GRUNDLÆGGENDE PROGRAMMERING

SQL II: Joins, Keys, Group-By, & Java+SQL!



Claus Brabrand

(((brabrand@itu.dk)))

Associate Professor, Ph.D. (((Software and Systems)))

IT University of Copenhagen

AGENDA

- Recap
- Join (samkøring af data)
- Indices
- Left and Right Join
- Group by
- Java + SQL!

SQL: Main Commands

Data definition:

- CREATE TABLE
- DROP TABLE
- SHOW TABLES
- DESCRIBE

// creates a new table

- // deletes a table
- // see tables defined
- // info about a table

Data manipulation:

- INSERT INTO .. VALUES (..) // insert record(s)
- SELECT .. FROM .. WHERE // retrieves info !!!
- DELETE FROM .. WHERE // delete record(s)
- UPDATE .. SET .. WHERE // changes record(s)

Information Retrieval

students:

name	id	address
Anna	1234	Somewhere 3
Brian	0001	Homestreet 4
Claire	0002	Nowhere 9b

■ SELECT-FROM:

SELECT * FROM students ;



name	id	address	
Anna	1234	Somewhere 3	
Brian	0001	Homestreet 4	
Claire	0002	Nowhere 9b	

■ SELECT-FROM-WHERE:

SELECT name, address FROM students WHERE id > 1;



name	address	
Anna	Somewhere 3	
Claire	Nowhere 9b	

AGENDA

- Recap
- Join (samkøring af data)
- Indices
- Left and Right Join
- Group by
- Java + SQL!

Lov om offentlige myndigheders registre

Lov nr. 294 af 8. juni 1978:

- § 4. Registre, der føres for en statslig myndighed, må kun oprettes i henhold til godkendelse af vedkommende minister efter forhandling med finansministeren. [...]
- Stk 3. Bestemmelserne [...] gælder tilsvarende ved samkøring af registre, der er oparbejdet med henblik på varetagelse af forskellige opgaver.
 - (Dette gælder dog ikke samkøring, der udelukkende foretages med henblik på uddrag i statistisk eller videnskabeligt øjemed.)

mailing_list:

name	email
Claus	brabrand@itu.dk
Barack	obama@hotmail.com
John	jdoe@notmail.com

phone_numbers:

email	phone	
brabrand@itu.dk	7218 5076	
obama@hotmail.com	212-555-0000	
jdoe@notmail.com	123-456-7890	

name	email	email	phone
Claus	brabrand@itu.dk	brabrand@itu.dk	7218 5076
Barack	obama@hotmail.com	brabrand@itu.dk	7218 5076
John	jdoe@notmail.com	brabrand@itu.dk	7218 5076
Claus	brabrand@itu.dk	obama@hotmail.com	212-555-0000
Barack	obama@hotmail.com	obama@hotmail.com	212-555-0000
John	jdoe@notmail.com	obama@hotmail.com	212-555-0000
Claus	brabrand@itu.dk	jdoe@notmail.com	123-456-7890
Barack	obama@hotmail.com	jdoe@notmail.com	123-456-7890
John	jdoe@notmail.com	jdoe@notmail.com	123-456-7890

"join" operation:

gives all combos!

Join (1/3)

Join (2/3)

mailing_list:

name	email
Claus	brabrand@itu.dk
Barack	obama@hotmail.com
John	jdoe@notmail.com

phone_numbers:

email	phone	
brabrand@itu.dk	7218 5076	
obama@hotmail.com	212-555-0000	
jdoe@notmail.com	123-456-7890	

SELECT *

	name	email	email	phone
√	Claus	brabrand@itu.dk	brabrand@itu.dk	7218 5076
X	Barack	obama@hotmail.com	brabrand@itu.dk	7218 5076
X	John	jdoe@notmail.com	brabrand@itu.dk	7218 5076
X	Claus	brabrand@itu.dk	obama@hotmail.com	212-555-0000
√	Barack	obama@hotmail.com	obama@hotmail.com	212-555-0000
X	John	jdoe@notmail.com	obama@hotmail.com	212-555-0000
X	Claus	brabrand@itu.dk	jdoe@notmail.com	123-456-7890
X	Barack	obama@hotmail.com		

"join" operation:

gives all combos!

now, only the rows we want :-)

name	email	email	phone
Claus	brabrand@itu.dk	brabrand@itu.dk	7218 5076
Barack	obama@hotmail.com	obama@hotmail.com	212-555-0000
John	jdoe@notmail.com	jdoe@notmail.com	123-456-7890

jdoe@notmail.com

John

Join (3/3)

mailing_list:nameemailClausbrabrand@itu.dkBarackobama@hotmail.comJohnjdoe@notmail.com

phone_numbers:

7218 5076

212-555-0000

123-456-7890

email	phone	
brabrand@itu.dk	7218 5076	
obama@hotmail.com	212-555-0000	
jdoe@notmail.com	123-456-7890	

SELECT name, mailing_list.email, phone
FROM mailing_list, phone_numbers
WHERE mailing_list.email = phone_numbers.email;

_				
	name	email 🗸	email 🗡	phone •
/	Claus	brabrand@itu.dk	brabrand@itu.dk	7218 5076
X	Barack	obama@hotmail.com	brabrand@itu.dk	7218 5076
X	John	jdoe@notmail.com	brabrand@itu.dk	7218 5076
X	Claus	brabrand@itu.dk	obama@hotmail.com	212-555-0000
√	Barack	obama@hotmail.com	obama@hotmail.com	212-555-0000
X	John	jdoe@notmail.com	obama@hotmail.com	212-555-0000
X	Claus	brabrand@itu.dk	jdoe@notmail.com	123-456-7890
X	Barack	obama@hotmail.com	jdo	
	lohn	idaa@natmail.aam	name email	phone

Claus

Barack

John

brabrand@itu.dk

obama@hotmail.com

jdoe@notmail.com

"join" operation:

gives all combos!

now, only the rows we want :-)

and only the col's we want :-)

√ John

jdoe@notmail.com

EXERCISE 1: Join

Database:

suspects:

name	dna_profile
Jack the Ripper	ACTGC
Jason	ACGTC
Amagermanden	ACTTC

^{+ 200.000} more people

crimescenes:

place	item	dna_found
Living Room	Table	ACCTC
Living Room	Lamp	AGTGC
Hallway	Chainsaw	ACCTC
Back Alley	Knife	ACTGC
Bathroom	Shampoo	ACCTC
Kitchen	Milk Carton	ACTTC

Join to determine (possibly multiple) perpetrators?



name	place	item
Jack the Ripper	Back Alley	Knife
Amagermanden	Kitchen	Milk Carton

EXERCISE 2: Join

customers:

cpr	name	phone	address
321085-1111	Anna	212-555-7755	Somewhere 4
221188-2222	Brian	212-555-5050	Anystreet 7
010190-3333	Claire	212-555-0707	Some Other Place 15
020290-4444	Danny	212-555-9999	Anywhere 1a

orders:

cpr	total
010190-3333	1,525.24
321085-1111	195.56
040587-5555	1,020.30
221188-2222	897.20

Find names & phone #s of "good customers" ("good customer" = spent more than 500 DKK):



name	phone	total
Brian	212-555-5050	897.20
Claire	212-555-0707	1,525.24

Three-way Join

Three-way joins:

```
SELECT students.id, students.name
FROM students, courses, enrollment
WHERE students.id = enrollment.id
AND enrollment.id = courses.id;
```

The above query can be abbeviated to (via "AS" aliases):

```
SELECT s.id, s.name
FROM students AS s , courses AS c , enrollment AS e
WHERE s.id = e.id
AND e.id = c.id ;
```

AGENDA

- Recap
- Join (samkøring af data)
- Indices
- Left and Right Join
- Group by
- Java + SQL!

How fast is a query?

- A join of two tables can take a very long time!
 - every record in one table needs to be compared to every record in the other table

- If both tables have 10,000 records (not uncommon!), we get 10,000 x 10,000 = 100,000,000 comparisons
 - This can easily take 20 seconds!
- However, if id1 and id2 have been indexed
 - This only takes 0.05 seconds!
 - (Note: the bigger the table, the more important with indexing)

Types of Indices (= Index'es)

Three types of indices:

type	needs to be unique?	is allowed to be null?
PRIMARY KEY	yes	no
UNIQUE	yes	yes
INDEX	no	yes

(Note: only one PRIMARY KEY allowed per table)

Example:

```
students:
CREATE TABLE students
                                                   stud id
     stud id INT PRIMARY KEY,
                                                          name
                                                                 age
                                                    1234
                                                                 20
                                                          Anna
    name VARCHAR (50) NOT NULL,
                                                          Brian
                                                                 25
                                                    5678
    age INT NOT NULL,
                                                    9999
                                                          Claire
                                                                 23
```

When to use Indices

Indices are best used on columns that are...:

...frequently used in the WHERE part:

```
SELECT * FROM table WHERE age > 20 ;
```

...frequently used in the ORDER BY part:

```
SELECT * FROM table ORDER BY name ;
```

...used as part of joins:

Foreign Keys

mailing_list: phone_numbers:

name	email < -	email	phone
Claus	brabrand@itu.dk	brabrand@itu.dk	7218 5076
Barack	obama@hotmail.com	obama@hotmail.com	212-555-0000
John	jdoe@notmail.com	jdoe@notmail.com	123-456-7890

Ensure "referential integrity":

i.e., that phone number.email is

always in mailing list.email!

Recall:

```
CREATE TABLE mailing_list (
   name VARCHAR(50) NOT NULL,
   email VARCHAR(100) NOT NULL
);
CREATE TABLE phone_numbers (
   email VARCHAR(100) NOT NULL,
   phone VARCHAR(20) NOT NULL
);
```

We want to link:

```
■ mailing list.email ← phone numbers.email
```

We can use a foreign key:

```
CREATE TABLE mailing_list (
    name VARCHAR(50) NOT NULL,
    email VARCHAR(100) NOT NULL
) ENGINE=InnoDB;

NB: some (annoying) stuff we have to write b/c we use Foreign Keys [long story...]

CREATE TABLE phone_numbers (
    email VARCHAR(100) NOT NULL,
    phone VARCHAR(20) NOT NULL,
    FOREIGN KEY(email) REFERENCES
    mailing_list(email)
    ) ENGINE=InnoDB;
```

AGENDA

- Recap
- Join (samkøring af data)
- Indices
- Left and Right Join
- Group by
- Java + SQL!

Recall "Join"

Database:

suspects:

name	dna_profile
Jack the Ripper	ACTGC
Jason	ACGTC
Amagermanden	ACTTC

crimescenes:

place	item	dna_found
Living Room	Table	ACCTC
Living Room	Lamp	AGTGC
Hallway	Chainsaw	ACCTC
Back Alley	Knife	ACTGC
Bathroom	Shampoo	ACCTC
Kitchen	Milk Carton	ACTTC

Join to determine (possibly multiple) perpetrators?



name	place	item
Jack the Ripper	Back Alley	Knife
Amagermanden	Kitchen	Milk Carton

One problem with joins

Database:

suspects:

name	dna_profile
Jack the Ripper	ACTGC
Jason	ACGTC
Amagermanden	ACTTC

crimescenes:

place	item	dna_found
Living Room	Table	ACCTC
Living Room	Lamp	AGTGC
Hallway	Chainsaw	ACCTC
Back Alley	Knife	ACTGC
Bathroom	Shampoo	ACCTC
Kitchen	Milk Carton	ACTTC

Suppose we instead would like to generate:

That is, <u>always</u> generate a record from the right table (not just when we happen to have a match)

```
SELECT place, item, name
FROM
   suspects RIGHT JOIN crimescenes
   ON dna_profile = dna_found;
```

place	item	name (if match)
Living Room	Table	NULL
Living Room	Lamp	NULL
Hallway	Chainsaw	NULL
Back Alley	Knife	Jack the Ripper
Bathroom	Shampoo	NULL
Kitchen	Milk Carton	Amagermanden

LEFT JOIN vs. RIGHT JOIN

Database:

suspects:

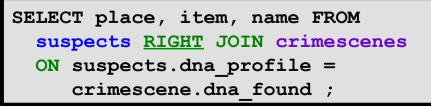
name	dna_profile
Jack the Ripper	ACTGC
Jason	ACGTC
Amagermanden	ACTTC

crimescenes:

place	item	dna_found
Living Room	Table	ACCTC
Living Room	Lamp	AGTGC
Hallway	Chainsaw	ACCTC
Back Alley	Knife	ACTGC
Bathroom	Shampoo	ACCTC
Kitchen	Milk Carton	ACTTC

Same thing for LEFT JOIN (just swap arguments):

```
SELECT place, item, name FROM
    crimescenes LEFT JOIN suspects
    ON suspects.dna_profile =
        crimescene.dna_found;
```



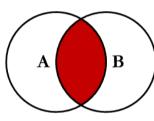
place	item	name (if match)
Living Room	Table	NULL
Living Room	Lamp	NULL
Hallway	Chainsaw	NULL
Back Alley	Knife	Jack the Ripper
Bathroom	Shampoo	NULL
Kitchen	Milk Carton	Amagermanden

Join (normal, left, right)

<u>a:</u>		
name	id	
Anna	1	
Brian	2	
Claire	3	

D:	
id	course
1	GRPRO
2	KKRT
4	BPAK

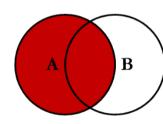
Normal join:



name	id	course
Anna	1	GRPRO
Brian	2	KKRT

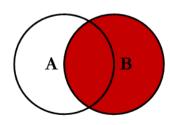
Left join:

```
SELECT name, a.id, course
FROM a LEFT JOIN b ON a.id = b.id ;
```



name	id	course (if match)
Anna	1	GRPRO
Brian	2	KKRT
Claire	3	NULL

Right join:



name (if match)	id	course
Anna	1	GRPRO
Brian	2	KKRT
NULL	4	BPAK

RIGHT JOIN

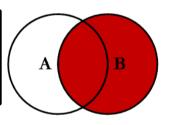
suspects:

name	dna_profile
Jack the Ripper	ACTGC
Jason	ACGTC
Amagermanden	ACTTC

crimescenes:

place	item	dna_found
Living Room	Table	ACCTC
Living Room	Lamp	AGTGC
Hallway	Chainsaw	ACCTC
Bedroom	Knife	ACTGC
Bathroom	Shampoo	ACCTC
Kitchen	Milk Carton	ACTTC

SELECT place, item, name
FROM suspects RIGHT JOIN crimescenes
ON suspects.dna_profile = crimescene.dna_found;



We, of course, have that:

place	item	name (if match)
Living Room	Table	NULL
Living Room	Lamp	NULL
Hallway	Chainsaw	NULL
Bedroom	Knife	Jack the Ripper
Bathroom	Shampoo	NULL
Kitchen	Milk Carton	Amagermanden

AGENDA

- Recap
- Join (samkøring af data)
- Indices
- Left and Right Join
- Group by
- Java + SQL!

GROUP BY

- Given table:
 - CREATE TABLE
 - INSERT INTO

expenses:

expense	dept	year	amount
salary	research	2011	490000
salary	sales	2012	1500000
salary	research	2012	500000
coffee	research	2013	800
coffee	sales	2013	300
salary	sales	2013	1600000
salary	research	2013	510000

Let's calculate total expenses per dept.:

SELECT dept, SUM(amount) FROM expenses GROUP BY dept ;

"GROUP BY" causes the records to be sorted and grouped wrt. their dept value:

expense	dept	year	amount
salary	research	2011	490000
salary	research	2012	500000
coffee	research	2013	800
salary	research	2013	510000
salary	sales	2012	1500000
coffee	sales	2013	300
salary	sales	2013	1600000

THEN the result is produced which involves calculating the 'SUM(amount)' values:

dept	SUM(amount)
research	1500800
sales	3100300

GROUP BY

The sum of expenses grouped by expense and department:

ZAPC113C31				
expense	dept	year	amount	
salary	research	2011	490000	
salary	sales	2012	1500000	
salary	research	2012	500000	
coffee	research	2013	800	
coffee	sales	2013	300	
salary	sales	2013	1600000	
		i	i	

research 2013

expenses:

salarv

SELECT expense, dept, SUM(amount) AS total FROM expenses
GROUP BY expense, dept;

The table is first grouped by expense:

dept exp. vear amount coffee research 2013 800 coffee | sales 2013 300 2011 490000 salary research salary 2012 1500000 sales 2012 500000 salary research salary sales 2013 1600000 2013 510000 salarv research

THEN, the table is sub-grouped by dept.:

exp.	dept	year	amount
coffee	research	2013	800
coffee	sales	2013	300
salary	research	2011	490000
salary	research	2012	500000
salary	research	2013	510000
salary	sales	2012	1500000
salary	sales	2013	1600000

FINALLY, the result is produced:

exp.	dept	total
coffee	research	800
coffee	sales	300
salary	research	1500000
salary	sales	3100000

510000

GROUP BY

The sum of expenses grouped by expense and department 2012-13:

ex	pe	n	S	es	

expense	dept	year	amount
salary	research	2011	490000
salary	sales	2012	1500000
salary	research	2012	500000
coffee	research	2013	800
coffee	sales	2013	300
salary	sales	2013	1600000
salary	research	2013	510000

SELECT expense, dept, SUM(amount) AS total FROM expenses
WHERE 2012 <= year AND year <= 2013 GROUP BY expense, dept;</pre>

The table is first grouped by expense:

dept exp. year amount coffee research 2013 800 2013 300 coffee | sales 490000 research 2012 1500000 sales salary 500000 2012 salary research 2013 1600000 salary sales 2013 510000 salarv research

THEN, the table is sub-grouped by dept.:

exp.	dept	year	amount
coffee	research	2013	800
coffee	sales	2013	300
salary	research	2011	490000
salary	research	2012	500000
salary	research	2013	510000
salary	sales	2012	1500000
salary	sales	2013	1600000

FINALLY, the result is produced:

exp.	dept	total
coffee	research	800
coffee	sales	300
salary	research	1010000
salary	sales	3100000

What is the total amount spent per expense?

expense	dept	year	amount
salary	research	2011	490000
salary	sales	2012	1500000
salary	research	2012	500000
coffee	research	2013	800
coffee	sales	2013	300
salary	sales	2013	1600000
salary	research	2013	510000

expense	SUM(amount)		
coffee	1100		
salary	4600000		

What is the average amount per expense?

expense	dept	year	amount
salary	research	2011	490000
salary	sales	2012	1500000
salary	research	2012	500000
coffee	research	2013	800
coffee	sales	2013	300
salary	sales	2013	1600000
salary	research	2013	510000

expense	AVG(amount)		
coffeee	550		
salary	920000		

What are the total expenses each year?

expense	dept	year	amount
salary	research	2011	490000
salary	sales	2012	1500000
salary	research	2012	500000
coffee	research	2013	800
coffee	sales	2013	300
salary	sales	2013	1600000
salary	research	2013	510000

year	SUM(amount)
2011	490000
2012	2000000
2013	2111100

What are the expenses per year per department?

expense	dept	year	amount
salary	research	2011	490000
salary	sales	2012	1500000
salary	research	2012	500000
coffee	research	2013	800
coffee	sales	2013	300
salary	sales	2013	1600000
salary	research	2013	510000

year	dept	total
2011	research	490000
2012	sales	1500000
2012	research	500000
2013	research	510800
2013	sales	1600300

What are the expenses each year in 2012-13?

expense	dept	year	amount
salary	research	2011	490000
salary	sales	2012	1500000
salary	research	2012	500000
coffee	research	2013	800
coffee	sales	2013	300
salary	sales	2013	1600000
salary	research	2013	510000

year	SUM(amount)		
2012	2000000		
2013	2111100		

What is the highest expenditure for each dept?

expense	dept	year	amount
salary	research	2011	490000
salary	sales	2012	1500000
salary	research	2012	500000
coffee	research	2013	800
coffee	sales	2013	300
salary	sales	2013	1600000
salary	research	2013	510000

dept	MAX(amount)	
research	510000	
sales	1600000	

What is the highest exp. for each dept per year?

expense	dept	year	amount
salary	research	2011	490000
salary	sales	2012	1500000
salary	research	2012	500000
coffee	research	2013	800
coffee	sales	2013	300
salary	sales	2013	1600000
salary	research	2013	510000

dept	year	MAX(amount)
research	2011	490000
research	2012	500000
research	2013	510000
sales	2012	1500000
sales	2013	1600000

AGENDA

- Recap
- Join (samkøring af data)
- Indices
- Left and Right Join
- Group by
- Java + SQL!

```
import java.sql.*; // Import required packages
                                                          Java + SQL!
public class MySQL {
  // JDBC driver name and database URL
  static final String JDBC DRIVER = "com.mysql.jdbc.Driver";
                                                                 SELECT * FROM mailing list ;
  static final String database name = "grpro";
  static final String DB URL = "jdbc:mysql://mysql.itu.dk/" + database name;
 // Database credentials
                                                                mailing list:
  static final String USER = "brabrand";
                                                                          email
                                                                 name
  static final String PASS = "*******;
                                                                 Barack
                                                                          obama@hotmail.com
 public static void main(String[] args) {
    Connection connection = null;
                                                                 Michelle i m.obama@hotmail.com
    Statement statement = null;
    try {
     DriverManager.registerDriver(new com.mysql.jdbc.Driver()); // Register driver
     connection = DriverManager.getConnection(DB URL, USER, PASS); // Open connection
     statement = connection.createStatement(); // Create statement
     String sql = "SELECT * FROM mailing list"; // NB: implicit semi-colon!
     ResultSet rs = statement.executeOuery(sql); // *** EXECUTE OUERY! ***
     // STEP 5: Extract data from result set
     while (rs.next()) { // Retrieve data by column name
        String email = rs.getString("email");
        String name = rs.getString("name");
       // int id = rs.getInt("id");
       System.out.println("Name: '" + name + "', Email: '" + email + "'"); // Display data
     rs.close(); // close query
     connection.close(); // close connection
    } catch(Exception e) {
     e.printStackTrace(); // handle errors
```

Compile and Run

Jar (you have to download it):

mysql.jar

Compile:

%> javac -cp mysql.jar MySQL.java

Run:

```
%> java -cp mysql.jar:. MySQL
```

Name: 'Barack', Email: 'obama@hotmail.com'

Name: 'Michelle', Email: 'm.obama@hotmail.com'

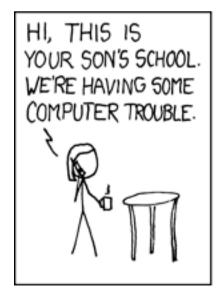
SQL Injection Attacks :-)

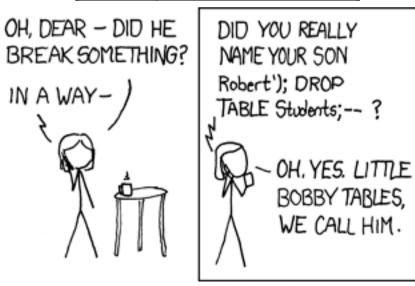
"Inject" following SQL command:

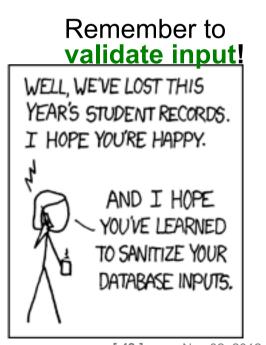
DROP TABLE mailing list ;

mailing_list:

name	email	
Claus	brahrand@itu.dk	
Barack	obama@nctmail.com	
John	jdoe@notmail.com	







Claus Brabrand, ITU, Denmark

SQL Injection Attacks:-)





OR 1=1:--

Exercises 9 (for Oct 28, 2014)

Exercise 9.1:

- Create own MySQL database (hosted on the ITU MySQL server)
- Log into it via the MySQL text client

Exercise 9.2:

- Create three tables (each with its own 'id' column):
 - "Courses", "Students", and "Enrollment"
- ...and populate your database with sample real'ish data

Exercise 9.3:

Try out the various queries from these slides (GRPRO-15/16)

Exercises 9 (for Oct 28, 2014)

Exercise 9.4: (Three-way joins)

- Join data from your three tables as explained below:
 - "Courses", "Students", and "Enrollment"
- a) Select all students attending "Grundlæggende Programmering"

student	course	
Bent	Grundlæggende Programmering	
Anna	Grundlæggende Programmering	

(don't simply use the key: "GRPRO")

student	teacher
Bent	Claus
Carl	Mathias
Anna	TBA
Anna	Claus

b) Compute the student-teacher relation:

Exercise 9.5:

- Find some (any) data online and turn it into at least two tables.
- Perform a join on it. Finally send your SQL commands to me. :-)

Thx!

Questions?