

# 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 30, 2021

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

The  $\LaTeX$  source is composed of *two* milestone files (`milestone1.tex`, `milestone2.tex`), *two* feedback files (`feedback1.tex`, `feedback2.tex`), and a bibliography file (`bibliography.bib`). Include only `milestone1.tex` into the first report, and both milestone 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 2: Implementation

This milestone consists of two parts: (1) the actual implementation of the project specified in the previous milestone and (2) the corresponding report. In the report, you are required to provide precise information on

1. helpful resources on techniques/systems/tools/... you used,
2. how to set up a machine such that your application can be deployed,
3. how your data set generation and/or import is implemented,
4. the implementation of the application itself,
5. design decisions or problems encountered on the way (*optional*), and
6. an alternative implementation based on another database/processing system (*optional*).

Citing small code snippets in the main text is fine. Please move long code snippets into the appendix of the report and reference them.

## 2.1 Resources

You will consult various (online) resources while you are implementing your project. Reference at least *four* resources and briefly discuss in which respect the respective resources were useful.

1. ...resource with description ...
2. ...resource with description ...
3. ...resource with description ...
4. ...resource with description ...
5. ...any further resource ...(even without description)

## 2.2 Setup

Describe your setup thoroughly. This subtask is supposed to provide precise information about the architecture/pipeline of your project.

Using this documentation, it should be possible to reproduce your setup and run your application.

Write down important facts of the (virtual) machine you are using, for example, the operating system, the prerequisites in order to reproduce your setup on a different machine (third-party tools/libraries, ...). If you adapted configuration files, write down your changes to these files (provide code snippets of the changes) and why the respective adaptations were necessary.

*Hint:* You may want to check out Ansible, a great tool for automating, documenting, and reproducing system configurations.

...

## 2.3 Datasets

### 2.3.1 New Datasets

If you found additional or more interesting data sets than the data sets described in the previous milestone, you may use them. However, describe the characteristics of the new data sets in your report (as required in the previous milestone).

Analogously, if you generate synthetic data sets, you may adjust the data set generation process. In this case, explain and justify your changes with respect to the previous report.

1-2 paragraphs per new dataset...

### 2.3.2 Generation

*Note:* This part only applies to groups that generate their own synthetic data sets.

Provide details about the data generation process (including code snippets). You should answer at least the following questions:

- Which programming languages/tools did you use?
- How does your data set generator enforce the desired data characteristics?

1-2 paragraphs per synthetic dataset...

### 2.3.3 Import

Describe the process of importing the data sets (provide code snippets). Did you apply any transformations before the data was imported into your database system? For example, in a traditional relational database system, you would provide the schema of the relations and the mapping between the data and the schema.

For static data sets (rather than stream-based data sets): How long did it take you to import the data sets? Did you put some effort into optimizing the import process? If so, what optimizations did you apply and which approach did (not) improve the import process?

...

## 2.4 Implementation

### 2.4.1 Detailed Description

Describe all important parts of the actual implementation of your application (provide code snippets; two pages in total – text *and* figures). This section is the main part of the report.

...

### 2.4.2 Key System Features

Discuss at least *two* key features of your *database system* that you use to implement your application (min. 1 paragraph each). Recall that this is a database class and the main objective is to get insights about the underlying database system. Therefore, this part is mostly about the internals of the database (rather than the user interface or other functionalities of your application). For example, for indexes, you would provide information about what is indexed, describe the type of the index, and discuss its purpose.

...

## 2.5 Design Decisions or Problems

*Optional.* If you had to make design decisions, summarize them here and justify your decisions briefly, e.g., by exposing possible trade-offs. If any, discuss problems you encountered during the implementation process. Briefly describe how you resolved the problems.

...

## 2.6 Alternative Implementation

*Optional.* Based on the application description in the previous milestone: consider an alternative implementation of your application that is based on a different database system. What would change? This includes (but is not limited to) limitations, advantages, disadvantages, and any other interesting aspect with respect to performance, scalability, flexibility, ...

Focus on the underlying database system and its characteristics.

This subtask is *conceptual* (rather than practical). Although you are free to implement your alternative approach, you are *not* required to do so.

...

## Grading

	<b>Category</b>	<b>Max. points</b>
2.1	Resources	4 (1 each)
2.2	Setup description/documentation	6
2.3	Datasets and import process	6
2.4	Implementation details	16
	Key database/processing system features	6 (3 each)
2.5	<i>Design decisions or problems (optional)</i>	2
2.6	<i>Alternative implementation (optional)</i>	4
	<b>Max. points</b>	<b>38 + 6</b>

## 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