

Project Report

PS Non-Standard Database Systems
Summer Term 2021

Department of Computer Sciences
University of Salzburg

Group Number

Lastname1 Firstname1, StudentID1

Lastname2 Firstname2, StudentID2

Lastname3 Firstname3, StudentID3

March 8, 2021

This document includes the instructions for the checkpoints and also serves as a template for your report.

The \LaTeX source is composed of *two* checkpoint files (`milestone1.tex` and `milestone2.tex`), *two* feedback files (`feedback1.tex` and `feedback2.tex`), and a bibliography file (`bibliography.bib`). Include only `milestone1.tex` into the first report, and both checkpoint files into the second report.

Before submitting the report, please hide the instructions by setting the variable `showassignments` (that you find in the source code) to `false`.

The template contains the most important elements your report is supposed to cover. However, it is recommended that you mention anything that helps the reader (your instructor) to understand the contents of the report.

Milestone 1: Planning Phase

For this milestone, 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 load into your database, and (5) a roadmap for your project (*optional*).

1.1 Survey

Explore literature and other trustworthy (online) resources for information on different non-standard database systems. In our context, *non-standard* database systems include all databases except relational, disk-based systems like PostgreSQL. Non-standard database

systems also include processing frameworks that are closely related to databases (e.g., Big Data processing systems). After this literature survey, you should have a good overview of available non-standard database systems.

Enumerate (at least) *four* references and briefly summarize the main insight they provided (1 paragraph each). All references should also be included in the Section *References* that concludes the report.

1. ...
2. ...
3. ...
4. ...
5. Additional references ...

1.2 System Choice

Discussion of System Choice

After reviewing the literature and (some of) the respective database/processing systems, you will choose one (open-source/freely-available) system to base your project on.

Motivate your choice: Why do you find the chosen database system interesting? Please focus on the underlying database technology rather than, for example, tools for designing nice user interfaces.

We choose ... for our project.

Why is this type of system interesting to you? (1-2 paragraphs)

Key Features

Also name and discuss *four* key features/properties of your system. Please focus on the underlying database technology rather than, for example, tools for designing nice user interfaces.

In the following, discuss *four* interesting features/properties of the chosen system (1 paragraph each).

1. ...
2. ...
3. ...
4. ...
5. Additional features/properties ...

1.3 Application Description

Workload and Application

Now that you have chosen a database 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.

Describe the application and the workload your application has to deal with (e.g., OLTP,

OLAP, ...).

Description of workload and application scenario. (2-3 paragraphs)

Why a Good Fit?

Explain why this application/workload fits your database/processing system of choice.

Why is the application/workload a fit for your database/processing system of choice? (1-2 paragraphs)

Architectural Overview

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?

Architectural overview. (1-3 paragraphs)

1.4 Experimental Data

You will need data to test and evaluate your application. We distinguish between real world and synthetic data:

1. *Real world data*: Since you probably do not have your own data, download datasets from online sources like Twitter ^a, Kaggle ^b, Google Data Set Search ^c, ...
2. *Synthetic data*: Data generated with a program to fit your needs.

In both cases, you are required to describe its origin, its contents, important properties of the experimental data, and why this data set is good to test and evaluate your application. If you find multiple interesting data sets, you can receive a 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 be large enough to challenge the systems.

^a<https://developer.twitter.com/en/docs.html>

^b<https://www.kaggle.com/>

^c<https://toolbox.google.com/datasetsearch>

Describe your test data (2-3 paragraphs).

Optional. Other test data with description (optional, 2-3 paragraphs each).

1.5 Roadmap

Optional. 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. You may use the vroadmap environment (example below).

...

YYYY-mm-dd	•	Step 1
YYYY-mm-dd	•	Step 2
...	•	...

Grading

	Category	Max. points
1.1	Survey: references with summaries	4 (1 each)
1.2	Discussion of system choice	2
	Key features	4 (1 each)
	Workload and application	4
1.3	Why a good fit?	4
	Architectural overview	2
1.4	Experimental data	4
	<i>Additional datasets (optional)</i>	2
1.5	<i>Roadmap (optional)</i>	2
	Max. points	24 + 4

Feedback

Please answer the following questions (optional):

- *Effort*: How much time did each of the group members spend on this assignment?
- *Instructions*: Were the instructions clear? Any suggestions for improvement?
- Any other comments?

References