

Provisioning*

Distributed Data Processing Environments

Lab Guide 1

This session aims at creating a provisioning recipe with Vagrant.

Steps

1. Download the Bundesliga dataset from HuggingFace.¹
2. Create a small Python program that loads the data file and counts how many goals were scored by some team. Provide the team name as a parameter.
3. Create a Vagrant recipe that installs software requirements to run the program, using `apt` and `pip`.
4. Change your recipe to automatically download the dataset directly into the VM.
5. Change your recipe to run the program on startup.
6. Remember to destroy the VM and cleanup with: `vagrant destroy`

Learning Outcomes Create simple provisioning recipes that install software, download data, and run a program. Apply command line utilities for various tasks.

*May use of AI tools for steps 2 and 4.

¹https://huggingface.co/datasets/KrisSommer/Bundesliga_Stats_2018-2024