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, February 21, 2020

**MSc Project: A Raspberry Pi Cluster for Teaching Big-Data Analytics**

Cluster management 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 design and develop a cluster made up of Raspberry Pi computers. This opens up the possibility to put students in full control of a small cluster in which all the different parameters can be tested and tuned (including crashing a cluster node).

The work is structured into the following tasks:

- The first task is to set up the basic infrastructure and compare different solutions to each other. This is broken down into the following subtasks:
  - The first subtask is about connecting the hardware, installing an operating system, and integrating the Raspberry Pi devices into a network.
  - The next step is to install cluster management software, such as YARN, on top of this infrastructure. The goal is to try out different platforms and figure out which platform is best suited for a Raspberry Pi cluster.
  - The next subtask, evaluating different platforms according to their suitability and performance, follows directly from this. The current assumption is that different platforms have different overheads, making them more or less suitable to a Raspberry Pi environment.
- After all the infrastructure has been put in place, the second big task is to design practi-

cal experiments that can be used in a classroom setting and to write the documentation necessary to run the experiments. The main system that will run on top of the cluster will be Apache Spark.

**Supervisor:** Sven Helmer

Department of Informatics, University of Zurich

Prof. Dr. Michael Böhlen