

Bachelor Thesis Contract for Jakob Joachim

Thesis Title: Methodology for Debugging Flink Applications

1. Problem and Objective:

In the modern time, enormous amounts of data are collected that make it incredibly hard for a single machine to process. Thus calls for a new way of building applications. Instead of using a single computer it is often better to spread the workload to multiple computers to reduce the running time. This method of programming is hard to execute without a framework to provide the necessary infrastructure for communication between the systems and to distribute the data on the systems. One of these frameworks is “Apache Flink”. This framework provides ways to debug the code as well. The challenge is that not all problems surface when the program is run on only a single machine. It is just a first step in making sure the program runs as intended. The second phase, which is much harder to achieve is finding and fixing errors that only occur when running in a distributed environment. This thesis tries to lay out a methodology for debugging a distributed program in the “Flink” framework as well as design and implement one part of the methodology.

2. Individual Tasks:

- overview of other approaches => Chapter 3
- design a methodology to debug Flink applications => Chapter 4
- design and implement one part of the methodology => Chapter 6
- test and evaluate the tool => Chapter 7

3. Organizational Matters

- **Timeframe:** now up to 31.8.
- **Advisor:** Prof. Dr. Olaf Zukunft
- **Secondary Advisor (to be discussed):** Prof. Dr. Klaus-Peter Kosakowski
- **E-mail:** jakob.joachim@haw-hamburg.de, mail@jakobjoachim.com
- **Telephone:** +49 176 661 774 03

4. The Temporary Structure of the Thesis:

1. Introduction
2. Fundamentals
3. Related Work
4. Concept of Methodology
5. Applications and Utilization of the Methodology
6. Concept of the Tool

7. Experiment/Evaluation of the Tool
8. Evaluation
9. Summary & Outlook

5. Milestones:

- until the end of April: completion of the contract, a draft of the “Related Work” chapter
- until the end of May: a first concept of the methodology
- until the end of June: finishing the methodology and the
- until the end of July: develop and test the tool.
- until the end of August: finish the last chapters and proofread everything

6. Literature:

- Ivan Beschastnikh, Patty Wang, Yuriy Brun, and Michael D, Ernst. 2016. Debugging Distributed Systems. Queue 14, 2, pages 50 (March 2016), 20 pages.
- Zeller, A. 2007. Beautiful Debugging. In Wilson, G. and Oram, A. Beautiful Code Chapter 28. O'Reilly Media, Inc., 26 Jun 2007
- Muhammad Ali Gulzar, Matteo Interlandi, Seunghyun Yoo, Sai Deep Tetali, Tyson Condie, Todd Millstein, and Miryung Kim. 2016. BigDebug: debugging primitives for interactive big data processing in spark. In Proceedings of the 38th International Conference on Software Engineering