Spring 2022 courses

DONE Quiz 1: Introduction to the course and to databases

This is the first weekly quiz. Time = 5 minutes, 10 points max.

Literate programs contain documentation + code + output

TRUE

> There is more to literate programming, but this is what you should remember. Programming in this way, e.g. by using interactive notebooks, directly helps you develop your computer and data literacy.

Match GitHub action and definition

ACTION	MEANING
Create repository	Place for your code
Create branch	Create a modified copy of a main repo
Open pull request	Request for changes of main repo
Merge branch	After review, merge changes into main repo
Submit issue	Make a comment on a repository
Commit change	Make change permanent

How will this course be graded?

TRUE:

- Class assignments: 10%
- Weekly quizzes: 10%
- Monthly tests: 30%
- DataCamp assignments: 30%
- Final exam: 20%

FALSE:

- Attendance: 10%
- Group project: 20%
- Essay: 20%
- Mid term exam: 30%
- Presentation: 20%

What will you learn in this course?

TRUE:

- Learn relational database query languages
- Learn how to use and program the shell
- Learn how to design database applications

FALSE:

- Learn how to hack your way into databases
- Learn how to make predictions from public datasets
- Learn how databases relate to machine learning

What's GitHub?

TRUE:

- Platform for software developers
- Owned by Microsoft
- Ca. 77 mio users + 200 mio software projects

FALSE:

- Version control program | Git is the version control software, which runs on GitHub
- Created by Linus Torvalds | Torvalds wrote Git, not GitHub
- Ca. 7 mio users + 20 mio software projects | You're off by a factor 10

  GitHub is a cloud platform for developers - there are many other platforms like that, all using the free, open source software Git, but GitHub is probably the largest.

How does MooCall sensor operate?

Bring the following processes in the right order.

1. Edge computing: data are generated and pre-processed locally
2. Cloud computing: data are processed globally to generate a signal
3. Local computing: signal is transmitted to the user

Feedback: The sensor is attached to the cow's tail. It records data as signals in the form of temperature, motion, etc. In relation to the cloud, the sensor (a microcontroller with minimal operating and processing capabilities) operators "on the edge" (of the cloud). It has likely an SQLite database on board, or perhaps something even simpler. Gathering the data and writing them to a file if only for transport to the cloud is the "preprocessing". In the next step, the signal for the end user is prepared once the data (from different cows, and over a longer period of time) are now "in the cloud". In practice this means that they are now in a database on a server, very likely in tabular form. The processing in the cloud generates a signal for the end user when the evidence is conclusive, and when the software used to process the data has reached a positive prediction: "Cow is ready to calve". Only then the end user, the farmer, will be notified. In another scenario, the farmer is continuously informed about the state of the cow. But this does not change anything in the data processing pipeline.

What's a database system?

TRUE:

- Database management system + database

FALSE:

- Database + user application
- File system + database application

  The database management system (DBMS) sits between the user application and the data storage. Data management based on a file system predates the development of DBMS software.

A database is defined within a specific business process or problem setting

TRUE

Data themselves are independent of their use - they simply come from the real world. A database is a rational, systematic attempt to store the data, and reflects a specific purpose - beginning with the type of data to be stored (e.g. text data, numerical data, video data etc.)

Which data items should be stored in the database?

As an example scenario, assume the "MooCall" sensor application, which we looked at and discussed in class, to support cow calving.

We're looking for entities that make sense for a MooCall database given its purpose.

TRUE:

- Cow ID
- Cow temperature
- Cow location and motion data
- Farmer's phone number
- Farmer's user ID and password

FALSE:

- Size of cow pasture
- Cow weight
- Cow height

  ID, temperature, and geo data, are all locally obtained through the sensor. User data are necessary for the signal submission (via SMS). Pasture size, weight and height of the cow may be relevant for some other setting but they cannot be obtained through the sensor and either not directly relevant to calving (pasture size), or not volatile enough (they change too slowly for the setting of interest).

What's part of the 3V definition of Big Data?

TRUE:

- Data volume
- Data variety
- Data velocity

FALSE:

- Data value
- Data veracity

  Data "value" and "veracity" are often mentioned as part of a "5V" definition of big data. As discussed, they are too volatile, and not directly measurable, and therefore less useful for a definition.

Author: Marcus Birkenkrahe
Created: 2022-03-17 Thu 15:18
Validate