PS Non-Standard Database Systems

Summer term 2020

Checkpoint 02
Implementation

Due date: Wed, June 17, 2020

General

Submit your checkpoint report until **Wed, June 17, 2020** using our submission system [1]. You may repeatedly upload new versions: only the **latest submission** is assessed.

Support

The preferred way of communication is the Slack channel `#nsdb-ps-2020ss` in the `dbteaching` workspace[2].

You may use the channel to get support if the instructions are unclear or if you run into problems in the course of your project. The channel is also open for topic-related discussions among students.

1 Task Description

This checkpoint consists of two parts: (1) the actual implementation of the project specified in the previous checkpoint 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,

---

[1] https://abgaben.cosy.sbg.ac.at
[2] https://dbteaching.slack.com

5. problems encountered on the way (*not graded*), and
6. an alternative implementation based on another database/processing system (*optional*).

1.1 Resources

Most probably, you will consult some (online) resources while you are implementing your project. Reference at least *four* resources and briefly discuss in which respect the respective resources were useful (0.25 points each; max. 1 point in total).

1.2 Setup

Describe your setup thoroughly (2 points). This subtask is supposed to provide precise information about the architecture/pipeline of your project.

Essentially, this section serves as a documentation of your setup. 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.

1.3 Data Sets

1.3.1 New Datasets

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

Analogously, if you generate synthetic data sets, you may adjust the data set generation process described in the previous checkpoint. In this case, describe and justify your changes in the Checkpoint 02 report.

1.3.2 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?

1.3.3 Import

Describe the process of importing the data set(s) (provide code snippets) (2 points). Did you apply any transformations before the data was imported into your database system? For example, in a traditional relational database system, you would define the schema of the relations here.

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?

1.4 Implementation

1.4.1 Detailed Description

Describe all important parts of the actual implementation of your application (provide code snippets) (3 points). This section is supposed to constitute the majority of the report.

1.4.2 Key System Features

Discuss at least *two* key features of your *database system* that you use to implement your application (1 point each; max. 2 points in total). 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?

1.5 Problems Encountered

Note: This part does not influence your grade.

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.

1.6 Alternative Implementation

Note: This part is optional. However, you will receive a bonus point if you work on this subtask.

Based on the application description in the previous checkpoint: 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, . . . In any case, focus on the underlying database systems and their characteristics.

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

2 Grading Scheme

	Category	Max. points
1.1	Resources	1.0 (0.25 each)
1.2	Setup description/documentation	2.0
1.3	Dataset import process	2.0
1.4	Implementation details	3.0
	Key database/processing system features	2.0 (1.00 each)
Bonus	Alternative implementation	1
	Max. points	**10 + 1**

3 Feedback

You can help us to improve this class (even for the current semester). Therefore, please answer the following questions.

- How much time did each of the group members spend on this assignment? Please use the Effort Collector tool [3] to anonymously answer this question; you find the access data in the slack channel of the course.
- Are there any hints/references we should provide for future students? 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?

Note: Feedback is optional. Your answers will have no impact on the grade.

---

[3]`https://aufwand.cosy.sbg.ac.at`