Python:

* download all mtg cards
  + api: [https://api.magicthegathering.io/v1/cards?pageSize=[PAGESIZE]&page=[PAGENR](https://api.magicthegathering.io/v1/cards?pageSize=%5bPAGESIZE%5d&page=%5bPAGENR)]
  + declare PAGESIZE = 150, so that one page contains 150 cards
  + declare PAGENR (Min) = 1
  + declare PAGENR (Max) = 709, so that every card will be extracted. This value was determined by manual testing of the api
  + Function: get\_all\_mtg\_cards will iterate through every page; from page = 0 to page = 709 with pagesize 150, extract the data (mtg\_cards – english, mtg\_foreign\_cards – other languages) and put it into a json file in the airflow container. The json file will later be exported into the hdfs
* create ssh tunnel
  + ! Somehow only allows database queries via ssh tunnel
  + ! for more information have a look at the section “MySQL Queries”
* execute mysql queries
* export data from hive to mysql table

Hadoop

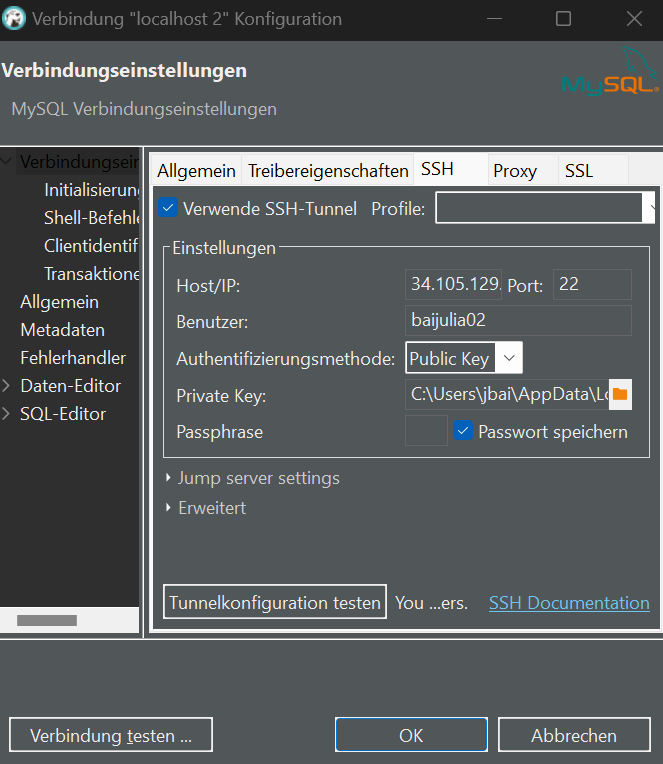
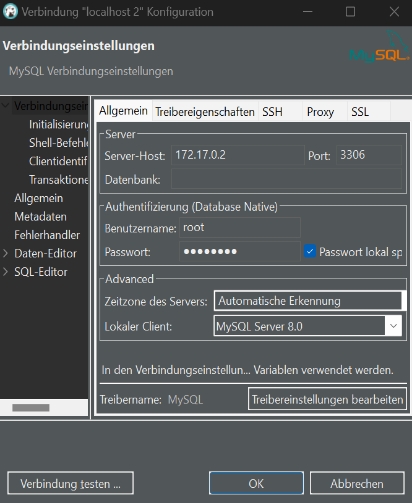
* Setup
  + Pull Hadoop Image:
    - docker pull [name]
  + Run/Start Hadoop container
    - docker run […]
    - docker start [name]
  + Initialize Hadoop Container
    - docker logs [name]
  + Get into the Hadoop container
    - docker exec -it [name] bash
* Start Hiveserver
  + Switch to Hadoop user
    - Sudo su hadoop
    - cd
  + start Hadoop cluster
    - start-all.sh
  + connect to the hiveserver2 (database)
    - hiveserver2

Airflow

* Setup
  + Pull Airflow image
  + Run/Start Airflow Container
  + Initialize Airflow Container
  + Get into the Airflow container
    - docker exec -it [name] bash
* Get into Magic.py for DAG
  + sudo su airflow
  + cd
  + Create a dag:
    - vi /home/airflow/airflow/dags/Magic.py
* Download dependencies:
  + For building a connection to the ssh server
    - Pip install mysql-connector
    - Pip install pandas
    - Pip install paramiko
    - Pip install sshtunnel

MySQL Setup

* Purpose: Load reduced table data to enduser database in mysql
* Get docker container of mySql
  + Pull mySql image: sudo docker pull mysql/mysql-server:latest
  + Deploy Mysql container
  + Connect to mysql Container
    - DBeaver: SSH Konfiguration, Server-Host IP mittels docker inspect [name] ermitteln



* + Configure mysql container
    - bind-adress = 0.0.0.0
  + Database:
    - Name: mtg\_db
    - User: root
    - Port: 3306
    - Root password: password

MySQL Queries

* Build a connection to the ssh server
* Mysql queries are executed from python via ssh tunnel
  + Requirements:
    - pip install sshtunnel
    - pip install pymysql
    - ssh address must be updated in code in variable “ssh\_host”

HiveSQL Queries

* Build a connection to the ssh server
* Requirements:
  + Pip install pyhive
  + sudo apt-get install libsasl2-dev
  + pip install sasl
  + pip install thrift-sasl
  + pip install thrift

Hive to MySQL:

* Select all data from cards\_reduced table in the HiveSQL database via HiveSQL
* Insert the data into the mtg\_cards table in the MySQL database via mySQL
* Requirements:
  + Save ssh private key in: “/home/airflow/airflow/dags/big-data”

Website:

* frontend: NodeJS
  + npx create-react-app frontend
  + App.Js and App.Css changes
  + Deletion of unnecessary files
* Backend:
  + Install node and required libs
    - npm init
    - npm i express
    - npm i mysql2
    - npm i cors
  + folder:
    - routes
      * glue between URI and corresponding function in “services/\*”
    - services
      * houses all services
      * db.js talks with database
      * mtg\_db.js contains methods to get Data from the mysql mtg enduser db
* frontend-backend-communication:
  + Axios:
    - http requests: GET with axios.get()