NIE-PDB: **Advanced Database Systems**

http://www.ksi.mff.cuni.cz/~svoboda/courses/241-NIE-PDB/

Practical Class 0

Organization

Martin Svoboda

martin.svoboda@fit.cvut.cz

24. 9. 2024

Charles University, Faculty of Mathematics and Physics

Czech Technical University in Prague, Faculty of Information Technology

Organizational Aspects

Course Webpage

Course webpage

- https://courses.fit.cvut.cz/NIE-PDB/
- Contains all **course materials** and information
  - Semester plan
  - **Lecture slides**
  - **Lab materials**
  - **Homework assignments**
  - Formal requirements
  - Important links
  - …

User Credentials

User credentials

- **Sent in an e-mail** at the beginning of the semester
  - If you have not received it …
    - First, check your **spam folder**
    - If not successful, contact the teacher
- **Remember all the information**!
  - Namely your …
    - **Personal number**
    - **Login name**
    - **Initial password**
  - You will need them repeatedly during the whole semester
    - In order to **access the NoSQL server**
    - I.e., to attend the **classes** and work on the **assignments**

User Credentials

User credentials (cont'd)

- **Personal number**
  - Randomly **generated and fully <u>anonymous</u>**
    - Allows for tracking points and topic selection (see further)
  - E.g.: 3587
- **Login name**
  - User **account name** within the **NoSQL server and databases**
    - Based on the standard university login name
  - E.g.: pdb241_svobom25
- **Initial password**
  - Generated password for the NoSQL server and databases
    - These passwords can be changed later on (see further)
  - E.g.: Wx94i3

Google Table

Google table with points

- https://bit.ly/pdb-241-points
  - Allows to transparently **track your progress**
    - Throughout the whole semester
  - As well as **compare your results** with your colleagues
- Particular content
  - Points from **homework assignments**
  - Points from **exam tests**
  - …
- Identification based on the **personal numbers**
  - And so everything really is **anonymous**

Formal Requirements

Attendance

Attendance during **lectures** and **practical classes**

- <u>**Not**</u> **compulsory**
- Yet **sincerely recommended**
  - Since it allows you to …
    - Prepare continuously
    - Understand everything
    - Gain experience
    - Ask questions
    - Discuss solutions
    - …
  - However, only if **attending actively**, of course

Homework Assignments

Homework assignments

- Altogether **7 assignments**
  - XQuery, SPARQL, …
- **None is compulsory** on its own
  - Nevertheless, you will actually need to have almost all of them
    - To gain a sufficient number of points
- Must be within a **distinct individual topic**
  - You will choose it at the beginning of the semester
  - And you must then stick to it throughout the semester
- Strictly an **individual work**
  - I.e., **do not cheat**
    - Since it will prevent you from passing the course successfully
    - Not speaking of other possible consequences …

Homework Assignments

Assignment evaluation

- Solutions are **awarded by points**
  - Each up to 20 or 15 points depending on the difficulty
  - You can altogether get **130 points**
- **Quality shortcomings** are penalized
  - <u>Each</u> depending on the magnitude and relevancy
    - Usually by 1 to 3 points
    - But even more in case of major issues
- Each assignment is **evaluated only once**
- ⇒ **follow the instructions** carefully
  - Based on the experience …
    - Almost all deductions are unnecessary and easily avoidable
  - I.e., if you do what is expected, you will get the maximum

Homework Assignments

Submission delays

- Delays are penalized
    - Delay of **one day** is **penalized by 5 points**
    - **Shorter delays** are penalized proportionally
    - **Longer delays** are not further penalized
        - I.e., the penalty stays the same and no longer increases
        - Even if it is at the level of weeks or months
- ⇒ **follow the deadlines**
    - And do not postpone everything until the last moment

Course Credit

Course credit

- **At least 100 points** are required for the credit to be granted
  - Out of 130 points possible

Bonus points

- The more points you actually get, the better
  - Since **all the points above the credit boundary** …
  - … are **transferred to the exam as bonus points**
    - And so they can improve your overall result
- I.e., active work during the semester will be rewarded

Course Credit

Second chance

- If you are unable to get enough points for the course credit …
  - You can submit improved versions of your assignments
    - However, only **one correction for each assignment**
    - And deductions for delays will be preserved intact
- Moreover, if you use such an option …
  - Your semester points will be **truncated at the credit boundary**
    - And so **you will <u>not</u> get any bonus points**
  - I.e., it really is just an option to save your credit

Final Exam

Exam test

- Course credit is required
  - You will otherwise not be able to enroll for the exam
- **Written test** for **100 points**
  - **Practical part**
    - Various exercises similar to those from the practical classes
  - **Theoretical part**
    - Open and multiple choice questions
- **At least 30 % of points** from **<u>each</u> of the two parts** is required
  - You will otherwise not be successful
    - Regardless of how many points in total you would have

Final Exam

Overall result

- **Bonus points** are first applied
  - I.e., they are added to the result from the test
- **Homework solutions** must be submitted before the exam
  - I.e., the exam date you intend to participate
    - Otherwise the points (if any) cannot be considered
  - Note that it really suffices to just submit such solutions
    - I.e., they do not necessarily need to be assessed yet
- Standard university **grading scheme** is then used
  - **90+ for A, 80+ for B, 70+ for C, 60+ for D, and 50+ for E**
  - Having 49 or even less points means you were not successful

NoSQL Server

NoSQL Server

NoSQL server

- nosql.kti.in.fit.cvut.cz:42222
  - Server we will work with on a regular basis
- Hosts various **NoSQL databases** and **other systems** we need
  - XQuery, SPARQL, MapReduce, Riak, MongoDB, Cassandra, Neo4j
- Only allows **SSH and SFTP access**
  - Other types of connections are not possible
  - I.e., the listed systems cannot be accessed directly
    - Only via this server acting as a mediator
- **Login and password**
  - Sent by e-mail (as already discussed)
    - I.e., `pdb241_login`

Remote Access

Faculty VPN (**Virtual Private Network**)

- **NoSQL server** is **not** accessible remotely
  - I.e., from outside of the faculty private network
    - E.g., when you are at home
  - However, this may also include **labs at school**!
    - Specifically if you are connected via **Eduroam**
- **VPN connection** **is thus required**
  - Get acquainted with all instructions
    - https://help.fit.cvut.cz/vpn/en/
  - **Download and install OpenVPN**
  - **Configure the faculty connection**
    - So that you can establish it effortlessly later on

SSH Connection

Secure Shell (**SSH**)

- Allows to establish shell connections to remote servers
  - Learn how to do so …

Linux

- `ssh -p port login@host`
  - **Port number** must be specified via a separate option
  - Type `exit` to terminate the session
  - E.g.: `ssh -p 42222 \`
    `pdb241_login@nosql.kti.in.fit.cvut.cz`

Windows

- **Download and install PuTTY**
  - https://www.chiark.greenend.org.uk/~sgtatham/putty/
  - Configure and save your connection

Password Modification

Change your initial password

- And, of course, remember it …
- **Execute a special script** for this purpose
  - `password`
- Not only it changes your **NoSQL server password** as such
  - But also all your individual **database passwords**
    - E.g., passwords to MongoDB, Cassandra, …
  - This script is not able to tackle all the potential errors, though
    - I.e., check the output for any suggestions of **partial failures**
    - And contact the teacher if anything went wrong

SFTP Connection

Secure File Transfer Protocol (**FSTP**)

- Allows to **upload or download files** to / from remote servers

Linux

- `sftp -P port login@host`
  - Note that the **port number option** must be `-P` and not `-p`
  - Basic commands: `cd directory`, `ls`, `put local remote`, `get remote local`, `exit`
  - E.g.: `sftp -P 42222 \`
    `pdb241_login@nosql.kti.in.fit.cvut.cz`

Windows

- **Download and install WinSCP**
  - https://winscp.net/
  - Configure and save your connection again

Server Directories

Browse important directories on the server

- `/home/pdb241_login/`
  - **Personal directory** with all your data
    - There are no backups in place and so keep copies of everything
    - Furthermore, everything will be deleted at the end of the year
  - `assignments` subdirectory
    - Working directory for your **assignments**
  - `submissions` subdirectory
    - Copies of your **submissions** and **evaluation comments**
- `/home/PDB/`
  - **Shared directory** with course data

Homework Assignments

Individual Topics

Individual topic

- You must select an **individual topic**
  - Not later than on the day of the **first practical class**
- This topic must be **unique**
  - **First come, first serve principle** is applied
    - To avoid any conflicts
- Do not start working on the assignments without the topic
  - I.e., it must be reported and approved in advance
- Possible **examples**
  - Library, cinema, cookbook, university, …
  - Basically anything will work, do not worry
    - The more creative, the better
    - Just follow your interests

Individual Topics

Prohibited topics

- Certain topics are prohibited, though
  - Those **closely related to course materials** (lectures and labs)
  - Or those **frequently chosen in past**
- In particular, anything related to …
  - Movies, actors, TV series
  - Airports, airlines
  - Football / ice hockey / other sport clubs / leagues / matches

Individual Topics

Topic selection

- https://bit.ly/pdb-241-topics
    - **Simple interactive web form**
    - Use it for your **topic selection**
        - You can use it even repeatedly
        - Should you want to change your mind
    - Authentication
        - Use your **personal number** and **initial** **password**
- Contains an extensive **list of topic suggestions**
    - In fact, hundreds of them …
- You can even come with your **own proposal**
    - In such a case, **send it by e-mail** for approval

Submission Process

Submission steps

- **Prepare your solution**
  - I.e., create all the required files, test everything
- Open an **SFTP** / **WinSCP** connection
- **Upload all the files** to the NoSQL server
  - Use the corresponding assignment directory
    - `/home/pdb241_login/assignments/assignment/`
- Open an **SSH** / **PuTTY** connection
- **Verify your solution**
  - By following the **execution** steps
    - As described in the assignment **instructions**
- **Submit your solution** …

Submission Script

Solution submission

- Go to the assignments directory
  - `/home/pdb241_login/assignments/`
- **Execute the submission script**!
  - `submit assignment`
  - Submits everything in a subdirectory of a given name
  - This name must correspond to the assignment name
    - E.g.: `submit xquery`
- If you do not execute this script …
  - You have <u>not</u> made your submission at all!

Submission Script

Solution submission (cont'd)

- If anything goes wrong …
  - You will get a **red error message**
    - E.g.: `File <...> is required and missing`
  - You then need to remedy the situation and try again
- In other words …
  - You need to get a **green success confirmation**
    - E.g.: `Assignment <...> SUBMITTED by <...> SUCCESSFULLY`
  - Otherwise you have <u>not</u> made your submission at all!

Submission Script

Solution submission (cont'd)

- Having made your submission ...
  - You can then check it
    - Just for your convenience should you want to be extra sure
  - Go the the **submissions directory**
    - `/home/`pdb241_login`/submissions/`
  - Verify that a new submission directory was created
    - E.g.: `xquery-`pdb241_login`-20240922-163500`
    - It will contain copies of all the files you have really submitted

Test Assignment

Solution submission (cont'd)

- Try it by yourself …
  - Just to be sure you understood the whole process
    - And so you avoid any undesired surprises later on
- Just use a special **fake assignment** called `test`
  - **submit** `test`
  - Browse the contents of your `submissions` directory

Solution Requirements

Quality requirements

- Your homework solutions must be …
  - Within the **individual topic**
  - **Original**, **realistic**, and **non-trivial**

Technical requirements

- Respect **names of files** **to be submitted**
  - They are **case sensitive**
- Place all the files into the root directory of your submission
- **Do not submit anything that was not requested**
  - E.g., shared libraries, …
- Preserve **standard and error outputs** in your **shell scripts**
  - I.e., do not suppress or redirect anything

Evaluation Process

Solution evaluation

- **Number of submissions** is not limited
  - I.e., you can make even **repeated submissions**
    - Should you need to improve anything
- Only the **latest submission** **will be assessed**
  - I.e., the very last version you submitted
- **Evaluation will never start before the deadline**
  - At least unless you make an explicit request
    - E.g., at the end of the semester
  - However, it can start any time after the deadline elapses
- Each assignment is **evaluated only once**!
  - I.e., you cannot resubmit corrected and improved versions
    - Unless you need the **second chance** for the credit

Evaluation Process

Evaluation result

- Once the **evaluation is completed**, you fill find …
    - **Number of points**
        - In our **Google table**
    - **Text comments**
        - In the submissions directory on the server
        - In a text file with the same name as your submission
- If you find anything assessed incorrectly …
    - Or you just have doubts about something
    - Simply do not hesitate to **contact the teacher**

Java IDE

Java development

- We will need to work with **Java** later on during the semester
  - In particular, during the **MapReduce** class and assignment
  - It will be just a **simple command line application**
    - With one source file only
    - I.e., do not worry even if you are a **complete beginner**
- You will need a **Java IDE**
  - And so **download and install** it
- Possible options
  - **Apache NetBeans** – https://netbeans.apache.org/
  - **IntelliJ IDEA** – https://www.jetbrains.com/idea/
    - You can use the university license
    - See details at https://download.cvut.cz/en/jetbrains/