Dokumentation des Pentaho Workflows

Contents

[Vorwort 1](#_Toc88747144)

[Voraussetzungen 1](#_Toc88747145)

[Starten der Docker Container und des Workflows 1](#_Toc88747146)

[Stoppen der Container 2](#_Toc88747147)

[Erklärung des Workflows 2](#_Toc88747148)

[Überblick 2](#_Toc88747149)

[Erklärung der einzelnen Schritte 3](#_Toc88747150)

[Region\_List\_Generation.ktr 3](#_Toc88747151)

[Download\_Addresses.kjb 3](#_Toc88747152)

[3](#_Toc88747153)

## Vorwort

Vorab gilt zu sagen, dass ich das Programm fehlerfrei auf meinem lokalen Docker getestet und vollständig hab laufen lassen. Bei Fragen bezüglich Fehlern, stehe ich Ihnen jederzeit zur Verfügung.

## Voraussetzungen

Alle Voraussetzungen, die erfüllt sein sollten, sind in der Readme Datei des Projektes genannt.

## Starten der Docker Container und des Workflows

1. Zuerst sollte docker-compose auf dem System installiert werden, hierfür siehe <https://docs.docker.com/compose/install/>
2. Um die Container zu starten, reicht *„docker-compose up -d“* im Projektverzeichnis, in welchem auch die „docker-compose.yml“ zu finden ist
3. Sobald im Container „pentaho“ und „hadoop“ die Zeilen „Container Startup finished“ zu sehen ist, kann sich in diese eingeloggt werden  
   Hadoop Container:
   1. „sudo su hadoop“
   2. „cd“
   3. „start-all.sh“
   4. „hiveserver2“

Pentaho Container:

1. „sudo su pentaho“
2. „*/home/pentaho/pentaho/data-integration/kitchen.sh -file=/home/pentaho/custom\_pdi\_jobs/Address\_Validation.kjb”*
3. Der Workflow sollte nun starten

## Stoppen der Container

Zum Stoppen reicht ein *„docker-compose down”* im Projektverzeichnis.

## Erklärung des Workflows

### Graphical user interface Description automatically generatedÜberblick

## Erklärung der einzelnen Schritte

## Region\_List\_Generation.ktr

Graphical user interface, text, application

Description automatically generated

## Timeline Description automatically generatedDownload\_Addresses.kjb

## Diagram Description automatically generatedRead\_Subfolder\_Names.ktr

## Diagram Description automatically generatedImport\_Hadoop.kjb

## Timeline Description automatically generatedCreate\_Hive\_Tables.kjb

## Add\_Hive\_Partitions.kjb

Timeline

Description automatically generated

## Create\_Final\_Hive\_Table.kjb

Timeline

Description automatically generated with low confidence

## Graphical user interface, application Description automatically generatedMove\_Files\_From\_Hive\_To\_RBMDS.ktr

## Cleanup.kjb

Timeline

Description automatically generated

# Testdaten

Daten für die Enduserdatenbank:

Host: localhost

User: postgres

Passwort: example

Database: default

Port: 5432

Funktionelle Testdaten:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Street | Number | City | Region | Postcode |
| 10TH AVE | 110 | ROSS | PA | 15229 |
| 2ND ST | 578 | PITCAIRN | PA | 15140 |
| WINDSOR WAY | 17 | EAST HANOVER | NJ | 07936 |
| KNAPP STREET | 25 | EASTON | CT | 06612 |