University of Reading
Department of Computer Science

Julian Kunkel

Exercise Week 5
CS3DP / Autumn 2020
150 Minutes Total
Discussion in Week 7: 2020-11-09

---

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 team) are struggling, reach out for help in Teams! You may also share your thoughts on the channel.
4. We recommend that you create a (private) GIT repository where you store your findings and outcomes while processing the exercises. This portfolio of work could be useful in the future.

Contents

Task 1: Performance Analysis (75 min)

As part of this exercise, you will further research the meaning of scalability in High-Performance Computing (HPC) and discuss one example from HPC and one from Big Data or distributed systems.

Proceed as follows:

1. Research the meaning of strong and weak scaling by using the references:

   `https://www.archer.ac.uk/training/course-material/2016/12/mpi_scaling_manc/Slides/Scaling.pdf`

   `https://www.sharcnet.ca/help/index.php/Measuring_Parallel_Scaling_Performance`

2. Find a research paper or blog which evaluates strong/weak scaling for an HPC application (or benchmark).
3. Look at the performance measurements and compare the theoretic performance and the observation. What do you think about the quality of the evaluation?
4. Now, find a research paper or blog which discusses the performance of a distributed application (Big Data, IoT, Cloud, ...).
5. Look at the performance measurements and compare the theoretic performance and the observation. What do you think about the quality of the evaluation?

Portfolio (directory: `5/performance`)

`5/performance/scaling.txt`	Description of the strong/weak scaling and the meaning in terms of HPC (50+ words)
`5/performance/example-hpc.pdf`	Discussion of the performance in the example research paper/blog for the HPC application (max. 1 page and you may attach images)
`5/performance/example-distributed.txt`	Discussion of the performance in the example research paper/blog for the distributed application (max. 1 page and you may attach images)

Task 2: MongoDB (75 min)

In this task, you will create a data model to store individual Wikipedia articles in MongoDB together with some derived data such as categories and links among the articles. Assume you want to store the whole Wikipedia data inside MongoDB. Make sure that the articles can be navigated and searched in a similar fashion than provided by Wikipedia. Model the derived data in such a way that additional data can easily extend the existing data model. Then, test the creation and search of such documents using either the MongoDB shell or Python.

Portfolio (directory: `5/mongodb`)

`5/mongodb/data-model.txt`	A textfile including JSON that describes how Wikipedia data can be mapped to the document model provided by MongoDB
`5/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.
- You may use the MongoDB version provided in the image to test your mapping.