



FAKULTA ELEKTROTECHNICKÁ

České vysoké učení technické v Praze

B4M36DS2 – Database Systems 2

Practical Class 1

Introduction: Organization. Relational VS NoSQL Databases

Yuliia Prokop

prokoyul@fel.cvut.cz, Telegram [@Yulia_Prokop](https://www.instagram.com/Yulia_Prokop)

Based on **Martin Svoboda**'s materials (<https://www.ksi.mff.cuni.cz/~svoboda/courses/211-B4M36DS2/>)



CourseWare Wiki

<https://cw.fel.cvut.cz/b231/courses/b3b36prg/start>

Basic course information

Lectures: Monday, 9:15 – 10:45

Practical classes: Monday, 12:45 – 14:15, 14:30 – 16:00, 16:15 – 17:45

Homework – maximum 130 points

Course credit – minimum 80 points

Exam – maximum 100 points

- **written exam** (mandatory) + **oral exam** (optional)

[CourseWare Wiki](#) – course materials

[BRUTE](#) – upload reports on the homework

NoSQL Server – submit and execute homework

Lab KN:E-328 user accounts

- Set your new password

<https://www.felk.cvut.cz/labpass/>

NoSQL server

- SSH and SFTP access
- **nosql.felk.cvut.cz**
 - MapReduce, Redis, Cassandra, MongoDB, Neo4j
- Login and password: sent by e-mail

Linux

- `ssh login@host` – login to remote server
`ssh login@nosql.felk.cvut.cz`
`exit`
- `sftp -P port login@host`
 - `cd directory` – change remote directory
 - `lcd directory` – change local directory
 - `ls` – list remote directory contents
 - `lls` – list local directory contents
 - `put local remote` – copy a local file to the remote directory (sftp access)
 - `get remote local` – copy a remote to the local directory (sftp access)
 - `bye` or `exit` – disconnect

Windows

- **PuTTY** – <http://www.chiark.greenend.org.uk/~sgtatham/putty/>
- **WinSCP** – <http://winscp.net/>

Change your initial password

- password
 - enter the current password
 - enter the new password

Browse important directories

- /home/login/ – personal directory with your data
- /home/DS2/ – shared directory⁷ with course data

1. Submit your home assignments

- Upload your submission files to the **NoSQL server**
- Put these files into a sub-directory `~/assignments/name/`, where **name** is a name of a given homework
- This **name** parameter must also correspond to one of the predefined assignment names: `formats`, `mapreduce`, `redis`, `cassandra`, `mongodb`, `neo4j` (case sensitive)
- Use **ssh** or **PuTTY** to open a remote shell connection to the NoSQL server
- Based on the instructions provided for a given homework assignment, verify that everything is working as expected

Homework Assignments: submission

- Go to the `~/assignments/` directory and execute:

`sudo submit_execute` `name`

where `name` is the name of the homework

- Wait for the confirmation of success, otherwise, your homework is not considered to be submitted
- If any complications appear, write to me or send your solution by e-mail to prokoyul@fel.cvut.cz.
- Just for your convenience, you can check the submitted files in the `~/submissions/directory`

9

Example of submission

HOMEWORK ASSIGNMENT SUBMISSION SCRIPT

Initializing assignment submission...

Preparing <sparql> assignment submission...

Submitting <sparql> assignment...

```
-rw-rw-r-- 1 f221_serhii f221_serhii 0 Aug 28 18:40 data.ttl
-rw-rw-r-- 1 f221_serhii f221_serhii 0 Aug 28 18:40 query1.sparql
-rw-rw-r-- 1 f221_serhii f221_serhii 0 Aug 28 18:40 query2.sparql
-rw-rw-r-- 1 f221_serhii f221_serhii 0 Aug 28 18:40 query3.sparql
-rw-rw-r-- 1 f221_serhii f221_serhii 0 Aug 28 18:40 query4.sparql
-rw-rw-r-- 1 f221_serhii f221_serhii 0 Aug 28 18:40 query5.sparql
```

Assignment <sparql> SUBMITTED by <f221_serhii> SUCCESSFULLY

Check your submission at <~/submissions/f221_serhii-sparql-20220828-184814>

If one of the required files is absent:

```
f221_serhii@database:~/assignments$ sudo submit_execute cassandra
```

HOMEWORK ASSIGNMENT SUBMISSION SCRIPT

Initializing assignment submission...

Preparing <cassandra> assignment submission...

File <cassandra/script.cql> is required and missing

Requirements:

- Respect the prescribed names of individual files to be submitted
(**case sensitive**)
- Place all the files in the root directory of your submission
- Do not include shared libraries or files that are not requested
- I.e. do not submit files that were not explicitly requested
- Do not redirect or suppress both standard and error outputs in your shell scripts
- All your files must be syntactically correct and executable without errors

2. Upload .pdf of your home assignment to BRUTE

- Use the filename: **username_number.pdf**, where the number is the homework number.
- The report must fulfill the requirements and have the following structure:
 - ✓ A verbal description of the task
 - ✓ Code (copy from the file uploaded to the server)
12
 - ✓ Screenshot of code execution on the server
- If there are several problems, repeat the same for each of them

Students database

A student has a name, phone number, date of birth and gender.

A course has a name and number of credits.

Students can take courses and receive grades as a result.

Alexander got a B in Data Mining, Roman got an A in Artificial Intelligence and a B in Database systems, Tereza got a C in Artificial Intelligence.

Name	Phone	DateOfBirth	Sex	Course title	Credits	Grade
Alexander	666-555-444	06.03.2000	M	Data Mining	5	B
Roman	777-666-555	23.08.1999	M	Artificial Intelligence, Database systems	6, 5	A, B
Tereza	555-666-777	14.05.2000	F	Database systems	5	C

The table is not in the **1st Normal Form** (attributes Course title, Credits and Grade are not single)

Relational databases: Normal forms

Name	Phone	DateOfBirth	Sex	Course title	Credits	Grade
Alexander	666-555-444	06.03.2000	M	Data Mining	5	B
Roman	777-666-555	23.08.1999	M	Artificial Intelligence	6	A
Roman	777-666-555	23.08.1999	M	Database systems	5	B
Tereza	555-666-777	14.05.2000	F	Database systems	5	C

The table is in the **1st Normal Form** (all attributes are simple)

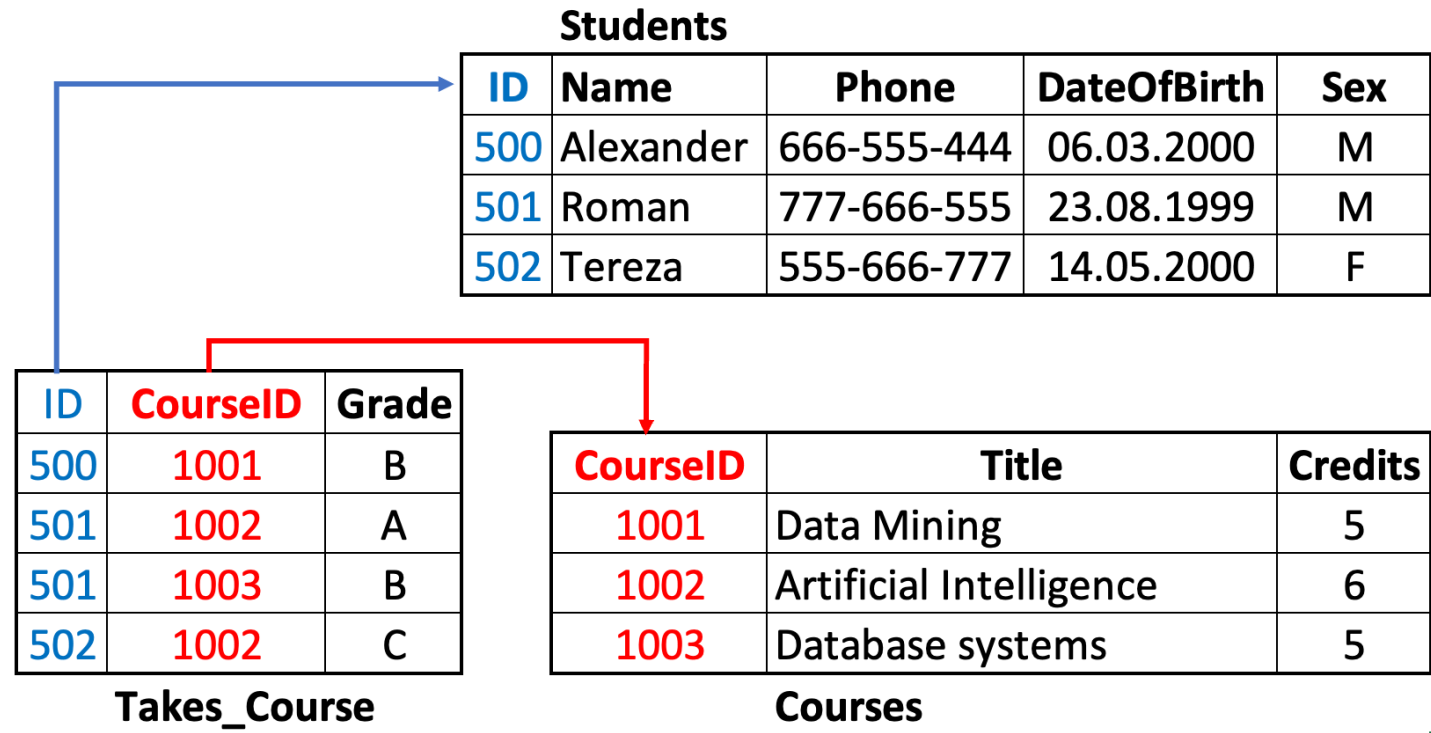
The table is not in the **2nd Normal Form** (The primary key is Name+Course title.

But Phone only depends on Name, Credits only depends on Course title and

Grade depends on both Name and Course title).

Relational databases : relationships

- ✓ Relational databases allow you to define **relationships** between different data sets.
- ✓ **Foreign keys** are used to define the relationships among the tables.



Representing a LinkedIn profile using a relational schema

<http://www.linkedin.com/in/williamhgates>



Bill Gates
Greater Seattle Area | Philanthropy

Summary
Co-chair of the Bill & Melinda Gates Foundation. Chairman, Microsoft Corporation. Voracious reader. Avid traveler. Active blogger.

Experience
Co-chair • Bill & Melinda Gates Foundation
2000 – Present
Co-founder, Chairman • Microsoft
1975 – Present

Education
Harvard University
1973 – 1975
Lakeside School, Seattle

Contact Info
Blog: thegatesnotes.com
Twitter: @BillGates

user_id	first_name	last_name	summary	region_id	industry_id	photo_id
251	Bill	Gates	Co-chair of ... blogger	us:91	131	57817532

id	user_id	job_title	organization
458	251	Co-chair	Bill & Melinda Gates F...
457	251	Co-founder, Chairman	Microsoft

id	industry_name
43	Financial Services
48	Construction
131	Philanthropy

id	user_id	school_name	start	end
807	251	Harvard University	1973	1975
806	251	Lakeside School, Seattle	NULL	NULL

id	region_name
us:7	Greater Boston Area
us:91	Greater Seattle Area

id	user_id	type	url
155	251	blog	http://thegaresnotes.com
156	251	twitter	http://twitter.com/BillGates

Source: https://ebrary.net/64616/computer_science/object_relational_mismatch

- ✓ **Structured data.** Data is in the form of tables with strong structure.
- ✓ **1NF, 2NF, 3NF BCNF**
- ✓ The tables have **primary keys** and are linked with **foreign keys**.
- ✓ **SQL** (Standart Query Language) exists for querying.
- ✓ **Transactions.**

Representing a LinkedIn profile as a JSON document

<http://www.linkedin.com/in/williamhgates>



Bill Gates

Greater Seattle Area | Philanthropy

Summary

Co-chair of the Bill & Melinda Gates Foundation. Chairman, Microsoft Corporation. Voracious reader. Avid traveler. Active blogger.

Experience

Co-chair • Bill & Melinda Gates Foundation
2000 – Present

Co-founder, Chairman • Microsoft
1975 – Present

Education

Harvard University
1973 – 1975

Lakeside School, Seattle

Contact Info

Blog: thegatesnotes.com
Twitter: @BillGates

```
{
  "user_id": 251,
  "first_name": "Bill",
  "last_name": "Gates",
  "summary": "Co-chair of the Bill & Melinda Gates... Active blogger.",
  "region": "Greater Seattle Area",
  "industry": "Philanthropy",
  "photo_url": "/p/7/000/253/05b/308dd6e.jpg",
  "positions": [
    {
      "job_title": "Co-chair",
      "organization": "Bill & Melinda Gates Foundation»,
      "job_title": "Co-founder, Chairman", "organization": "Microsoft"
    }
  ],
  "education": [
    {
      "school_name": "Harvard University", "start": 1973, "end": 1975},
    {
      "school_name": "Lakeside School, Seattle", "start": null, "end": null}
  ],
  "contact_info": {
    "blog": "http://thegatesnotes.com",
    "twitter": "http://twitter.com/BillGates"
  }
}
```

Source: https://ebrary.net/64616/computer_science/object_relational_mismatch

Create a dataset for a relational database related to tourism and journeys. Insert into tables the following data:

Alice Brown with email `alice.brown@email.com` in 2015 travelled to Italy and visited Roma, Milan, and Venice. In 2016 she travelled to Poland and visited Warsaw and Krakov and also the Czech Republic and visited Prague and Brno.

Alex Fisher in 2016 travelled to Germany and visited Berlin, Erfurt and Koln.

Betty Fox with email `bet.fox@email.com` travelled to Istanbul in 2015 and to Roma in 2017.

Excercise 1 – solution (relational DBS)

Table 1: Tourists

TouristID	FirstName	LastName	Email
1	Alice	Brown	alice.brown@email.com
2	Alex	Fisher	NULL
3	Betty	Fox	bet.fox@email.com
4	Mark	Fox	mark.fox@email.com

Table 4: Journey Members

JourneyID	TouristID
1	3
1	4
2	1
2	2
3	3
3	4
4	1
4	2
4	3
4	4
5	1
5	3

Table 3: Journeys

JourneyID	Name	Year
1	Our First J	2014
2	Lovely Trip '15	2015
3	Fox Merry Tour	2015
4	Italy the best	2016
5	Shoping 2017	2017
6	Central Europe	2017

Excercise 1 – solution (relational DBS)

Table 2: Destinations

DestinationID	City	Country
1	Roma	Italy
2	Milan	Italy
3	Venice	Italy
4	Warsaw	Poland
5	Krakov	Poland
6	Prague	Czechia
7	Brno	Czechia
8	Berlin	Germany
9	Erfurt	Germany
10	Koln	Germany
11	Istanbul	Turkey

Table 5: Journeys Plan

JourneyID	DestinationID
1	11
2	3
2	4
2	5
2	6
2	7
3	1
3	3
3	11
4	1
4	2
4	3
5	2
...	...

Denormalize the tables from Ex. 1

Excercise 2 – solution

FirstName	LastName	Email	Year	Destinations
Alice	Brown	alice.brown@email.com	2015	Roma-Italy, Milan-Italy, Venice-Italy
Alice	Brown	alice.brown@email.com	2016	Warsaw-Poland, Krakov-Poland, Brno-Czech Republic, Prague-Czech Republic
Alex	Fisher	NULL	2016	Berlin-Germany, Erfurt-Germany, Koln-Germany
Betty	Fox	bet.fox@email.com	2015	Istanbul
Betty	Fox	bet.fox@email.com	2017	Roma