Universität Göttingen
Department of Computer Science

Julian Kunkel

Exercise Week 6
HPDA / Autumn 2025
240 Minutes Total
Discussion in Week 7: 2025-12-08

1. The tasks described in this worksheet are part of the formative assessment. They serve the purpose to prepare you for the examination. We will discuss the solutions during the next **interactive session** after they are handed out – while they fit to the lecture of the week they are handed out, they might be discussed in two weeks time due to the bi-weekly exercise schedule.

2. Make sure to plan your time for the whole sheet carefully. The complete exercise should represent approximately three hours of independent study. The time limit indicates how much time you should spend on each task, and not how much time you may actually need; it is important that you engage with the material and not that you complete all tasks perfectly. Feel free to collaborate and team up.

3. The exercises are designed to challenge you and train you further as guided self-study. The time limit might be too ambitious for you; you may team up with colleagues. It is not an issue as long as you manage to at least partially resolve each task within the time budget. If you (and your team) are struggling, reach out for help in Teams! You may also share your thoughts via the Studip Forum.

4. We recommend that you create a (private) Git repository (see https://gitlab.gwdg.de) where you store your findings and outcomes while processing the exercises. This portfolio of work can be useful in the future.

Contents

Task 1: Researching Schema Creation (120 min)

Read the documentation and synthesize a short summary of how the performance in MongoDB depends on the architecture and appropriate data model (and schema) creation.

Use the following documents:

- Choosing a shard key

- Sharding Broadcast

- Best practices

- Indexing Performance

Portfolio (directory: `6/mongodb`)

`6/mongodb/shards.txt` Description of the influence of the schema.

Task 2: MongoDB for Wikipedia (120 min)

In this task, you will create a data model to store individual Wikipedia articles in a MongoDB together with some derived data such as categories and links among the articles. Assume you want to store the whole

Wikipedia data inside a MongoDB. Make sure that the articles can be navigated and searched similarly to Wikipedia. Model the derived data in such a way that additional data can easily extend the existing data model. Test the creation and search of such documents using either the MongoDB shell or Python. You may want to test how to create an index on a field.

Portfolio (directory: `6/mongodb`)

`6/mongodb/data-model.txt`	A textfile including JSON that describes how Wikipedia data can be mapped to the document model provided by MongoDB
`6/mongodb/code.py`	Code to insert and query some example documents. Either Python or code in the MongoDB shell

Hints

- Look at a few articles at Wikipedia first. Select some features to model such as title, text, category, and links. Start with the most basic model, then expand it further.

- To install MongoDB in Ubuntu see the `hpda-samples` GitHub repository.

- See https://www.mongodb.com/basics/examples for examples.