

RAW DATA

Portfolio Project 1
Introduction

Henrik Bulskov & Troels Andreasen

RAWDATA Project portfolio

- an independent activity
- closely linked with the course

❑ **Four subprojects**

- covered in the four corresponding sections of the course
- partly planned already
- requirements (not necessary consistent) will initiate each
- requirements are given by the **project owners: your teachers** 😊

❑ **One Reflexive synopsis**

- Discussing and summarizing the Project portfolio

❑ ***Hand-ins: 4 project reports + 1 reflexive synopsis report***

The Project Portfolio: Four subprojects

❑ Portfolio project 1: Database

- design and implement databases.

❑ Portfolio project 2: Web Service

- design and implement web services to access and manipulate data in databases implemented in Portfolio project 1

❑ Portfolio project 3: Information Retrieval

- draw on aspects of Information Retrieval to improve search and retrieval of data from our databases

❑ Portfolio project 4: Responsive applications

- develop applications that build on web services developed in Portfolio project 2 and extended in Portfolio project 3

The Project Portfolio

❑ Project Portfolio – Problem & Domain

- provide a tool to **help computer programmers** develop skills while they are working

❑ Key source of data

- The Q&A site **Stack Overflow**
- We will use a publicly available dump of data from this site

❑ What is Stack Overflow?

- A question and answer (Q&A) site for programmers,
- you can search the knowledge captured in the answers and comments given to the more than 16 million questions
- answers are ranked and generally of high quality

What is Stack Overflow?

pgAdmin copy database Search

124 results Relevance Newest More ▾

590 votes Q: Creating a copy of a database in PostgreSQL

What's the correct way to **copy** entire **database** (its structure and data) to a new one in **pgAdmin**? ...

16 answers

0 votes Q: I was looking for information around this

1

Creating a copy of a database in PostgreSQL

▲ What's the correct way to copy entire database (its structure and data) to a new one in pgAdmin?

590 postgresql

▼ share improve this question

★ 185 add a comment

edited Jan 11 '15 at 21:17 frian 4,497 ● 3 ● 14 ● 52

asked May 18 '09 at 7:00 egaga 11.2k ● 9 ● 40 ● 54

16 Answers

active oldest votes

▲ 914 Postgres allows the use of any existing database on the server as a template when creating a new database. I'm not sure whether pgAdmin gives you the option on the create database dialog but you should be able to execute the following in a query window if it doesn't:

▼ `CREATE DATABASE newdb WITH TEMPLATE originaldb OWNER dbuser;`

The Project Portfolio goal: SOVA application

❑ Our goal: a Stack Overflow Viewer Application (SOVA)

- 1) **Search for posts and comments** in Stack Overflow
- 2) **Present search results** by lists of items that link to (or can be unfold for) details
- 3) Keep track of **search history**
- 4) Provide a **marking** option for posts of special interest among posts presented in the search result and allow optional **annotation** to marked posts
- 5) Support **multiple users** such that each user only has access to own history, markings and annotation
- 6) Deal with **relevance** of objects to queries and use this **to provide better answers**
- 7) Present **search results** with focus **on words frequencies** rather than on the texts they occur in, such as ranked lists of most frequent words
- 8) Provide **visualizations** of search results and subsets of the database by means of **word clouds** and **word graphs** showing significant words and their relations

❑ Open-ended set of additional features

- see examples Portfolio 1 description
- ...
- Plenty of room for your own ideas and preferences here
- ...

however

The Project Portfolio goal: SOVA application

- ❑ Our goal: a Stack Overflow Viewer Application (SOVA)

- ❑ Challenge

 - to decide on a **small** but well chosen **set of features**

- ❑ we are **NOT** aiming for

 - the full functionality of StackOverflow

A multilayer architecture

❑ The presentation layer

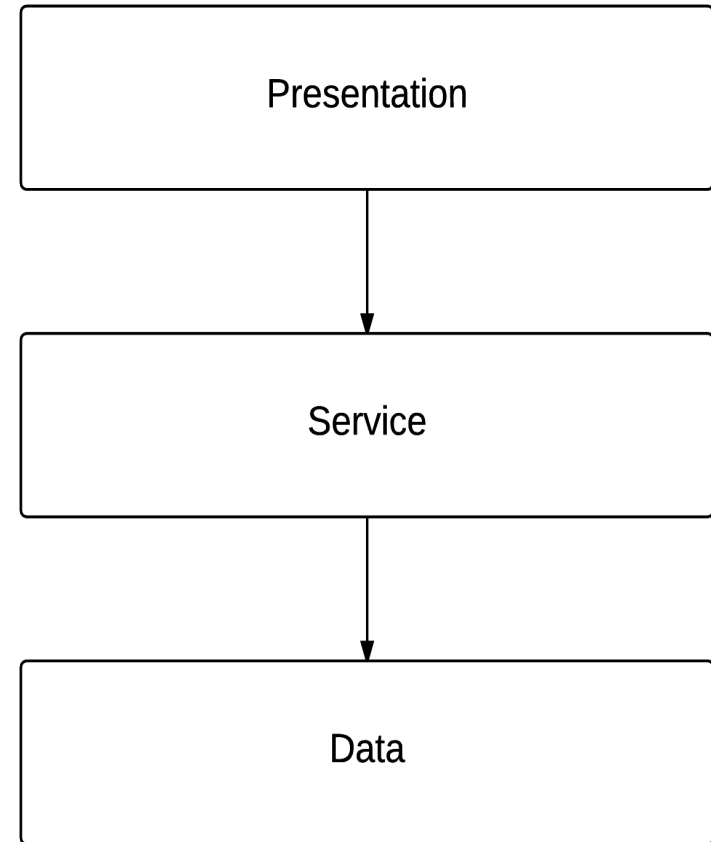
- A web based responsive application.

❑ The service layer

- defines the application logic.
- provide an interface to the presentation layer through web services.

❑ The data layer

- encapsulates storage and retrieval of data
- expose basic functionality related to this



Portfolio project 1

❑ The goal

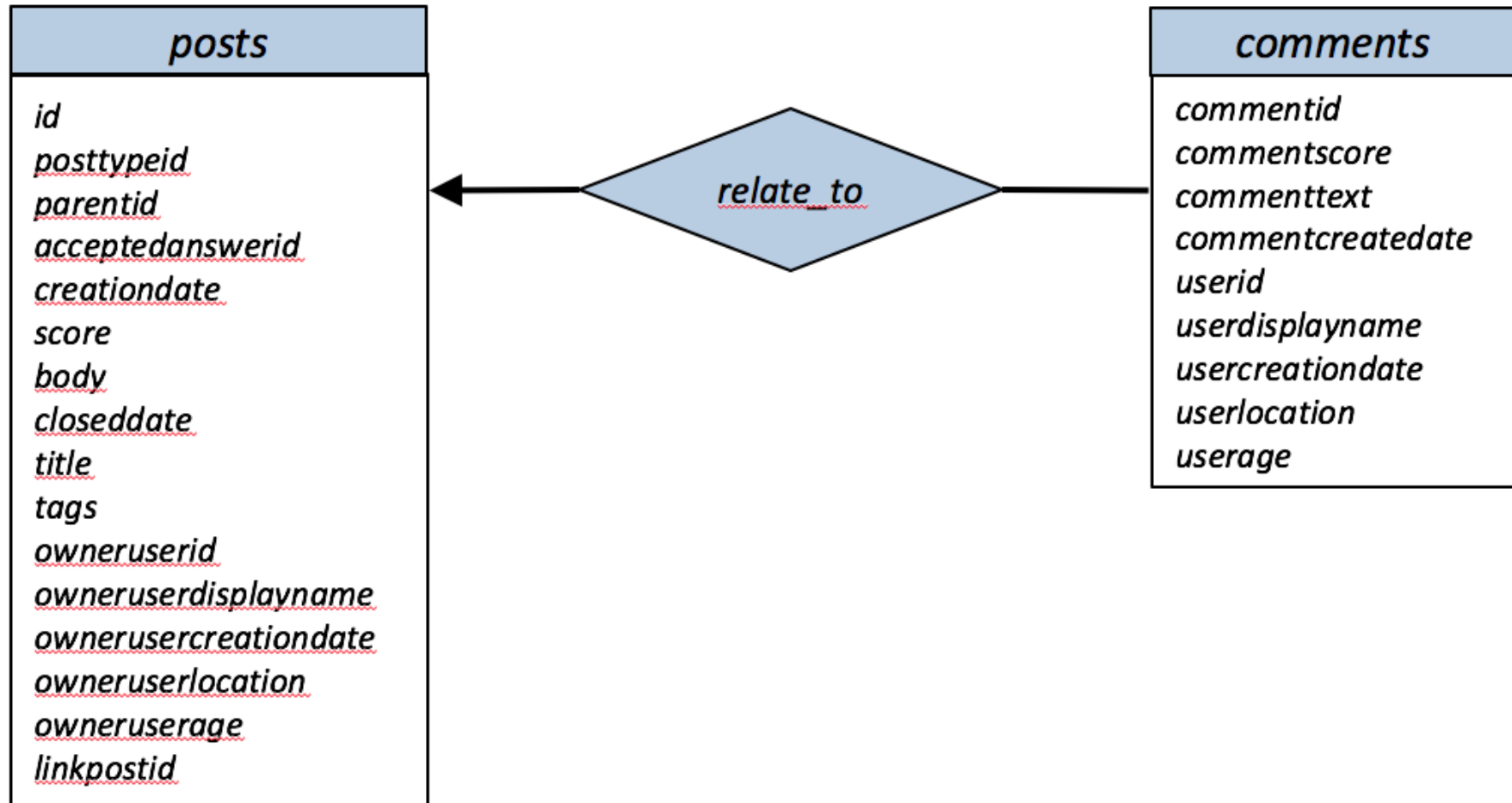
- to provide a database for the SOVA application and to develop an API with key functionality of the application.
- two independent data models
 - a QA-data model (to represent data from Stackoverflow) and
 - a framework model (users, markings, annotations, history)
- provide support for
 - browsing and search for answers and
 - individual for users: activity history, marking and annotation on posts of special interest

Starting point

- ❑ Data provided in a two-table database including a **comments** and a **posts** table, on moodle:

- [stackoverflow_universal.backup](#)

Model for this simple database:



What to do

❑ A. Application design

- Sketch a preliminary design of your application

❑ B. The QA-model

- Develop a good design so that
 - all data from the source database (stackoverflow_universal.backup) can be represented
 - your own preferences regarding search functionality are met

❑ C. Framework model

- Design a (complementary) model to meet
 - the basic requirements of the history, marking, annotation on individual user basis of your application

❑ D. Functionality

- Design and implement key functions to be exposed as API from the data to the service layer

❑ E. Testing

- Demonstrate by examples that the results of D work as intended. (More elaborate testing later)

The project report

❑ Work in groups

- each group:
 - one Portfolio Project 1 report with appendices
 - a selection of script to redo, what you have done
 - a product, including the database and the implemented functionality, implemented on rawdata.ruc.dk

❑ The report

- size around 6-10 normal-pages (8*2400 characters) excluding appendices
- your Portfolio project 1 report is not supposed to be revised later
- your product may be subject to revision later if documented in report 2, 3 or 4

❑ The submission deadline

- for the report as well as the product is 3/10-2018.