DB Agenda

Agenda DB330 Database Theory and Applications

README

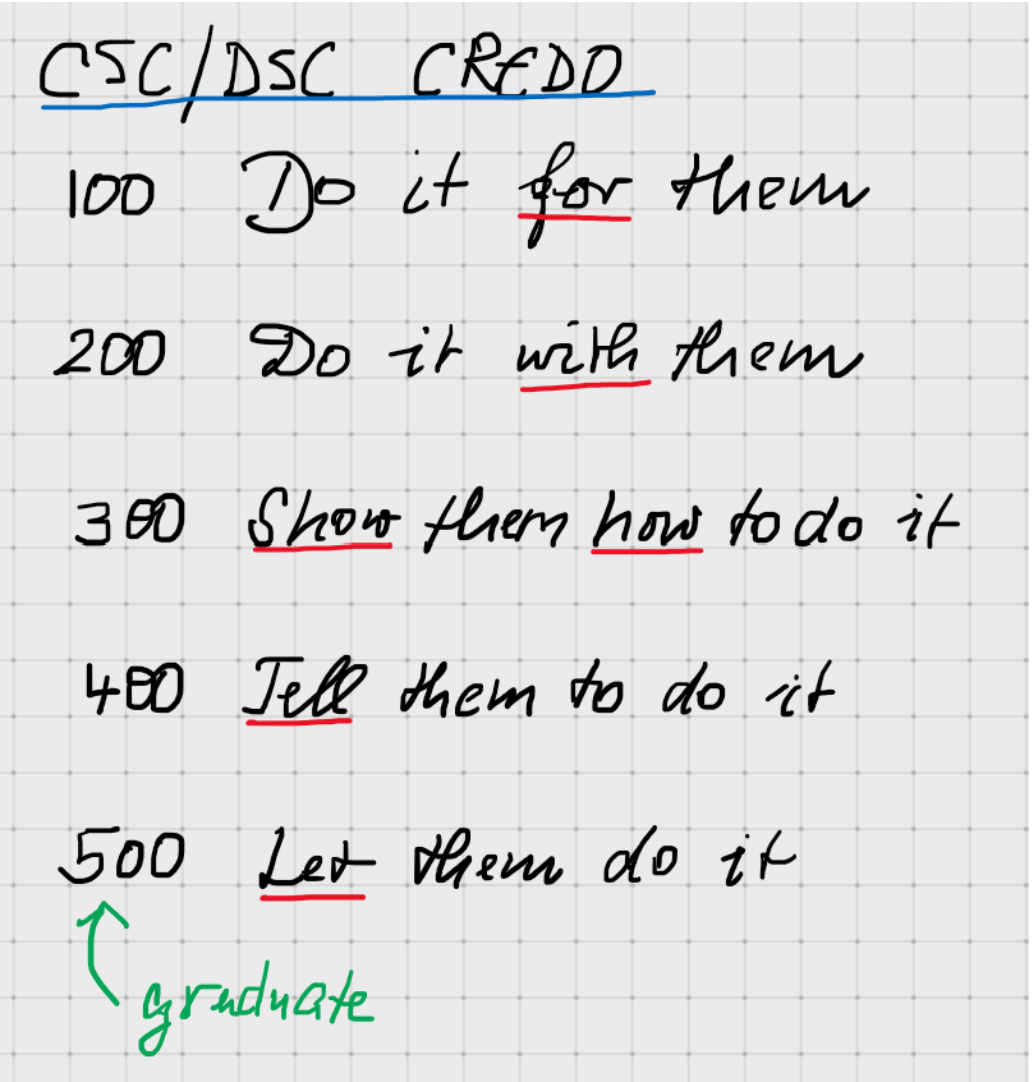
This file contains the agenda overview (what I had planned), the objectives (what we managed to do) and (much of the) content of each taught session of the course. I want to avoid splitting the content up over many files - so that you have to navigate as little as possible (like a book)!

The companion file to this file, less structured and with the captain's log, is the notes.org file.

Intro and GitHub - w1s1 - January 11, 2022

- Aspiration and ambition (Lyon Data Science program)
- Introduction to the course & the lecturer
- Homework assignments: **GitHub**, DataCamp, Emacs
- What's next?

My new didactic credo



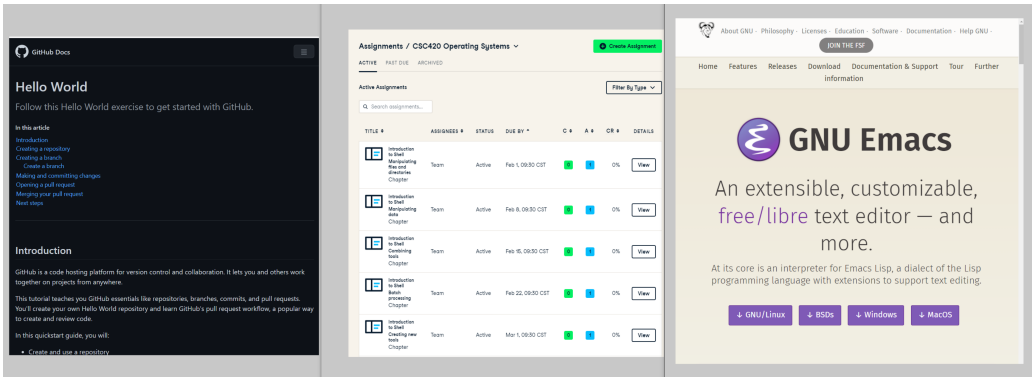
Aspirations and ambitions (DS program 2021-2023)

CLASS	CODE	TERM	Topics
Data Science Tools and Methods	DSC 101	Fall 2021	R, Basic EDA, Base R
Introduction to Advanced Data Science	DSC 205	Spring 2022	R, Advanced EDA, Tidyverse
Database Theory and Applications	CSC 330	Spring 2022	SQL, SQLite
Operating Systems	CSC 420	Spring 2022	Bash, awk, sed, regular expressions
Data Visualization	DSC 302	Fall 2022	D3, Processing, Javascript, Bokeh
Machine Learning	DSC 305	Spring 2023	Predictive algorithms, neural nets
Digital Humanities	CSC 105	Spring 2023	Data science applications

Introduction to the course & the lecturer

- PhD theoretical particle physics / WWW development
- SQL since 2005 (Why? Particle data = unstructured)
- Professor, Business Informatics @Berlin Univ
- Visiting Assoc Prof for Data Science @Lyon (2021-23)
- Syllabus for this course (Schoology)

Homework assignments week 1 (11-Jan/13-Jan-2022)



- **GitHub Hello World Exercise (Info: FAQ) - by Thursday 13-Jan!**
- DataCamp platform registration (Link: Schoology)
- GNU Emacs installation (Info: FAQ)

GitHub

- What is it?
  - Software development platform (like GitLab, BitBucket, SourceForge, etc.)
  - Built around Git by Linus Torvalds
  - Bought by Microsoft in 2018 (like OpenAI - home of GPT3)
  - 77 mio users (developers) + 200+ mio software projects

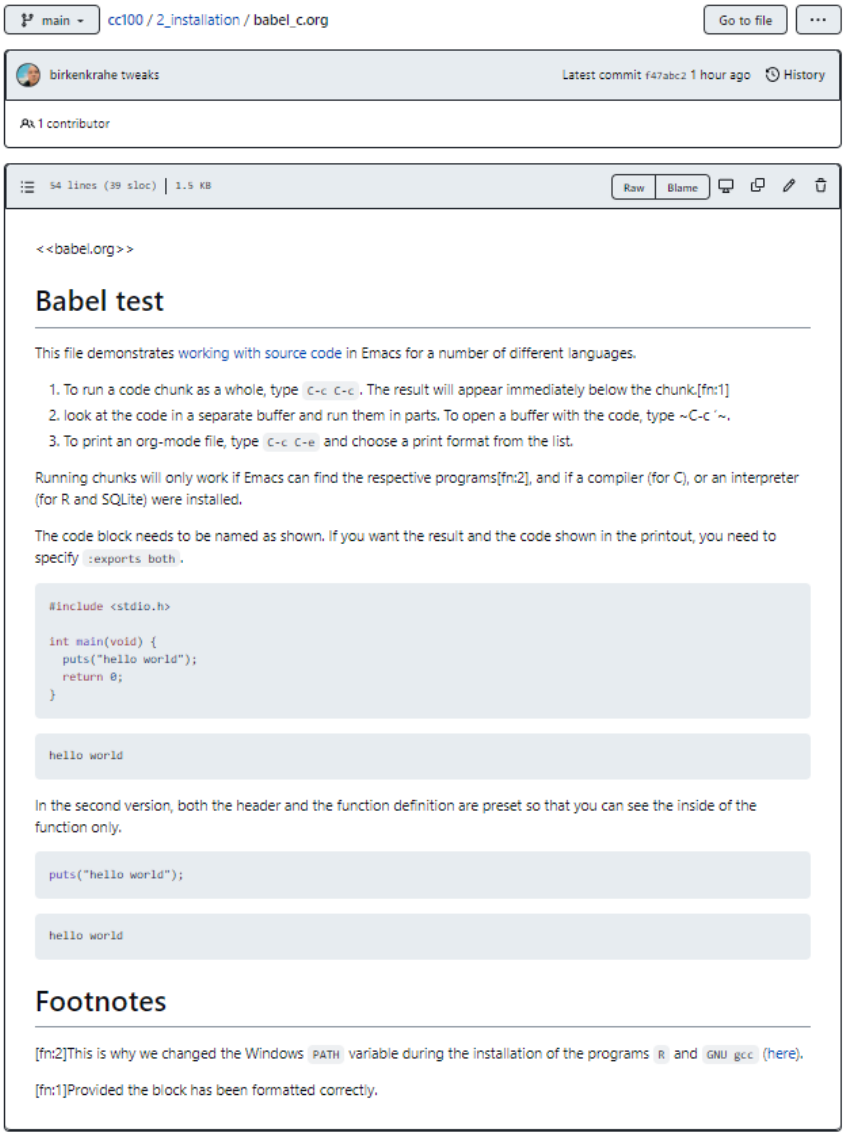
- AI support (e.g. GitHub Copilot)

Watch: "What is GitHub?" (GitHub, 2016)

Gif: "So long binder of requirements" Source: GitHub

- Why are we using it?

  Image: Org-mode file in GitHub


main cc100 / 2_installation / babel_c.org Go to file

birkenkrahe tweaks Latest commit f47abc2 1 hour ago History

1 contributor

54 lines (39 sloc) | 1.5 KB Raw Blame

<<babel.org>>

Babel test

This file demonstrates working with source code in Emacs for a number of different languages.

1. To run a code chunk as a whole, type `C-c C-c`. The result will appear immediately below the chunk.[fn:1]
2. look at the code in a separate buffer and run them in parts. To open a buffer with the code, type ~C-c '~.
3. To print an org-mode file, type `C-c C-e` and choose a print format from the list.

Running chunks will only work if Emacs can find the respective programs[fn:2], and if a compiler (for C), or an interpreter (for R and SQLite) were installed.

The code block needs to be named as shown. If you want the result and the code shown in the printout, you need to specify `:exports both`.

```
#include <stdio.h>

int main(void) {
  puts("hello world");
  return 0;
}
```

```
hello world
```

In the second version, both the header and the function definition are preset so that you can see the inside of the function only.

```
puts("hello world");
```

```
hello world
```

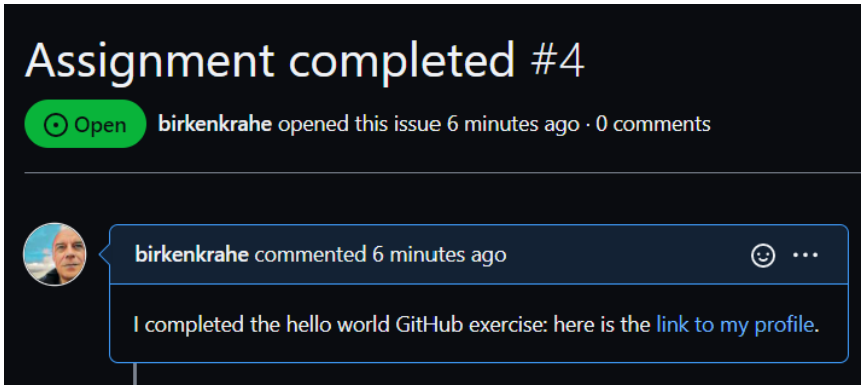
Footnotes

[fn:2]This is why we changed the Windows `PATH` variable during the installation of the programs `R` and `GNU gcc` (here).

[fn:1]Provided the block has been formatted correctly.

  - It's free
  - To host course materials
  - Upload assignments (esp. Emacs Org-files)
  - Discussion
  - Wiki for collaboration
  - Complements Schoology
- What will you have to do?
  - Sign up with GitHub - use Lyon Email
  - Pick an available username **using your own first and last name**, e.g. `MarcusBirkenkrahe`, or `DonaldTrump`
  - Complete the "Hello World" exercise (FAQ)
  - Give me your GitHub username so that I can add you as a collaborator to my private `db330` repository
  - Create an issue from the `db330` repository like in the example below (except from your account instead of mine).

Image: Issue "Assignment completed"



If you do have a GitHub account already, do the exercise anyway using your existing account (it takes 10 min)! Make sure you let me know what your user name is so that I can add you to my repo.

- What else can you do?
  - You can fork the `db330` repository
  - You can watch the `db330` repository - and set `Notifications` to `Participating and @mentions` so that you see my comments (see image below).

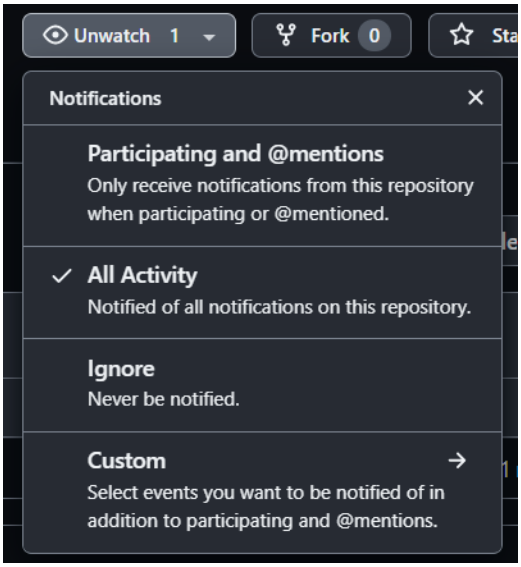
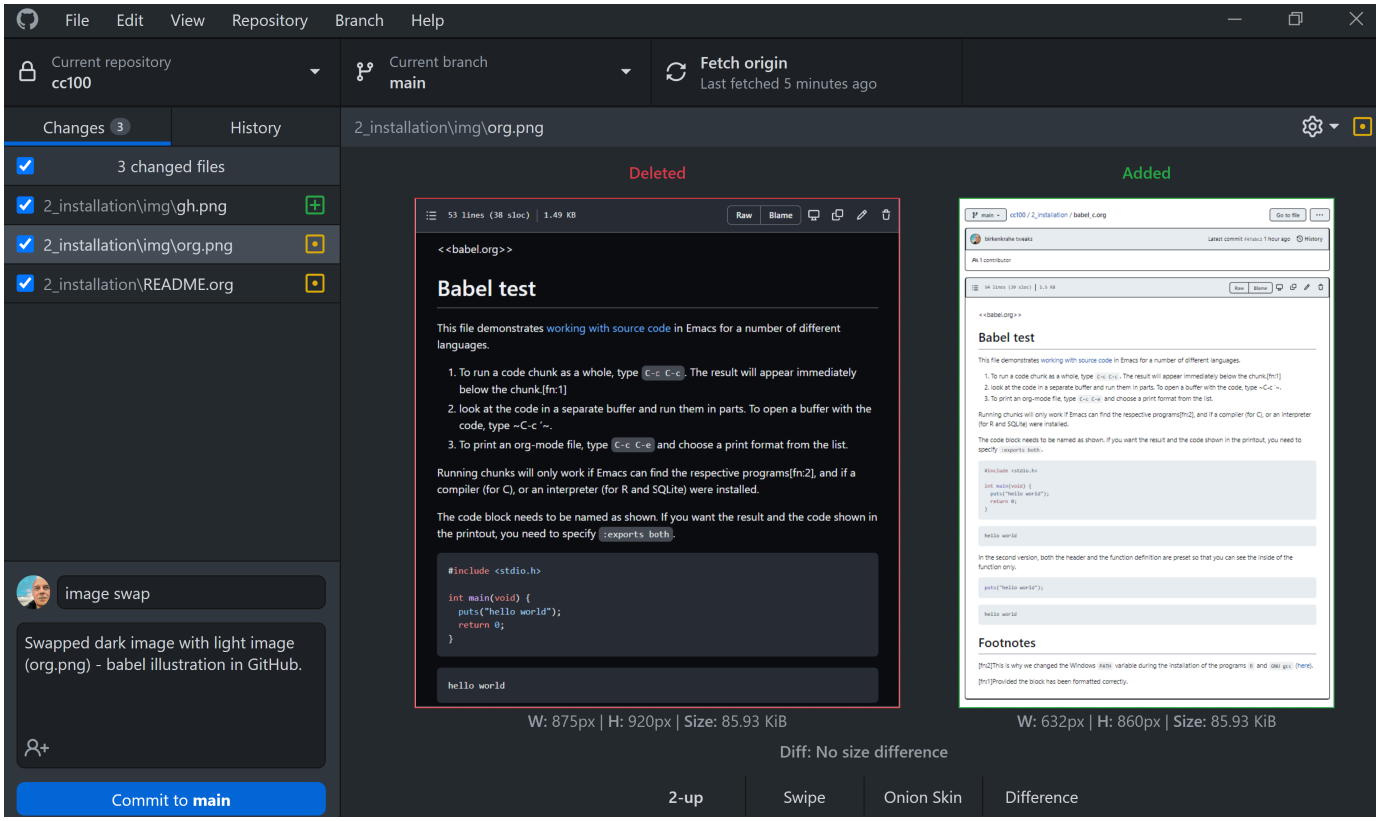


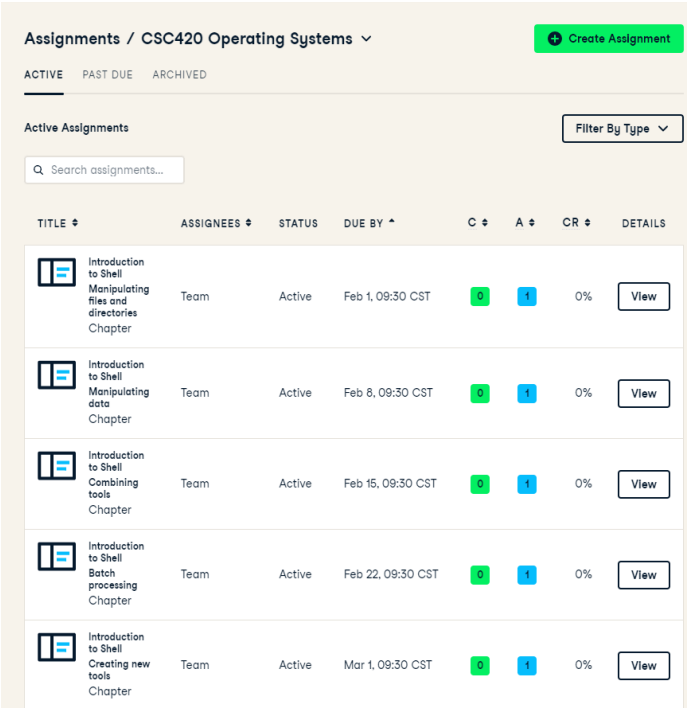
Image: Notifications settings when watching a repository

  - You can submit issues from the repository (e.g. if you notice mistakes or if you want extra information, or to share a link)
  - You can participate in discussions (sometimes I will make you)
  - You can add to the wiki (e.g. comments and links to interesting resources)
  - You can install the mobile app on your smartphone
  - You can use it as a platform for projects or coding
  - You can download the desktop client to manage repos on your PC (see image below).

Image: GitHub desktop client commit



DataCamp



- Why are we using it?
- How are we using it?

- What will you have to do?

GNU Emacs (1976…1985)

- Why are we using it?
  - To mix documentation + code + output = literate programming (1984)
  - It's the same thing as an interactive computing notebook (Jupyter) … except open to ALL languages and outputs
- How are we using it?
- What will you have to do?

What's next?

- See schedule (GitHub)
- Watch online lecture on "Systems"
- Everything else = online summary
- See you (hopefully) Thursday in class! (Lyon 104)

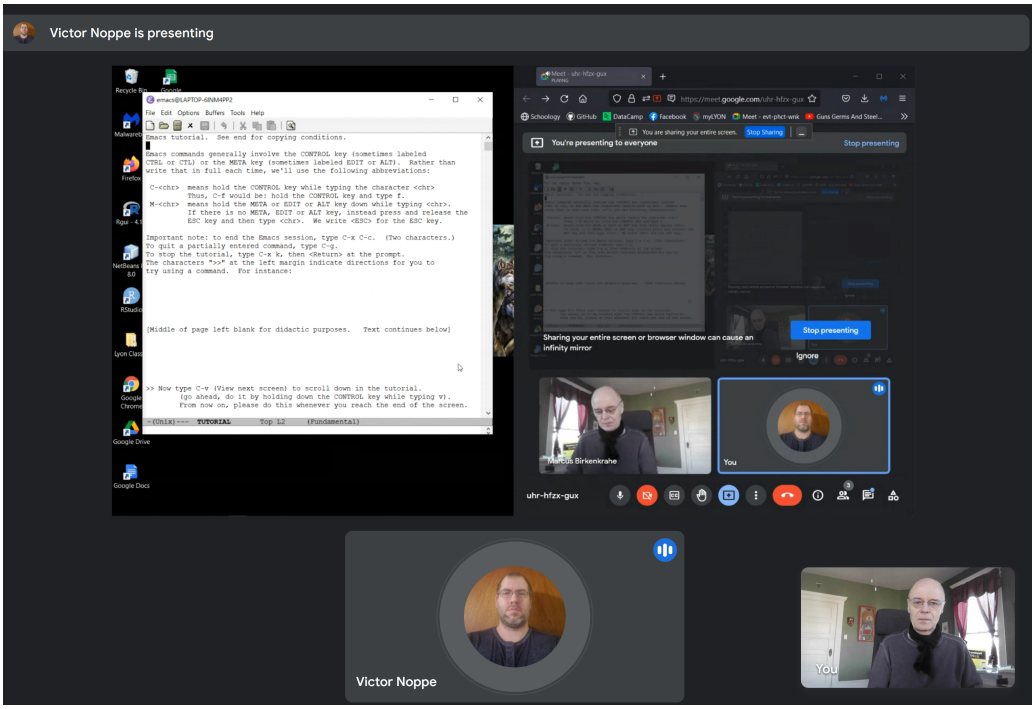


Figure 10: Victor N. installs Emacs @10:50 AM

DataCamp, History of DB, MooCall - w1s2 - January 13 , 2022

Overview

HOW	WHAT
Review	GitHub Hello World exercise (FAQ)
Video	History of Databases (Codd, Oracle, IBM)
Lecture	DB applications and basic definitions
Application	MooCall Calving Sensor App (IoT network)
Practice	Install GNU Emacs (ESS or vanilla) (FAQ)
Demo	Emacs guided tour
Self	Work through the Emacs onboard tutorial

Objectives

- [X] Review the basics of Git and GitHub
- [X] Know basic definitions of database management systems
- [X] See a current example of an IoT-based DB application
- [X] Install the GNU Emacs editor on your OS
- [ ] Understand how GNU Emacs works
- [ ] Make GNU Emacs work for you

DB elements, GNU Emacs - w2s3 - January 18, 2022

Overview

HOW	WHAT
Review	Quiz 1: Intro to course / databases
Lecture	Elements of Database Systems
Demo	Emacs guided tour
Self	Work through the Emacs onboard tutorial

Objectives

- [X] Review last week's content with a quiz
- [X] Review file vs. db approach to data management
- [X] Learn about the elements of a database system
- [X] Understand DB system design structure/users
- [X] Understand how GNU Emacs works (guided tour)
- [ ] Make GNU Emacs work for you

What's next?

- Take a look at the Emacs tutorial (CTRL-h t)
- GNU Emacs practice exercises (in class)
- DataCamp assignments beginning next week (online)

SQLite installation - w2s4 - January 20, 2022

Overview

HOW	WHAT
Demo	Installing `sqlite`[1] (done[2])
Practice	Emacs guided tour (tour)(web-tutorial)
Self	Work through the Emacs onboard tutorial

Objectives

- [X] Understand how GNU Emacs works (guided tour)
- [X] Make GNU Emacs work for you (tutorial)

What's next?

- Architecture and classification of databases
- DataCamp assignments due next week (online)
- Cloud computing - relevance for databases

Cloud Computing for Everyone Cloud Computing for Everyone Introduction to Cloud Computing I forgot to mention this in the last class. This is the first DataCamp assignment. It is informational and very simple - nothing but drag and drop practice and a few videos. This should not take you longer than 15-20 minutes.

We'll pick up on past assignments in class - short review including questions for the audience (you!) Completing this assignment on time gets you 10 points (100%). Late completion (after the due date): 5 points (50%).

Cloud computing intro - w3s5 - January 25, 2022

Overview

HOW	WHAT

HOW	WHAT
Review	Quiz 2 - database foundations / Emacs
	DataCamp assignment 1: cloud computing intro
Quiz (opt)	Inside Google's Data Center (Thu)
Demo	Emacs Org-/ code blocks
Practice	Create literate Org-mode file
	Run `sqlite` program in Emacs
Assignment	Do this on your PC (extra credit by Thu 1PM)

Objectives

- [X] Review db foundation and GNU Emacs (quiz 2)
- [X] Review introduction to cloud computing (DataCamp 1)
- [X] Understand GNU Emacs Org-mode
- [X] Know how to create a literate Org-mode file
- [X] Know how to run a literate Org-mode file

DataCamp assignment: Cloud Computing Introduction

- What's the main message of this lesson?
- What does cloud computing have to do with databases?
- What did you think about the assignment?
- What did I think about the assignment? (see notes)

Figure 11: Schoology assignment, Jan 21, 2022

What's next?

- Architecture and classification of databases (book)

- Next DataCamp assignment due Feb 1 ("Cloud deployment")
- Org-file assignment (in Schoology) coming your way

Cloud deployment - w4s6 - February 1, 2022

Overview

HOW	WHAT
Review	Quiz 3 - data centers / cloud computing / metadata
	DataCamp assignment 2: cloud deployment
	Emacs Org-mode assignment (see tutorial videos)
Lecture	Database foundations - 10 tenets
Practice	SQLite basics - creating/importing a database

Objectives

- [X] Review quiz 3 - how should you study, learn and rehearse?
- [X] Review deployment of cloud computing services (DataCamp 2)
- [X] Review database foundations (10 tenets)
- [X] Opened, closed SQLite (DBMS) and wrote on a db (see notes)

DataCamp assignment: Cloud Deployment

- What's the main message of this lesson?
- What infrastructure is required? Can you do this yourself?
- What're the greatest challenges of deployment?
- What did I think about the assignment? (notes)

Database management foundations - 10 tenets

1. A database (DB) is a collection of related data items within a specific business process or problem setting.
2. A database management system (DBMS) is the software package used to define, create, use and maintain a database.
3. We distinguish the file-based vs. DB approach to data management (data from different application stored in different files vs. managed by one application and one shared, central database)
4. Metadata are data (structure) definitions, like ownership or number of tables, and are stored in the DB catalog or dictionary.
5. DBMS provide DB languages (like SQL) that facilitate data definition (DDL), data manipulation (DML), data querying (DQL), and data control (DCL).
6. The database model (or schema) describes the DB data structure. It does not change easily, and is stored in the catalog. Examples are: entities stored (e.g. as tables), and entity aspects (e.g. as columns). To model, we use ERDs.
7. The database state, or set of instances, represents the data in the DB at a given moment in time, viewed using the DB language (DQL). Examples are DB records (rows) from subsequent observations.
8. DB follow a 3-layer design: an internal technical layer (e.g. server, file, network organisation), a conceptual/logical layer (e.g. the schema), and an external layer (views of the data provided e.g. by SQL queries).
9. There are different types of DB users, with different skill sets: information (cloud) architects; DB designers (ERD); DB Administrators (SQL, Linux); Application Developers (e.g. Web App); Business Users (SQL).

Practice: SQLite Basics

We'll learn more about SQLite in future sessions. This is just to get our feet wet, including some important file system aspects.

```
$ sqlite3
sqlite> .database
sqlite> .q
$ touch test.db  # this only works if you have 'touch' installed, e.g.
                 # via the cygwin utility bundle
$ fsutil file createNew test.db 0 # creates an empty file in ./
$ sqlite3 test.db
```

```
sqlite> .database
sqlite> .q
```

As you can see, it's never easy to do anything in Windows. We're better off writing SQLite code in Emacs where the `.db` file is automatically created (see assignment). If this doesn't work for you, contact me and we'll sort you out!

What's next?

- ~~Architecture and classification of databases (text book ch 2)~~
- Next DataCamp assignment due Feb 8 ("Cloud providers/cases")
- New DataCamp assignment due Feb 15 ("Introduction to SQL: SELECT")
- SQLite DDL practice

SQLite introduction - w5s7 - February 8, 2022

Overview

HOW	WHAT
Review	Cloud Providers and Case Studies
Lecture	Introduction to SQLite
Practice	Exploring `sqlite3`
Test info	Test 1 on Thu 10 Feb 1.30-2.15 pm

Objectives

- [ ] Understand setup for test 1 (online in class)
- [ ] Review DataCamp assignment
- [ ] Understand what SQLite is and why it's important

Test 1 info

- Online in Schoology
- Quiz 1-3 are not visible during the test
- The 10 hardest questions of quiz 1-3 (< 50%)
- 10 brand new questions
- Maximum time = 45 min

DataCamp assignment: Cloud Providers and Case Studies

- What's the main message of this lesson?
- How does one pick a cloud provider?
- What did I think about the assignment?

Introduction to SQLite

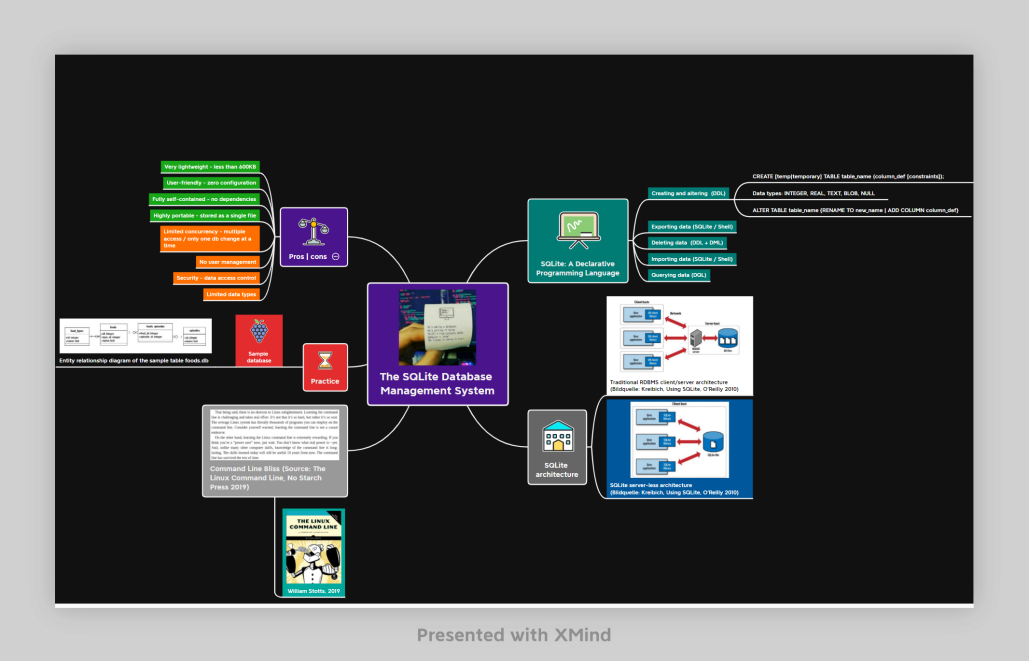


Figure 12: Introduction to SQLite (mindmap)

IN PROGRESS Practice: SQLite Basics (cont'd)

- We last looked at entering SQLite (notes)
- Today, we look at a few more commands:

```
$ sqlite3 sqlite.db  # starts SQLite on the Command shell
$ sqlite3 -help      # options list
sqlite> .databases   # prints current persistent database
sqlite> .show        # display and I/O options
sqlite> .tables      # check for tables
```

What's next?

- Test 1 - Thursday 1.30-2.15
- New DataCamp assignment due Feb 15 ("Intro to RDBM with SQL")
- SQLite DDL practice

References

- Lemahieu et al (2021). Principles of Database Management. Univ of Cambridge Press. URL: pdbmbook.com.
- TutorialCup (n.d.) System Catalog [website]. URL: www.tutorialcup.com.

Footnotes:

[1] For installation on your PC, see these instructions (PDF). The installation is simple: download the ZIP file, unpack it, and set the PATH variable so that sqlite can be found.

[2] In Lyon 104 (Computer lab), the PATH variable is unfortunately stored in the personal user's app data so that I could not set it properly for your account (you can do this on your own PC easily). However, I seem to have found a solution for this (see course FAQ).

Author: Marcus Birkenkrahe
Created: 2022-02-08 Tue 20:37
Validate