

Process Discovery using Big Data stack - Implementing the Alpha Algorithm with Map-Reduce Project Initiation

Martin Hashem, Xiangnan Chen

April 18, 2019
RWTH Aachen

1 Overview

2 Buisness Case

We develop a business case to disclose the potential of this project. The study is divided into three parts. First we explain one of the current algorithm in process mining. Furthermore we will define the scope of the project and point out the key benefits of our optimization.

2.1 Initial Situation

The current situation in process mining eventlogs using the Alpha algorithm is very centralized. All eventlogs are pushed together into one computation unit, eating up a big amount of resources, both in used space and computation time. The optimization of this is the main task.

2.2 Scope

The main goal of this project is to create a web application that provides callculations with the Alpha algorithm and can prepare the data using the Map-Reduce. The project contains these tasks:

- **Code required:** Integrating the system Hadoop into the pm4py interface and read files from it
- **Code required:** Run the Alpha algorithm utilizing map-reduce
- **Code required:** Reading in logs and upload them into Hadoop

2.3 Key Benefits

The current scope of mining in Big Data is slow and centralized. Using map-reduce in combination with the Alpha algorithm, callculations can be deffered to more, smaller instances and also reduce overhead in transferring the events. Depending on the amount of map-reduce stages that are implemented, there can be significantly less time used during the mining process, due to the parallized nature of the approach. [?]

3 Feasibility Study

3.1 Theoretical Point of View

A paper by Assadipour[?] suggests that using a decentralized approach for process mining using the Alpha algorithm by splitting the calculations into smaller tasks available to smaller instances.

The Alpha algorithm takes eventlogs and generates a Petri net model including an initial and final mark. The main purpose of the Alpha algorithm is to portray the relations between activities, but comes with the flaw of not finding self loops.

Map-Reduce is a technique to allow the user to have a scalable calculation on a distributed system. By first mapping events to an identifier and then reducing the map to its relevant values, the amount of needed data shrinks. Here the reductions prepare the data for the Alpha algorithm, removing the overhead of finding the traces first.

3.2 Technical Point of View

3.3 Risks and Mitigations

Project Managment Risks

Technical Risks

4 Project Plan

5 Project Team

References

- [BDP93] Mic Bowman, Saumya K. Debray, and Larry L. Peterson. Reasoning about naming systems. *ACM Trans. Program. Lang. Syst.*, 15(5):795–825, November 1993.
- [Bra91] Johannes Braams. Babel, a multilingual style-option system for use with latex’s standard document styles. *TUGboat*, 12(2):291–301, June 1991.
- [Cla91] Malcolm Clark. Post congress tristesse. In *TeX90 Conference Proceedings*, pages 84–89. TeX Users Group, March 1991.
- [Her93] Maurice Herlihy. A methodology for implementing highly concurrent data objects. *ACM Trans. Program. Lang. Syst.*, 15(5):745–770, November 1993.