



Private Public Partnership Project (PPP)

Large-scale Integrated Project (IP)



Social Data Aggregator Admin Guide

Project full title: Future Internet Core

Project acronym: FI-Core

Contract No.: 632893

Contents

Introduction	2
OVERVIEW	2
Source	2
FIWARE Social Data Aggregator - Installation and Administration Guide	3
SDA Installation	3
Requirements	3
Step 1: Install Java 8 sdk	3
Step 2: Install Maven	3
Step 3: Install and Configure Apache Spark	3
Step 4: Install MySQL	4
Step 5 (Optional): Install Apache Kafka	4
Step 6: Download and Install Social Data Aggregator	4

Introduction

Welcome the Admin Guide for the Social Data Aggregator Generic Enabler. The online documents are being continuously updated and improved, and so will be the most appropriate place to get the most up to date information on using this interface.

OVERVIEW

Over time all generations have come to embrace the changes social network has brought about. Nowadays online social media have gained astounding worldwide growth and popularity. They are generating a huge amount of records about users' activities and also attracting attention from variety of researchers and companies worldwide. Every day data collected by social networks are gathered and feed a variety of analytics (users behavior, habits, topic trends..) which are capable of extracting significant patterns and further analysis. The aim of Social Data Aggregator (SDA) GE is to retrieve data from different Social Networks and provide different analytics depending on user needs. The GE relies on Apache Spark for computation on data. The SDA comes with different built-in modules but custom modules can be added as well.

Source

The source code of this project can be cloned from the [GitHub Repository](#). Code for related libs can be found on [GitHub FiwareTIConsoft Group](#).

FIWARE Social Data Aggregator - Installation and Administration Guide

SDA Installation

This guide tries to define the procedure to install the Social Data Aggregator in a machine, including its requirements and possible troubleshooting that we could find during the installation.

Requirements

In order to execute the Social Data Aggregator, it is needed to have previously installed the following software or framework in the machine:

- Java 8 [<http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html/>]
- Maven 3.0.4 or above [<https://maven.apache.org/download.cgi/>]
- Apache Spark 1.4.0 [<http://spark.apache.org/downloads.html/>]
- Apache Kafka 0.8.1 [<http://kafka.apache.org/downloads.html/>]
- MySQL 5.6.14 or above [<http://dev.mysql.com/downloads/mysql/>]

Step 1: Install Java 8 sdk

If you do not have java 8 installed , please, follow instructions for your Operating System to download the correct package or to install it by a packet manager.

Step 2: Install Maven

If you do not have java 8 installed , please, follow instructions for your Operating System or download it on <https://maven.apache.org/download.cgi/>

Step 3: Install and Configure Apache Spark

Download Apache Spark

Download Apache Spark from <http://spark.apache.org/downloads.html/>. Choose:

- 1.4.0 as Spark Release
- If you have Hadoop already installed on your system choose the prebuild with your version of hadoop otherwise Pre-built for Hadoop 2.4 or later should be fine
- Direct download as download type

Configuration

Standalone

To launch a Spark standalone cluster with the launch scripts, you should create a file called *conf/slaves* in your Spark directory, which must contain the hostnames of all the machines where you intend to start Spark workers, one per line. If *conf/slaves* does not exist, the launch scripts defaults to a single machine (localhost), which is useful for testing. Note, the master machine accesses each of the worker machines via ssh. By default, ssh is run in parallel and requires password-less (using a private key) access to be setup. If you do not have a password-less setup, you can set the environment variable `SPARK_SSH_FOREGROUND` and serially provide a password for each worker.

For more information please refer to the following guide: <http://spark.apache.org/docs/1.4.0/spark-standalone.html/>

Yarn

If you have already an instance of yarn installed and configured please refer to the following guide for configuration: <http://spark.apache.org/docs/1.4.0/running-on-yarn.html/>

Remember to set the variable MASTER in SDA configuration files to yarn-cli or yarn-cluster.

Step 4: Install MySQL

To install MySQL Server, it is better to refer official installation page and follow instructions for the Operating System you use: <http://dev.mysql.com/downloads/mysql/>

You will need four packages:

- mysql-server
- mysql-client
- mysql-shared
- mysql-devel

After installation, you should create a user.

To add a user to the server, please follow official documentation: <http://dev.mysql.com/doc/refman/5.5/en/adding-users.html/>

Under **social-data-aggregator/db-data_model/** folder there are some .sql files with the sql schema that contains the tables needed from the connectors and consumers. Depending on your needs you can choose to import all or some of them. The file containing the database schema is **tw_stats_db.sql** To import a sql schema on your database by command line:

```
mysql -u root -p [DB_NAME] < tw_stats_db.sql
```

Then grant all privileges to the previously created user to this database.

Step 5 (Optional): Install Apache Kafka

To provide near real time data to consumers the Social Data Aggregator uses an internal bus in a publish/-subscribe pattern. The default connector expects apache kafka as internal bus. You can change the default behaviour by providing your own connector and modifying the configuration file *bus_impl.conf* on *configs/*. Anyway if you want to use the default connector you need to install and configure apache kafka.

Installation

Please refer to <http://kafka.apache.org/documentation.html#quickstart/>

Configuration

Please refer to <http://kafka.apache.org/documentation.html#configuration/>

Step 6: Download and Install Social Data Aggregator

Download the component by executing the following instruction:

```
git clone https://github.com/FiwareTIConsoft/social-data-aggregator.git
```

To deploy Social Data Aggregator from source go to the project main folder and launch the following command:

```
mvn clean package
```

Once built SocialDataAggregator with Maven, under the folder *scripts/your_os_env/* there is a script called *make-dist*. Run it with the following syntax (e.g. linux):

```
./make-dist.sh <output_folder>
```

where *_output_folder* is the folder inside which you want to create the SDA GE folder tree. When the script finish to run, you should see the following dir tree:

```
1 sda
2 - bin (contains all the binaries of sda in their respective folder)
3 - confs (contains the configurations of each specific sub-component)
4 - scripts (contains the launch scripts for each sub-component and a start-all script to start a
    ll components)
```

To configure and run SDA please refer to the User Guide.