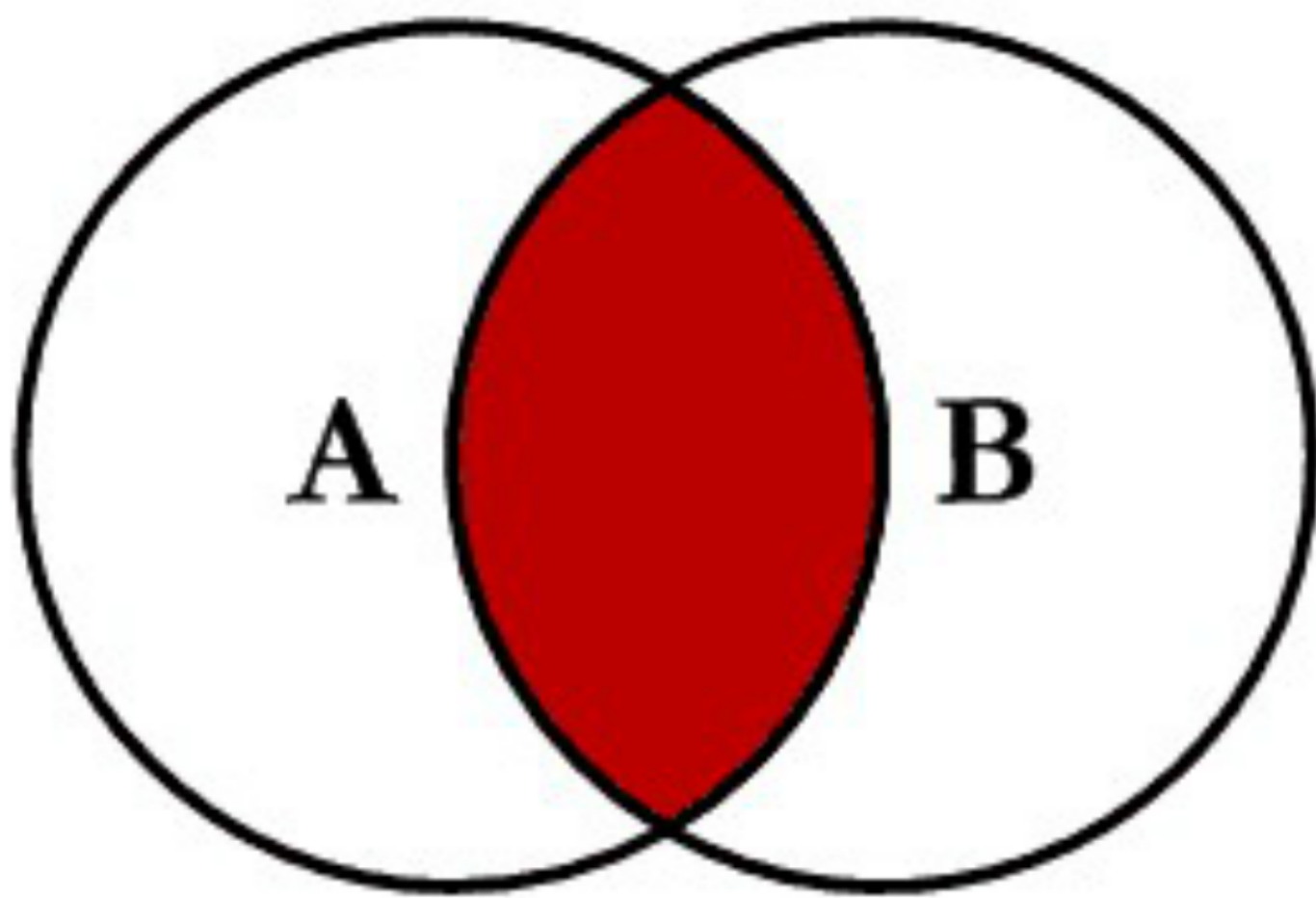
```
SELECT <select_list>  
FROM TableA A  
FULL OUTER JOIN TableB B  
ON A.Key = B.Key
```



```
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
```



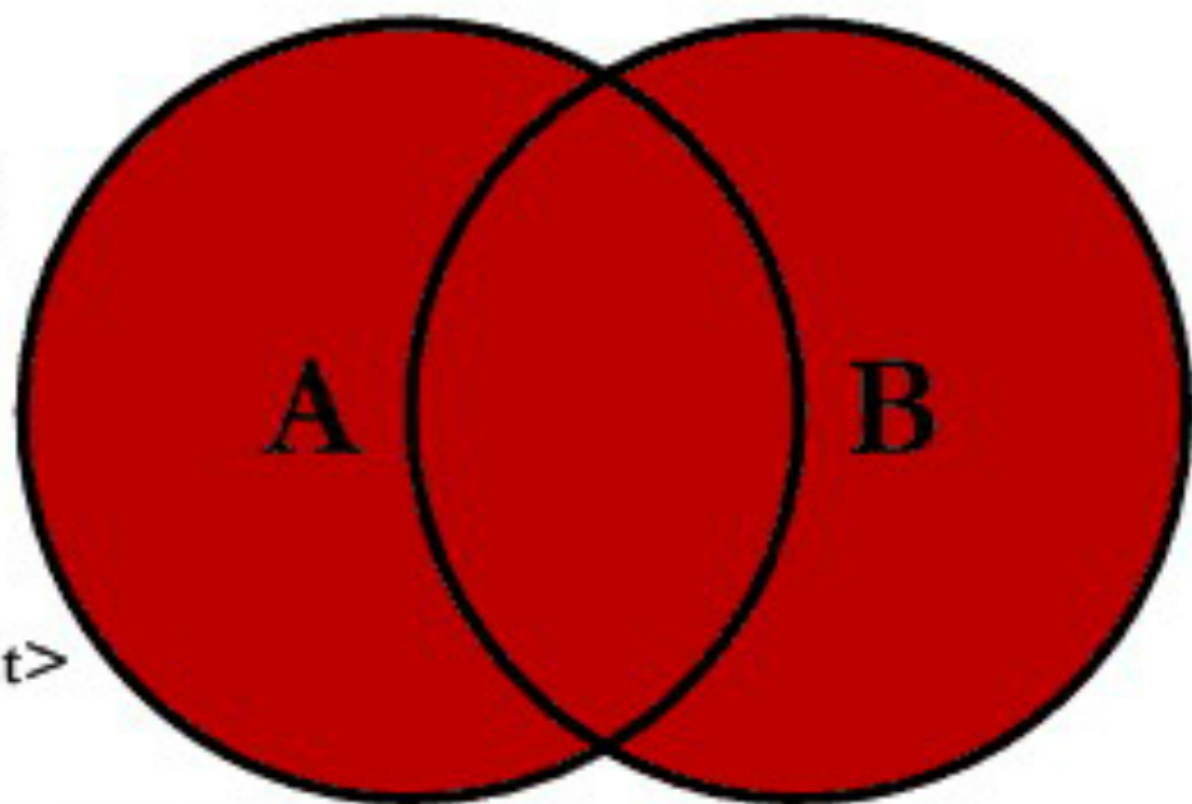
```
SELECT <select_list>  
FROM TableA A  
LEFT JOIN TableB B  
ON A.Key = B.Key
```

```
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 world's most advanced open source database



LibreOffice
The Document Foundation