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

User Manual

Xiangan Chen, Martin Hashem

July 4, 2019
RWTH Aachen

1 Introduction

Process mining is an approach to extract process models from event logs. Since the distributed nature of modern information systems, event logs are likely to be distributed among different physical machines. Map-Reduce is a scalable approach for efficient computations on distributed data. In this Python application we will present the main idea of a Map-Reduce implementation of the Alpha process mining algorithm, to take advantage of the scalability of the Map-Reduce approach.

This project is a python-based webapp, that integrates the big data capabilities of the Hadoop system into the process mining framework pm4py.

1.1 WebApp Requirements

Our WebApp was designed mainly on Unix-based operating system (MacOS and Debian of Linux), but can be accessed from every major browser.

We recommend to use the latest version of your browser. Both for security reasons and to guarantee the best experience.

1.2 Access WebApp

To use our WebApp, perform the following:

1. Navigate to the link
2. Enter your username and password
3. Click on the button Sign in

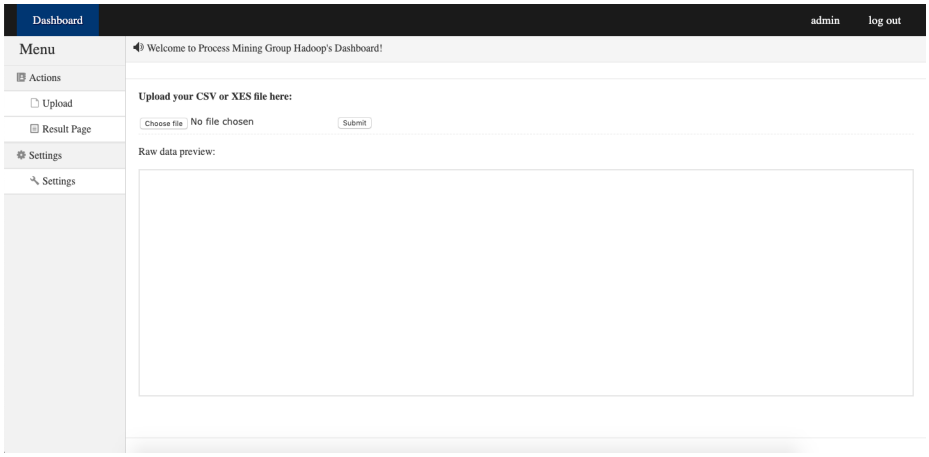

Fig. 1.

1.3 WebApp Overview

The WebApp consists of the following areas, which are showed in Fig. 1.

The main areas are:

1. Shortcut Bar contains shortcuts to the dashboard site itself on the left side. The user settings and logout button can be found on the right side of the shortcut bar.

2. Main Toolbar contains buttons with the most important functions for each applications like uploading the required files and the result page.

3. Main Window displays the main content of the application. On the upload page there are upload bar and raw data preview box. After selected the file of right format, the content of the uploading file will be showed in the box.

2 Installation

The version control of this project depends on Git, and our team chose to use GitHub to manage our source code and releases.

To deploy our WebApp on a server, you can just go to our GitHub releases page (*https://github.com/xianganc/PraktikumProcessMining/releases*), then download the *setup.sh* to local, run the contents inside the .sh file like in Fig. 2.

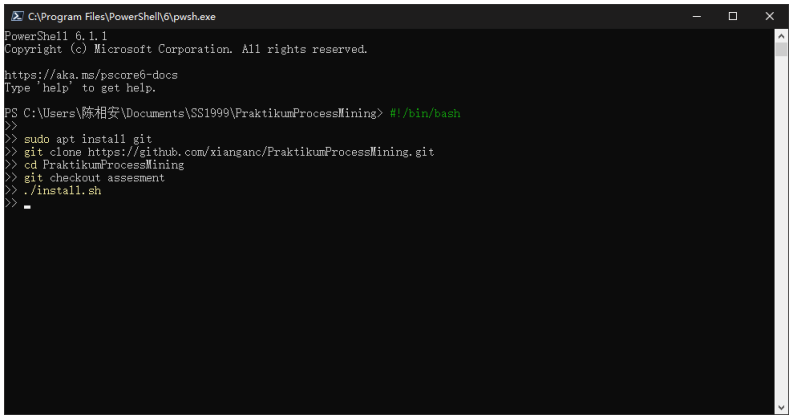

Fig. 2.

3 Upload

In this section we explain how to use the WebApp. After reading this section, you will be able to preview the to be uploaded data and upload a CSV or XES file to the server.

3.1 Uploading a CSV or XES File

1. Click the *Choose file* button on the top part of the main window.

2. To upload a CSV or XES file, choose the file from local disk.

3. If the selected file's format matches the server requirements, you will see the data inside the file in the text box below.

4. Press *Submit* button to upload the file to the server for further calculation.

3.2 Replace a wrong selected file

If you unexpectedly select a wrong file, you can do the followings:

2.2.1 Wrong File Format

If the selected file were in the wrong format (e.g. a pdf file), then a pop-up window will show up, to notify the user for choosing a wrong file like Fig. 3.

2.2.2 Wrong File Content

If the selected file were with the wrong content, maybe the wrong selected has a similar file name as the should be seleted one. When the user selected a wrong file, you can re-select the file by clicking the *Choose file* button or just refresh the website to re-select the file.

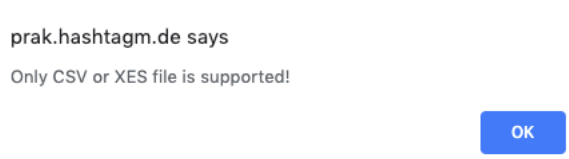

Fig. 3.

4 Result

After successfully submited the file, it should be uploaded to our server and being sent directly to calculation.

4.1 Check out the Output

Then you can simply click the *Result Page* on the left side in the main toolbar to check out the output.

5 Log out

If anytime the logged user want to log out, he can just click the *logout* button on the top right corner to log out.