

# PS Non-Standard Database Systems

Summer term 2019

## Checkpoint 01 Project Proposal

Due date: Thursday, 2019-03-28, 23:55

---

### 1 General

Submit your checkpoint report until **Thursday, 2019-03-28, 23:55** using our submission system<sup>1</sup>. 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**<sup>2</sup> (preferred way of communication)
2. Office hours: Upon request via email (dkocher@cs.sbg.ac.at)

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/processing system you would like to work with, (3) providing a concise, motivating application that is a good fit for the chosen database/processing system, (4) finding data sets to process and/or fill your database, and (5) a roadmap for your project (*optional*).

---

<sup>1</sup><https://abgaben.cosy.sbg.ac.at>

<sup>2</sup><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 include all but *traditional relational* database systems (e.g., PostgreSQL). This also includes processing frameworks that are no database systems per se but are closely related (e.g., Big Data processing systems). Afterwards, you should have a good overview on the landscape of available non-standard database systems.

Enumerate *four* references in the report (0.125 points each; max. 0.5 points in total) and briefly summarize the main insight they provided to you (0.125 points each; max. 0.5 points in total).

*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 database/processing systems*.

## 2.2 Choose a System

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

*Motivate your choice:* Why is this type of database/processing system interesting to you? Also name and (briefly) describe *four* key/interesting features/properties/capabilities of the chosen database/processing system (0.25 points each; max. 1 point in total). Be aware, understanding the details of the underlying database/processing system is the main objective of this class (rather than, for example, designing a nice user interface).

## 2.3 Application Description

Now that you have chosen a database/processing system, you need a suitable application that benefits from this kind of system. Try to come up with 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 meaningful and not just a dummy application. Try to define the workload(s) your application has to deal with (e.g., OLTP, OLAP, ...). Describe your application (1 point) and justify why this application/workload fits your database/processing system of choice (1 point). If possible, also provide an architectural overview: Which programming language will be used? 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 <sup>3</sup>, Kaggle <sup>4</sup>, Google Data Set Search <sup>5</sup>, ...)
2. Generate data that fits your needs.

<sup>3</sup><https://developer.twitter.com/en/docs.html>

<sup>4</sup><https://www.kaggle.com/>

<sup>5</sup><https://toolbox.google.com/datasetsearch>

In both cases, you are required to describe important properties of the experimental data and why this data set is good to test and evaluate your application (1 point). If you find multiple interesting data sets, you can receive a bonus point (1 bonus point).

You are not required to use data sets that are too large for a single machine. However, for the sake of a meaningful evaluation of your application, the data sets 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 (1 bonus point): 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 Scheme

	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/processing system	0.25 each (max. 1)
2.3	Description of application	1
	Reasoning	1
	Description of experimental dataset	1
Bonus	Additional datasets	1
	Roadmap	1
	<b>Max. points</b>	<b>5+2</b>

## 4 Feedback

*Note:* Answering the feedback questions is optional.

You can help us to improve this class (even for this semester). Therefore, please 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 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 or ambiguous?
- Do you have any suggestions for the instructors to support students more effectively?
- 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.