University of Zurich[UZH]

Department of Informatics

University of Zürich
Department of Informatics
Binzmühlestr. 14
CH-8050 Zürich
Phone. +41 44 635 43 11
Fax +41 44 635 68 09
www.ifi.uzh.ch/dbtg

UZH, Dept. of Informatics, Binzmühlestr. 14, CH-8050 Zürich

Prof. Dr. Michael Böhlen
Professor
Phone +41 44 635 43 33
Fax +41 44 635 68 09
boehlen@ifi.uzh.ch

Zürich, August 14, 2023

MSc Project: Adapting the Raspberry Pi Cluster for "Systems for Data Science" to 64 bits

Cluster management and distributed computing is part of the syllabus of many data analytics and data sciences modules. However, it is quite hard on the practical side to go into details, as on a typical cloud-based system many parameters are outside of the control of the user. In the best case, a user is able to configure some basic parameters via a dashboard and the cluster remains an abstract entity. Also, important properties, such as resilience, are hard to investigate, as users are usually not allowed to crash cluster nodes.

The goal of this project is to take the Raspberry Pi (RPi) clusters developed for the "Systems for Data Science" course and adapt them to a 64-bit environment (so far, the clusters were running 32-bit software). The move from 32 to 64 bits opens up new ways of deploying new systems and software directly on the cluster (some important frameworks are not available in a 32-bit version).

The work is structured into the following tasks:

- The first task is to set up the 64-bit infrastructure on a Raspberry Pi cluster. This is mainly about installing the 64-bit version of Raspbian on all the RPis of a cluster, create user accounts, and configure the network with fixed IP-addresses.
- The second task is to go through the existing classwork exercises for the "Systems for Data Science" course and identify all the systems, frameworks, and software that has to be deployed to be able to conduct the exercise on the new cluster. For instance, these include YARN, HDFS, Hadoop Map-Reduce, HBase, MongoDB, Spark, and so on. In a second step, these systems need to be installed on the cluster and configured appropriately. All this work needs to be documented, since ideally the students working with these clusters would recreate the deployment during the practical exercises (to get

a better understanding of how the different systems work).

- The third task is about adapting the existing coursework exercises to the new framework. The new version has to reflect the different setup, i.e., a lot of the instructions on how to do the set-up for an exercise will be different.
- There is an optional task to explore further experiments that could be done with the new version of the cluster.
- Finally, all the work needs to be written up in a report.

Supervisor: Sven Helmer

Department of Informatics, University of Zurich

Prof. Dr. Michael Böhlen