

PS Non-Standard Database Systems

Summer term 2018

Checkpoint 01 Project Proposal

Due date: Friday, 2018-04-06, 23:55

1 General

Submit your checkpoint report until **Friday, 2018-04-06, 23:55** using our submission system ¹. Please be aware that only the **last** submission is assessed.

1.1 Support

If there are any ambiguities or problems of understanding regarding the checkpoint, you have the following possibilities to clarify them:

1. Slack channel: **#nsdb** ² (preferred way of communication)
2. Office hours: Wednesday, 10AM - 11AM, room 0.26 (ground floor)

If you run into a problem, first, *try to resolve it yourself (as a group)*. If the problem remains unresolved, you should use one of the above possibilities *in time* in order to allow best possible support by the instructor.

2 Task Description

For this checkpoint, you will specify the project you are going to work on throughout this semester. This includes (1) a survey of different types of non-standard database systems, (2) the choice of the database system you would like to work with, (3) providing a concise, motivating application that is a good fit for the chosen database system, (4) finding datasets to fill your database, and (5) a roadmap for your project (*optional*).

¹<https://abgaben.cosy.sbg.ac.at>

²<https://dbteaching.slack.com>

2.1 Survey

The first subtask is to explore literature and other trustworthy (online) resources for information on different non-standard database systems. In our context, a *non-standard* database system is all but a *traditional relational* database system (e.g., PostgreSQL). This also includes processing frameworks which are not database systems per se but are closely related since they process Big Data. Afterwards, you should have a good overview on the landscape of available non-standard database systems.

Enumerate *four* references in the report and briefly summarize the main insight they provided to you.

Note: You may want to check out the suggested readings section (*Unterlagen*) on our course website

- to have a starting point for your investigation, and/or
- to find additional material/pointers on the topics covered in this class

Especially the *overview on non-standard database systems* may be helpful to get an overview on available types of systems.

2.2 Choose a System

After reviewing the literature and (some of) the respective systems, you have to choose one (open-source/freely-available) system. This system will form the basis of your project.

Motivate your choice: Why is this type of system interesting to you? Also name and (briefly) describe *four* key/interesting features/properties/capabilities of the chosen system.

2.3 Application Description

Now that you have chosen a system, you need a suitable application that benefits from using this system. Try to find such an application. This can be anything, for example, an everyday problem you always wanted to solve or a problem you encountered at work. However, make sure it is somewhat meaningful and not just a dummy application. Try to define the workload(s) your system has to deal with (e.g., OLTP, OLAP, ...). Describe your application and reason why this application/workload fits your system of choice. If possible, also provide an architectural overview: Which programming language will you be using? Do you plan to use any additional frameworks/libraries (e.g., for visualization)? If this is the case, briefly describe their role in your application pipeline. Will you deploy your application in a real/simulated distributed environment?

Experimental Data You will also need some data to test and evaluate your application. Basically, you have two choices here:

1. Find data; most probably online (e.g., Twitter)
2. Generate data that fits your needs.

In both cases, you are required to describe important properties of the experimental data and why this dataset is good to test and evaluate your application. If you find multiple interesting datasets, you can receive a bonus point.

You are not required to use datasets that are too large for a single machine. However, for the sake of a meaningful evaluation of your application, the datasets should not be too small.

2.4 Roadmap

Note: This subtask is optional. However, you can earn a bonus point if you work on this subtask.

Provide a roadmap for your project: Think about the actual implementation and try to identify necessary steps in order to reach the goal. The roadmap does *not* necessarily need to state explicit dates. In any case, you will benefit from planning your project before you start working on it.

3 Grading

	Category	Max. points
2.1	References	0.125 each (max. 0.5)
	Short summaries	0.125 each (max. 0.5)
2.2	Key features of chosen database system	0.25 each (max. 1)
	Description of application	1
2.3	Reasoning	1
	Description of experimental dataset	1
Bonus	Additional datasets	1
	Roadmap	1
	Max. points	5+2

4 Feedback

Note: Answering the feedback questions is optional.

You can help us to improve this class (even for this semester). Therefore, you are asked to answer the following questions:

- Was this checkpoint too easy/hard in any regard? Did it take too much/little time? Please indicate the average time per group member that was spent working on this checkpoint. The time that you indicate will have *no* impact on your grade.
- Are there any hints/references we should give future students? Conversely, did you find any of our guidance misleading?
- Do you have any suggestions for the instructors to more effectively support students?
- Any other comments?

If you would like to provide your feedback anonymously, you may also answer these questions in the evaluation at the end of the semester.