MLMP - Management Platform for Deep Learning Infrastructure

Degree programme : BSc in Computer Science | Specialisation : Computer Perception and Virtual Reality
Thesis advisor : Peter von Niederhäusern
Expert : Prof. Dr. Torsten Braun (University of Bern)

Deep learning requires a lot of computing power. Providing the appropriate BFH resources at the right time to successfully practice deep learning is the goal of this project. Over the course of the thesis, a web application has been developed to not only manage the availability, but to make reservations and access to these resources possible as well.

Introduction

A world without artificial intelligence can no longer be imagined. Whether we are talking about autonomous vehicles, supporting experts in the medical field or even aiding in solving cybercrime. One of the most recent advancements is deep learning, in which, contrary to its superset machine learning, manual feature engineering takes a step back. Deep learning creates its own features due to the brain like working principle. Next to requiring a considerable amount of data to generate reliable results, a lot of computing power is required to train and practice deep learning in a timely manner. The cpvrLab at BFH owns such hardware but it is distributed among several machines running with different operating systems and a variety of specialized processors called GPUs.

Results

Managing those resources is not easy and often done manually by system administrators. This incurs a high workload for the support staff. Solving this challenge is the goal of this thesis. A web application has been developed which allows users to reserve and access an environment. [1] This environment is running on a user-selected node which meets the resource needs of the respective project. Such nodes are joined in a Kubernetes cluster whose master node orchestrates everything.
Eventually, connecting to the environment can either be done via JupyterLab or a secure shell connection. To enable persistence of training as well as the user's data, a personal storage location on a dedicated network storage system is allocated which is available in each environment. [2]

Conclusion

Starting with a product requirements document to formulate requirements and goals the project launched smoothly guided by the Scrum methodology. On the road to completion, many obstacles had to be overcome. Using previously unfamiliar technology led to gaining more knowledge in the corresponding domains. The machine learning management platform came into existence and is now ready to be used by students, professors and data scientists alike.
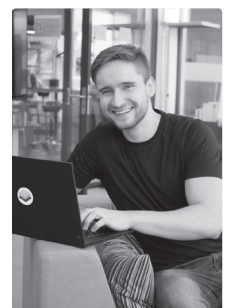
Lukas Zbinden
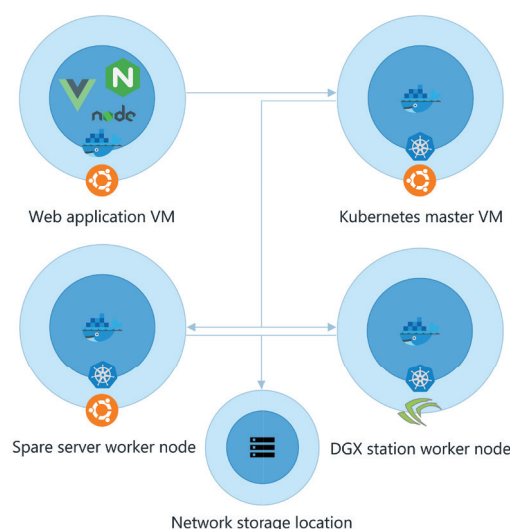luzbinden@gmail.com

Figure 1: Used technologies as a wordcloud

Figure 2: Web application system landscape