My Journey to Apache Wayang – Part 1

## Prerequisites

### BDE Blossom Development Environment

The Blossom Development Environment is a Docker based Jupyther Notebook which comes with a preinstalled version of Apache Wayang (v. and other usefull libraries for platform agnositic data processing.

BDE comes with a bundled Spark and Flink setup. This is all you need for a first test.

### Locally built version of Apache Wayang

We have to provide JDK11.

homebrew

sdkman

maven

mvn wrapper

HADOOP and SPARK must be installed on the local machine

### Data Processing Platform

I need 2 or more „data processing platforms” up and running.

Typicall, such a data processing platform is a Spark Cluster in, in single node mode or even a multi-node cluster, somewhere in the cloud. Alternatively, an Apache Flink cluster can be used.

|  |  |  |  |
| --- | --- | --- | --- |
| Platform Type |  |  | Repository |
| Spark Operator |  |  | https://github.com/GoogleCloudPlatform/spark-on-k8s-operator/blob/master/docs/user-guide.md |
| Flink Operator |  |  | https://nightlies.apache.org/flink/flink-kubernetes-operator-docs-main/ |

**Table 1** – Operators for managing data processing platforms on Kubernetes

An easy way of getting such data processing platforms at your fingertips is either by using the operators listed in table 1. You can also use one of the full service cloud providers as listed in table 2.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | Li |
|  |  |  |  |
|  |  |  |  |

**Table 2** – Full-service cloud providers for data processing platforms which can be used with Apache Wayang.