Bachelor Thesis Topic
Parent Restarting for Workflow Systems

Motivation and Background

Batch processing systems are systems that process data using an acyclic topology. These topologies consist of operators which pass data around. A problem these topologies can have is that bugs can propagate before surfacing. One of the most common fault recovery mechanisms is restarting the task that faulted. If a task was restarted several times and keeps faulting the usual approach is to just say the workflow failed[1]. As we know the task that fails might not be the task that generated the bug, just the task that could not propagate the task. For this reason we want to add a system, which restarts the parent operation(s) of the failing task, if it continues failing.

Goals

The goal of this thesis is to extend a workflow system (e.g. Nextflow) with a restart operation that restarts the parent instead of continuously restarting only the child.

Description of the Task

The specific tasks are:
- Understand the chosen batch processing system
- Implement the parent restarting operation
- Define and if possible implement improvements for the operation
- Evaluate the efficiency and effectivity of the approach

Research Type

Theoretical Aspects:	★★☆☆☆
Industrial Relevance:	★★★★☆
Implementation	★★★★★

Prerequisite

The student should be enrolled in the bachelor of computer science program, and has completed the required course modules to start a bachelor thesis.

Skills required

Programming skills in Java/relevant programming language, understanding of, or willingness to learn, the software engineering methods and tools (e.g., Nextflow) needed for the project.

Contacts

Jan Arne Sparka (sparkaar@informatik.hu-berlin.de)

Software Engineering Group, Institut für Informatik, Humboldt-Universität zu Berlin

References

[1] https://www.nextflow.io/docs/latest/process.html