



Alflytics

Open Source Business
Intelligence solution over Alfresco

Version 5.0 Early Access

<https://github.com/fcorti/Alflytics>

@ 2013-2017 Francesco Corti

Latest revision: 24 April 2017



Index

1. Introducing Alflytics	3
1.1 About the project	3
1.1.1 Open Source	3
1.1.2 Free of charge	3
1.1.3 License	3
1.1.4 Goals	3
1.2 The history	4
1.3 Overview of the architecture	4
1.4 The components	4
1.4.1 Alfresco platform	4
1.4.2 Pentaho Analytics platform	4
1.4.3 The Data Warehouse	4
1.4.4 Pentaho Data Integration (optional)	4
1.4.5 Pentaho Reporting (optional)	4
1.4.6 Pentaho CTools to develop dashboards (optional)	4
1.5 Compatibility matrix	4
1.6 References and resources	5
2. How to use Alflytics	5
2.1 Extracting from Alfresco	5
2.2 Publishing reports into Alfresco (optional)	7
2.3 Using analytics	8
3. How to install Alflytics	10
3.1 Preparing the installation	10
3.1.1 Minimum requirements	11
3.1.2 Installing Pentaho Analytics platform	11
3.1.2 Installing Saiku analytics (optional)	12
3.2 Getting Alflytics	16
3.2.1 Getting Alflytics using the Pentaho marketplace	17
3.2.2 Downloading Alflytics	17
3.2.3 Checking Alflytics	17
3.3 Installing Alflytics	18
3.3.1 Accessing the Alflytics wizard	18
3.3.1 Installing Alflytics using the wizard	19
3.3.2 Configuring the access to Alfresco	22
3.3.3 Troubleshooting	23
3.4 Manual installation of Alflytics (advanced)	23



3.4.1 Installing Alflytics on PostgreSQL	23
3.4.2 Installing Alflytics on MySQL	23
3.4.3 Installing Alflytics on HyperSql	23
3.5 About scheduling extractions and publications	26
4. Customizing Alflytics	26
4.1 Customizing reports	26
4.2 Customizing dashboards	26
4.3 Customizing the warehouse and the ETL	26
5. Scalability	27
5.1 The Alflytics Data Warehouse on PostgreSQL or MySQL	27
5.2 Tuning the Pentaho Analytics platform	27
5.3 Using the Pentaho Data Integration for ETL	27
5.3.1 Installing Pentaho Data Integration	27
5.3.2 Creating the connection to the Alflytics repository	27
5.3.3 Creating and scheduling the scripts to run the ETLs	28
6. How to contribute to the project	29
7. Support and contacts	29
7.1 Disclaimer	29
7.2 Preparing the request of support	29
7.3 How to get support from the community	29
7.4 How to get professional support	29
Appendix - FAQs	29



1. Introducing Alflytics

Alflytics is the Open Source Business Intelligence (OSBI) solution over Alfresco content and processes, built using the Pentaho platform. With Alflytics is provided a pure Business Intelligence solution to extract, store and enquiry content data (documents, folders, metadata, tags, categories, etc.), audit trail and workflows at a very detailed level, with the goal to be easily customized and extended with other entities coming from external Services like: ERP, CRM, custom Applications, etc.

Alflytics uses the data coming from Alfresco platform, the ability of the Pentaho platform to build reports, dashboards and free analysis, and the latest techniques in Data Warehousing, to define a complete set of analytics and insights on top of your preferred ECM+BPM solution.

1.1 About the project

Alflytics is invented, designed, developed and maintained by Francesco Corti, Developer and member of the Alfresco and Pentaho communities. Below some high level characteristics, defining the solution, to better understand the project and its nature.

1.1.1 Open Source

Alflytics is purely an **Open Source** project. This means that every source code, script or configuration is available and given as included in the release/distribution. Also the frameworks and platforms used by Alflytics are Open Source, even if the precise details depend on the Vendor and the Owner of the solution itself.

1.1.2 Free of charge

Alflytics is **free of charge** and not related to number of user, dimension of the repositories or any other detail of the final solution.

1.1.3 License

Alflytics is released under the **GNU Lesser General Public License (LGPL) Version 2.1**. This license has been selected to make easier the adoption of the solution from final users and Companies interested to use Alflytics in a vertical solution, for specific needs or specific use cases.

1.1.4 Goals

The goal of Alflytics is to become more adopted and make it **the most reliable Business Intelligence solution** over the entire Alfresco platform. To reach that goal we kindly ask you to contribute to the project (as described into one of the following chapters).



1.2 The history

TODO

In 2012 Analytics over Alfresco very far to be real.

AAAR is the origin

1.3 Overview of the architecture

One of the key concept of the solution is that is built as an integration of existing platforms, instead of “reinventing the wheel” and everything from scratch, one more time.

Picture of the architecture

1.4 The components

TODO

1.4.1 Alfresco platform

TODO

1.4.2 Pentaho Analytics platform

Alflytics uses also Saiku for free analysis.

1.4.3 The Data Warehouse

TODO

List of supported databases (Hsql, postgresql MySql)

1.4.4 Pentaho Data Integration (optional)

TODO

1.4.5 Pentaho Reporting (optional)

TODO

1.4.6 Pentaho CTools to develop dashboards (optional)

TODO

1.5 Compatibility matrix

TODO



1.6 References and resources

TODO

2. How to use Alflytics

Alflytics is designed to be easy to use, easy to be customized/extended and ready to scale on huge repositories. In this section we are going to see how to use Alflytics and what does it means that its use it is easy.

Using Alflytics means doing the following tasks.

- **Extracting** data from the sources (by default Alfresco) into the Alflytics Data Warehouse. This task is mandatory to see the data into the Alflytics reports, dashboards and free analysis.
- **Publishing** static reports directly into your Alfresco platform with your preferred format (pdf, xlsx, txt, cvs, ecc.). You can avoid this task if you want to have interactive reporting.
- **Using analytics** on top of the Data Warehouse, using Pentaho and in particular: interactive reports, dashboards and free analysis.

2.1 Extracting from Alfresco

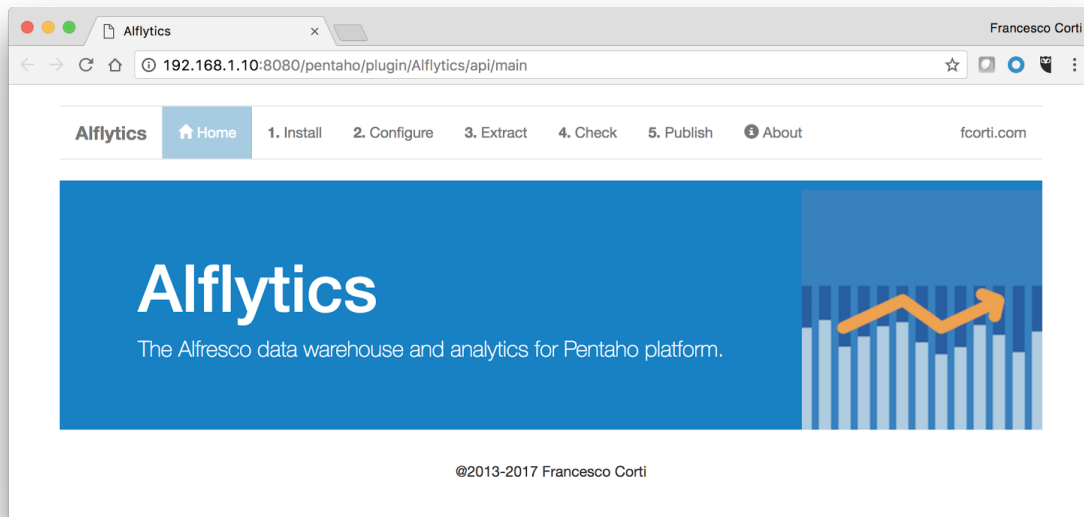
With the extraction task, a batch gets the data out from Alfresco into the Alflytics Data Warehouse. This task is mandatory to see the data into the interactive reports, dashboards and free analysis.

The extraction is requested through a web page, in a default installation, or executing a [Pentaho Job](#) if [Pentaho Data Integration](#) is used. In the following description, we will see how to extract data using a default installation and the Alflytics Administration Web pages.

The Alflytics Administration Web pages can be easily accessed using the following url, after the Alflytics solution is correctly installed (for further details on how to install Alflytics, check the related chapter).

```
http://<server>/pentaho/plugin/Alflytics/api/main
```

In the following screenshot, the Alflytics Administration Home page, used as starting point to manage the whole instance of Alflytics.



If you take a look at the upper menu, the item called `3.Extract` is the one used to get data from Alfresco. Click on it, to have access to the extraction page. As you can see directly in the target page, the extraction is very straightforward and does not require any difficult setup or question. Simply fill the Pentaho administrator login and password (by default `admin` and `password`) and click the `Extract` button to start the process.

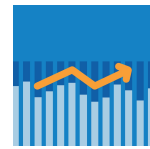


The `Extract` button only starts the extraction process, that could require few seconds or some minutes to be completed. If the web page replies with a result, this does NOT mean the extraction process is finished. The duration of the extraction process depends (mainly) on the Alfresco repository and the network.

Always check the `catalina.out` file stored into the `<pentaho-server>/tomcat/logs`, for further details on the extraction task.

Once the extraction task is finished, you can check how the extraction went, using the Alflytics Data Quality dashboard. You can access to the Alflytics Data Quality dashboard using the upper menu of the Administration Web pages. More in particular clicking the `4.Check` item.

The Alflytics Data Quality web page is available to check how the extractions went, with a easy to understand dashboards. In the dashboard you can select the Alfresco instance, the period to filter extractions and each extraction from an identifier, quite easy to read because based on a timestamp. Once the extraction id is selected, two tables show you the duration of the sub-extractions (Alfresco users, audits repository and processes) with their success/failure and a detail of the number of entities extracted. Below a screenshot showing how the Alflytics Data Quality web page look like.



Alflytics Home 1. Install 2. Configure 3. Extract 4. Check 5. Publish About fcorsi.com

Alfresco instance
Alfresco

Period
2017-04-05 > 2017-04-12

Extraction
source:Alfresco.id.1.at:2017-04-12.18:49:48:893

Extraction logs

Task	Start	End	Result
Alfresco users in staging	2017-04-12 18:49:50.0	2017-04-12 18:49:53.0	Success
Alfresco users in operational	2017-04-12 18:49:54.0	2017-04-12 18:49:55.0	Success
Alfresco users in warehouse	2017-04-12 18:49:56.0	2017-04-12 18:49:57.0	Success
Alfresco audit trail in staging	2017-04-12 18:49:57.0	2017-04-12 18:50:08.0	Success
Alfresco audit trail in operational	2017-04-12 18:50:09.0	2017-04-12 18:50:10.0	Success
Alfresco audit trail in warehouse	2017-04-12 18:50:11.0	2017-04-12 18:50:14.0	Success
Alfresco repository in staging	2017-04-12 18:50:17.0	2017-04-12 18:50:22.0	Success
Alfresco repository in operational	2017-04-12 18:50:23.0	2017-04-12 18:50:25.0	Success
Alfresco repository in warehouse	2017-04-12 18:50:26.0	2017-04-12 18:50:34.0	Success
Alfresco processes in staging	2017-04-12 18:50:37.0	2017-04-12 18:50:43.0	Success
Alfresco processes in operational	2017-04-12 18:50:44.0	2017-04-12 18:50:47.0	Success
Alfresco processes in warehouse	2017-04-12 18:50:49.0	2017-04-12 18:50:52.0	Success

Data quality

Entities	Alfresco	Staging	Operational	DataMart	Check
Users	8	8	8	8	Checked
AuditTrail	Cannot check	6245	6245	6245	Not checked
Nodes	279	279	279	279	Checked
Categories	337	337	337	337	Checked
OpenTasks	3	3	3	3	Checked
Processes	3 (not completed only)	4	4	4	Not checked

@2013-2017 Francesco Corti

You can check into this Alflytics Data Quality web page every time you want to understand how the extraction task went, also if the extraction is scheduled as an automatic job.

2.2 Publishing reports into Alfresco (optional)

Once the data are updated into the Alflytics Data Warehouse, you can decide to use the analytics (see the following paragraph for this purpose) or publish a static version of your reports (usually in PDF or Excel format), directly into the Alfresco repository.

Alflytics v5.0 Early Access

The Open Source Business Intelligence solution over Alfresco



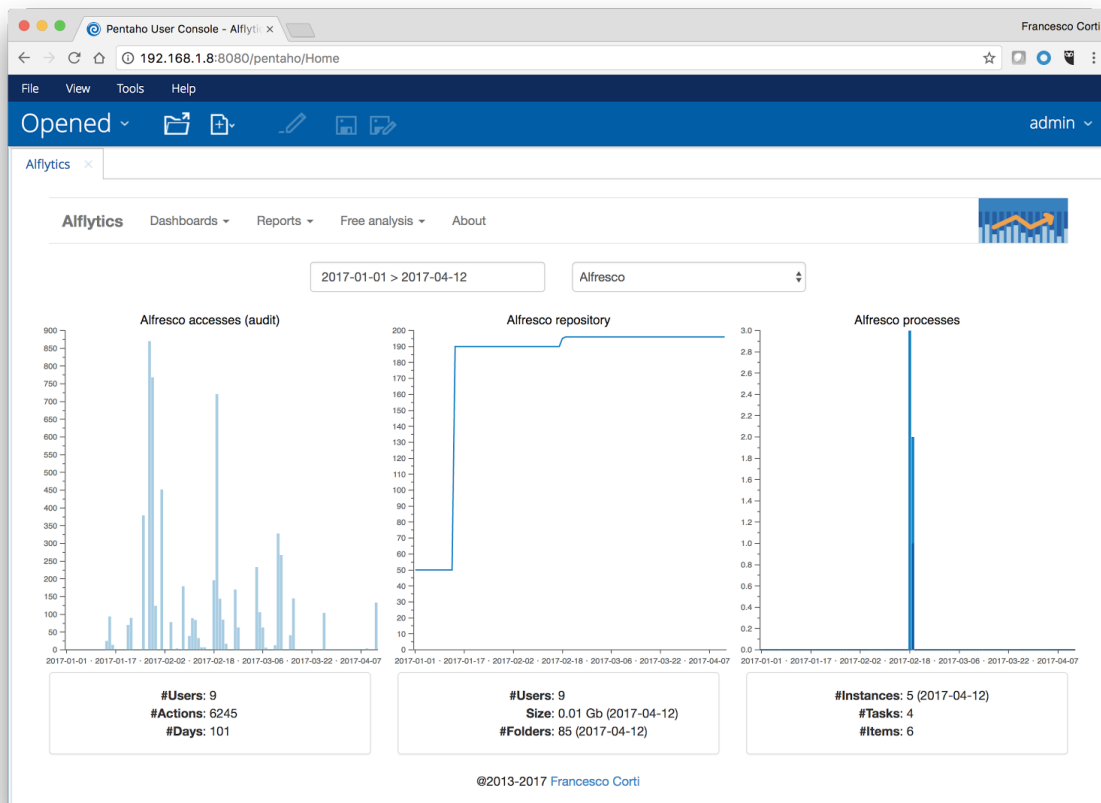
The publication of the reports can be done using a web interface or a command, launched from the terminal. The launch of the publication process from the web interface, is again very straightforward using the `5.Publish` item of the upper menu in the Alflytics Administration Home page.

Also in this case, the publication button only starts the process and you should always check the `catalina.out` file stored into the `<pentaho-server>/tomcat/logs`, for further details on the task.

2.3 Using analytics

Using a real Business Application you can benefit of several advantages like, for example, having a full featured platform available for your needs. Alflytics is released with a set of dashboards, interactive reports and free analysis, but these are only examples a solid starting point. You can easily build your own analytics tools, to better cover the use cases.

The starting point of you discovery is the main dashboard. The main dashboard is available using the Pentaho User Console, more in particular clicking the upper menu at the item `Tools`, and then `Alflytics`. Below a screenshot showing how the dashboard looks like.



If you want to integrate this dashboard into an external application or use it without the Pentaho User Console, you can have direct access using the following URL.



`http://<server>/pentaho/api/repos/%3Apublic%3AAlflytics%3ADashboards%3Amain.wcdf/generatedContent`

Directly into the dashboard you can see an upper menu, listing also reports and free analysis. Below a list of each available report and free analysis, as a first introduction.

- Reports
 - Alfresco users
 - Most active users
 - Audit per user (Excel)
 - Repository per creator
 - Repository per modifier
 - Repository per owner
 - Alfresco audit trail
 - Audit per actions
 - Most accessed content
 - Audit per day
 - Audit per hour
 - Audit trail (Excel)
 - Alfresco repository
 - Repository summary
 - Most used categories
 - Most used tags
 - Repository per category
 - Repository per tag
 - Repository per node type
 - Content per mime type
 - Creations per day
 - Modifications per day
 - Alfresco processes
 - Process instances
 - Process tasks



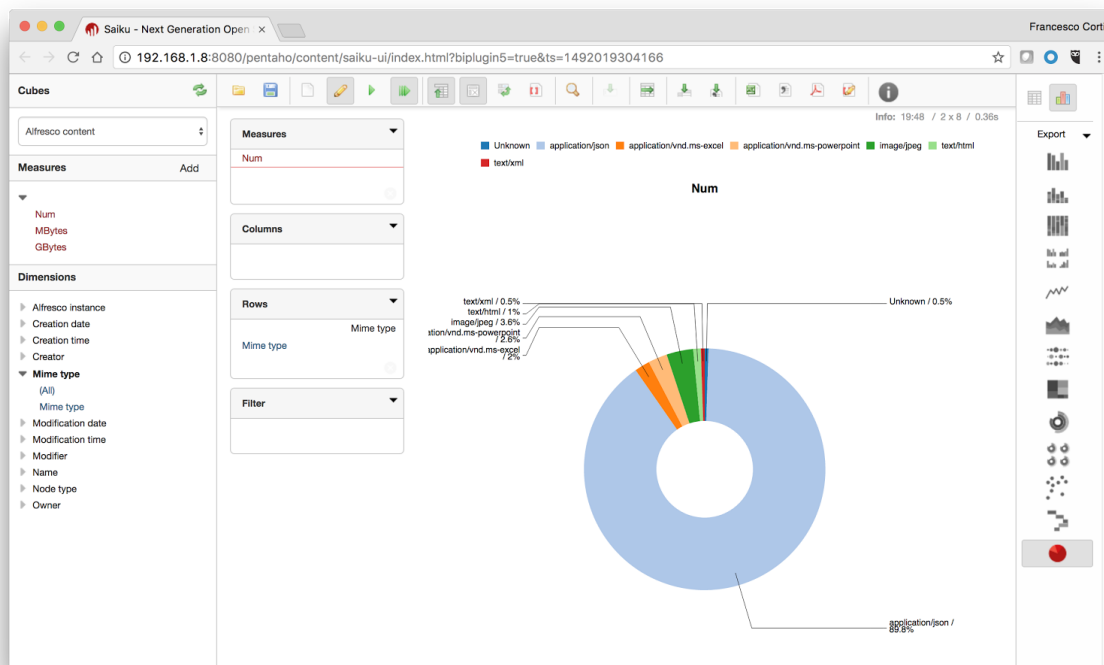
Free analysis will be available, if and only if the Saiku Analytics is installed.

- Free analysis
 - Alfresco audit trail
 - Alfresco audit trail (alfresco-access)
 - Alfresco repository
 - Alfresco content
 - Alfresco folders
 - Alfresco categories



- Alfresco tags
- Alfresco processes
 - Alfresco process instances
 - Alfresco process tasks

Please remember all the analytics are available for customizations and copies. This means the variety of analytics don't start and stop with this set, but can be improved (and customizes) as much as your fantasy will require.



3. How to install Alflytics

Installing Alflytics is very straightforward because it is released as a so called “Pentaho Application”, containing everything is requested to correctly configure and use the solution. No installations or configurations are requested on the Alfresco side and it can work on top of Pentaho Analytics platform, with an HyperSql, PostgreSQL or MySQL database. In this section we are going to described how to install it, using the wizard (simplest) or manually (advanced), depending on your needs.

3.1 Preparing the installation

As discussed above, Alflytics is a Pentaho Application developed using Sparkl (the application builder). In this section we are going to describe the setup of the working environment, to prepare the real installation of the Alflytics application.



3.1.1 Minimum requirements

Being a Pentaho Application deployed into the Pentaho Analytics platform, Alflytics minimum requirements are strictly related to the Pentaho version. Alflytics 5.0 in particular, runs on the Pentaho 7.0 platform and for this reason the minimum requirements are explicitly described into the [Pentaho's official documentation](#) and copied below as reference.

Hardware

Processor: Intel EM64T or AMD64 Dual-Core
RAM: 8 GB with 4 GB dedicated to Pentaho servers
Disk Space: 20 GB free after installation
(Data Warehouse is not included and can vary)

Operating System (64 bit)

Ubuntu Server 14.04 LTS & 16.04 LTS
CentOS 6 & 7
Red Hat Enterprise 6 & 7
SUSE Linux SLES 11 (SP3+)
Microsoft Windows 2008 Server R2 & 2012 Server R2



Alflytics is developed (and mainly tested) using a Linux platform, in particular Ubuntu Server 16.04 LTS. The adoption of a Linux based Operating System is highly recommended (because experienced as more stable) even there aren't reasons why the solution should not work on a Windows based OS.

3.1.2 Installing Pentaho Analytics platform

Installing Pentaho Analytics platform is very straightforward, very close to an “unzip” command of the distribution package. If you'd prefer a step by step tutorial, a simple [blog post](#) is what you need.

The only prerequisite for installing the Pentaho Analytics platform is having Java 8 on board. If you don't have it already or you don't how to install it, you can easily ask Google. It is a very basic task and there are a lot of tutorials in the web for every Operating System and distribution. If you are lazy to search, [this tutorial](#) is what can help you for a Linux platform.

After Java 8 is available into your environment you can go ahead downloading the Pentaho Analytics platform 7 from the [official website](#) or the [sourceforge web page](#). In our case, we are going to install the Pentaho Analytics platform 7 Community Edition.

Once the `pentaho-server-ce-7.0.0.0-25.zip` file is downloaded, you can unzip it into the desktop or everywhere else you will like (usually the `/opt` path is suggested for



Linux Operating System). All the Pentaho Analytics platform is now available in the `pentaho-server` folder.

Probably you cannot believe it, but this is enough to install Pentaho Analytics platform 7 into your environment.

Before considering completed this task, you should consider to reduce the default RAM, Pentaho is going to use. By default the Pentaho platform starts with 6Gb of RAM and in most cases is more than enough. If you don't have so much RAM into your environment, or you want more, you can tune it editing the `start-pentaho.sh/start-pentaho.bat` script, depending on your Operating System.

Pentaho Analytics platform 7 is mainly a web application deployed in a tomcat servlet container. You can easily run the Pentaho Analytics platform 7, opening a terminal and executing the `start-pentaho` script stored in the `pentaho-server` folder (the one with the `sh` extension for the Linux based Operating Systems and the one with the `bat` extension for the Windows based Operating Systems).

To check what is happening always look at the log file described below.

```
<pentaho-server>/tomcat/logs/catalina.out
```

Once, the application is started, you can access to the user interface using a web browser at the link below.

```
http://<server>:8080/pentaho
```

With Pentaho Analytics platform Community Edition is suggested to use Mozilla Firefox or Google Chrome instead of MS IE Explorer (not well supported today). To access as administrator you can use the `admin` user with password `password`.

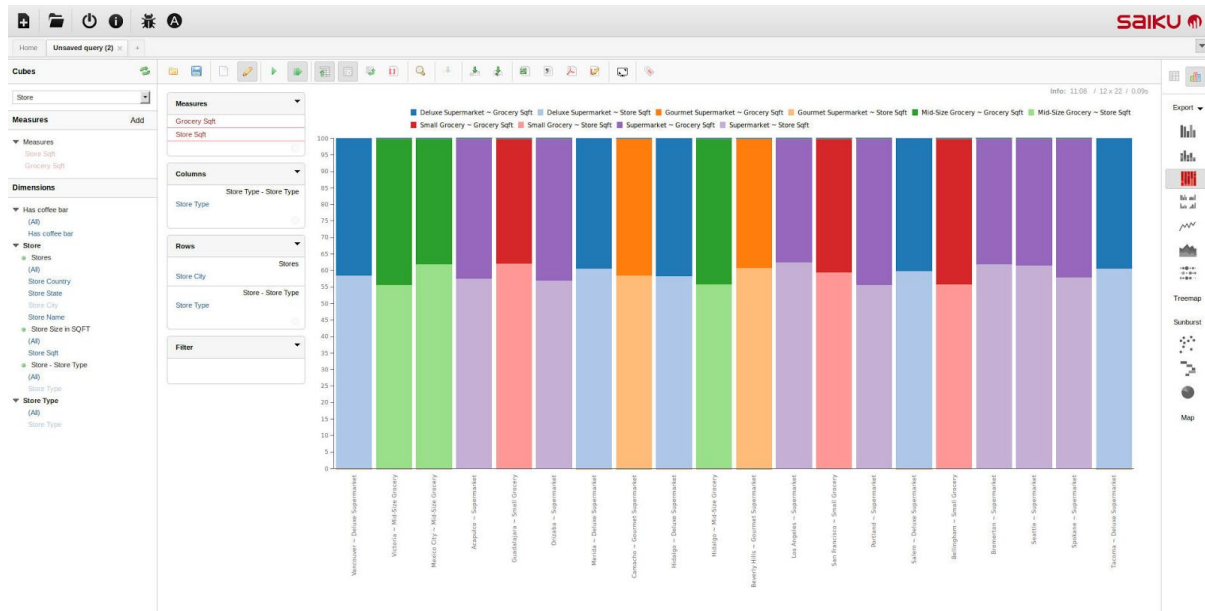
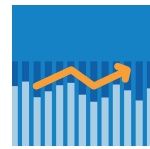
3.1.2 Installing Saiku analytics (optional)

[Saiku Analytics](#) allows business users to explore complex data sources, using a familiar drag and drop interface and easy to understand business terminology, all within a browser. Saiku analytics is used by the Alflytics solution to enable the user to the so called "free analysis". It is not mandatory to install (and use) the Saiku Analytics but it's recommended if you will want to use [OLAP](#) and [pivoting](#) over the data.



Alflytics v5.0 Early Access

The Open Source Business Intelligence solution over Alfresco

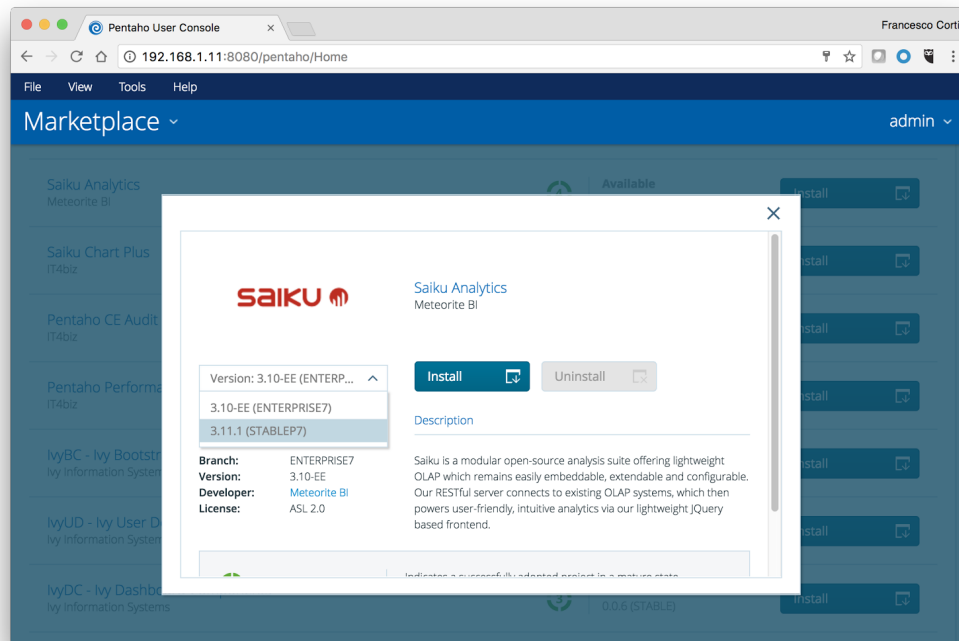
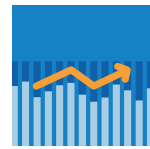


Saiku Analytics is released with a Community Edition and an Enterprise Edition. You can start evaluating the features of the Community Edition (free of charge) and upgrade to the Enterprise Edition later on.

Saiku Analytics installation is very straightforward, starting from an existing Pentaho Analytics platform. First of all access to the Pentaho User Console as administrator (otherwise you will not have permits to access to the marketplace). Click on the [Home](#) button and then on [Marketplace](#). In the marketplace list, search for the Saiku Analytics and be sure you are installing the latest Community Edition, not the Enterprise Edition. Below a screenshot showing how the user interface look like.

Alflytics v5.0 Early Access

The Open Source Business Intelligence solution over Alfresco



After selecting the latest Community Edition, click on the `Install` button. Once the Saiku Analytics is installed, a message box will tell you the result. As clearly described, a JAR library linked directly in the message box must be downloaded and copied into the `<pentaho-server>/tomcat/webapps/pentaho/WEB-INF/lib` folder.

Now it's time to stop the Pentaho Analytics platform, executing the `stop-pentaho` script stored in the `pentaho-server` folder.

Once the Pentaho Analytics platform is stopped, edit the file `<pentaho-server>/pentaho-solutions/system/importExport.xml`, adding the lines in bold described below.

```
...
<bean id="IRepositoryContentConverterHandler" ... >
  <constructor-arg>
    <util:map id="convertersMap">
      <entry key="mondrian.xml" value-ref="streamConverter"/>
      <entry key="jpeg" value-ref="streamConverter"/>
      <!-- Add the line below here!-->
      <entry key="saiku" value-ref="streamConverter"/>
    </util:map>
  </constructor-arg>
</bean>
...
<bean id="DefaultExportHandler" ... >
  <property name="repository" ref="unifiedRepository" />
</bean>
```



```
<property name="localeExportList">
  <list>
    <value>.xanalyzer</value>
    ...
    <!-- Add the lines below here! -->
    <value>.saiku</value>
  </list>
</property>
</bean>
...
<bean

class="org.pentaho.platform.plugin.services.importer.LocaleImport
Handler">
  <constructor-arg>
    <list>
      <value>xaction</value>
      ...
      <!-- Add the lines below here! -->
      <value>saiku</value>
    </list>
    ...
  </constructor-arg>
</bean>
```

Then edit the file `<pentaho-server>/pentaho-solutions/system/ImportHandlerMimeTypeDefinitions.xml`, adding the lines in bold described below.

```
...
<ImportHandler
class="org.pentaho.platform.plugin.services.importer.RepositoryFi
leImportFileHandler">
  <MimeTypeDefinitions>
    ...
    <MimeTypeDefinition mimeType="text/xml">
      ...
    </MimeTypeDefinition>

    <!-- Add the lines below here! -->
    <MimeTypeDefinition mimeType="application/json">
      <extension>saiku</extension>
    </MimeTypeDefinition>
  </MimeTypeDefinitions>
</ImportHandler>
```


Alflytics v5.0 Early Access

The Open Source Business Intelligence solution over Alfresco

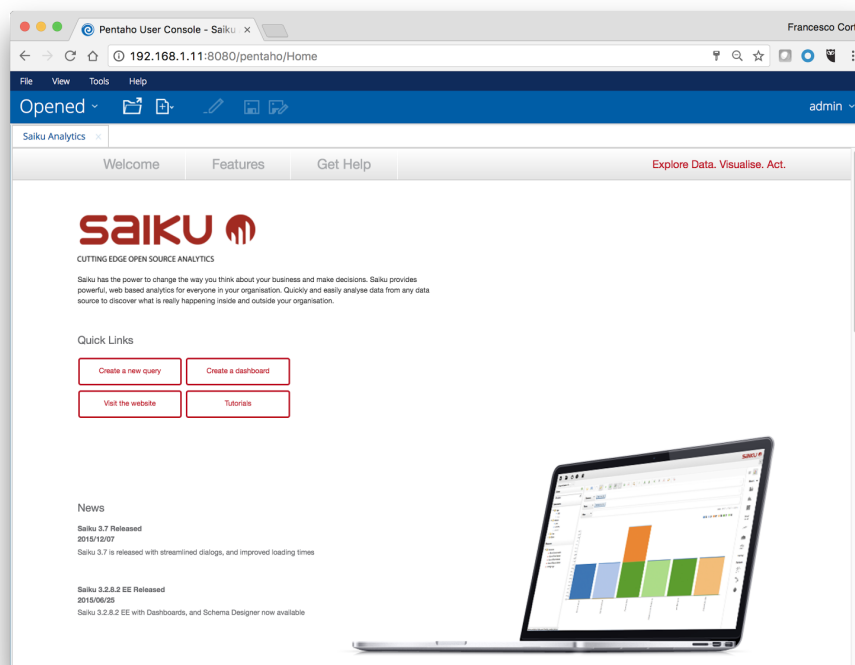


```
</MimeTypeDefinitions>  
</ImportHandler>
```

Last, but not least, there is the (free of charge) license to download and install into the `<pentaho-server>/pentaho-solutions/system/saiku` folder. Save the license file with name `license.lic`. To get the (free of charge) license, you can access directly to the url <http://licensing.meteorite.bi>. All the registration process and request of licenses for Community Edition is free of charge.

After this tasks are done, start again the Pentaho Analytics platform, executing the `start-pentaho` script stored in the `pentaho-server` folder.

To check if everything is working fine, access to the Pentaho User Console (<http://<server>:8080/pentaho>) and click on File, then New, then Saiku Analytics. Below a screenshot describes what the browser look like.



3.2 Getting Alflytics

There are two possible ways to get Alflytics: using the Pentaho Marketplace and downloading it manually. Below a brief description of each possible way. Both the possible paths get Alflytics into your Pentaho Analytics platform, with exactly the same result. Feel free to choose your preferred one, according to your network limitations and preferences.



3.2.1 Getting Alflytics using the Pentaho marketplace

Getting Alflytics using the Pentaho marketplace is similar to what we described for installing Saiku analytics.

First of all access to the Pentaho User Console as administrator (otherwise you will not have permits to access to the marketplace). Click on the `Home` button and then on `Marketplace`. In the marketplace list, search for the Alflytics. Click on the `Install` button and wait for the message box sharing the result of downloading.

Screenshot

Once the message box will come with a positive answer, it's time to restart the Pentaho Analytics platform executing the `stop-pentaho` script and the `start-pentaho` script later on. Both the scripts are stored in the `pentaho-server` folder.

3.2.2 Downloading Alflytics

Downloading and plugging Alflytics into the Pentaho Analytics platform is very straightforward.

To download the latest Alflytics distribution you can check into the [GitHub](https://github.com/fcorti/Alflytics/releases) project at the link below.

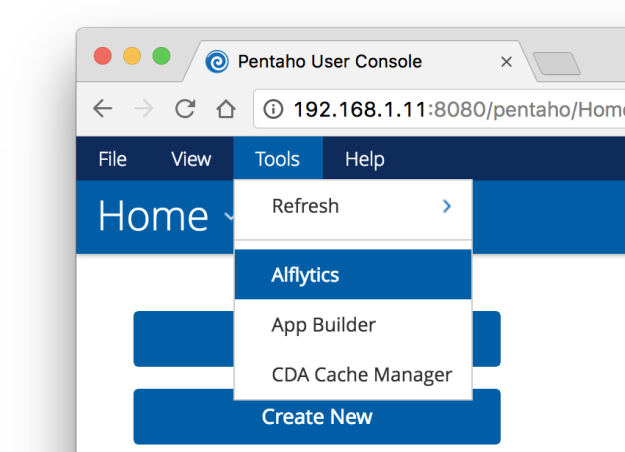
<https://github.com/fcorti/Alflytics/releases>

Download the zip file called `Alflytics_vX.X.zip` from the `Downloads` section and unzip it into the folder `<pentaho-server>/pentaho-solutions/system`. You should find then, a new folder called `Alflytics`.

Once the Alflytics distribution is plugged, it's time to restart the Pentaho Analytics platform executing the `stop-pentaho` script and the `start-pentaho` script later on. Both the scripts are stored in the `pentaho-server` folder.

3.2.3 Checking Alflytics

If you downloaded Alflytics using the Pentaho Marketplace or the manual setup, you can easily check the existence of Alflytics into your Pentaho Analytics platform, accessing to the Pentaho User Console and checking into the `Tools` item in the upper menu. If everything has been properly setup, the menu should contain an `Alflytics` item, as shown in the image below.



Clicking on the `Alflytics` item, a blank panel will be shown and an error will be logged into the `catalina.out`. In case you did it, don't worry, this is regular behaviour, considering you have not installed the dashboards yet.

3.3 Installing Alflytics

Now that Alflytics is plugged into the Pentaho Analytics platform, it's time to install everything is requested to have your fully featured Open Source Business Intelligence solution.

By default Alflytics comes with the Data Warehouse stored on a [HyperSql \(or HSQL\)](#) database. HSQLDB is a lightweight RDBMS written in Java, extremely useful for small databases and/or evaluation purpose. HSQLDB is particularly interesting for our use case because it is distributed in the Pentaho Analytics platform and does not require external or different installations.



[HyperSql](#) is recommended to be used on evaluation use cases only. If you plan to experiment Alflytics or to develop a POC on a small Alfresco repository, there are no concerns on using the default release. If you plan to use Alflytics in a production environment or a large repository, you should consider to use PostgreSQL or MySQL instead.

Below a detailed description of the tasks to follow to install the easiest version of Alflytics.

3.3.1 Accessing the Alflytics wizard

Alflytics has a wizard to make the installation task extremely easy. To access the installation wizard you simply have to access to the Administration Web pages, at the link below.

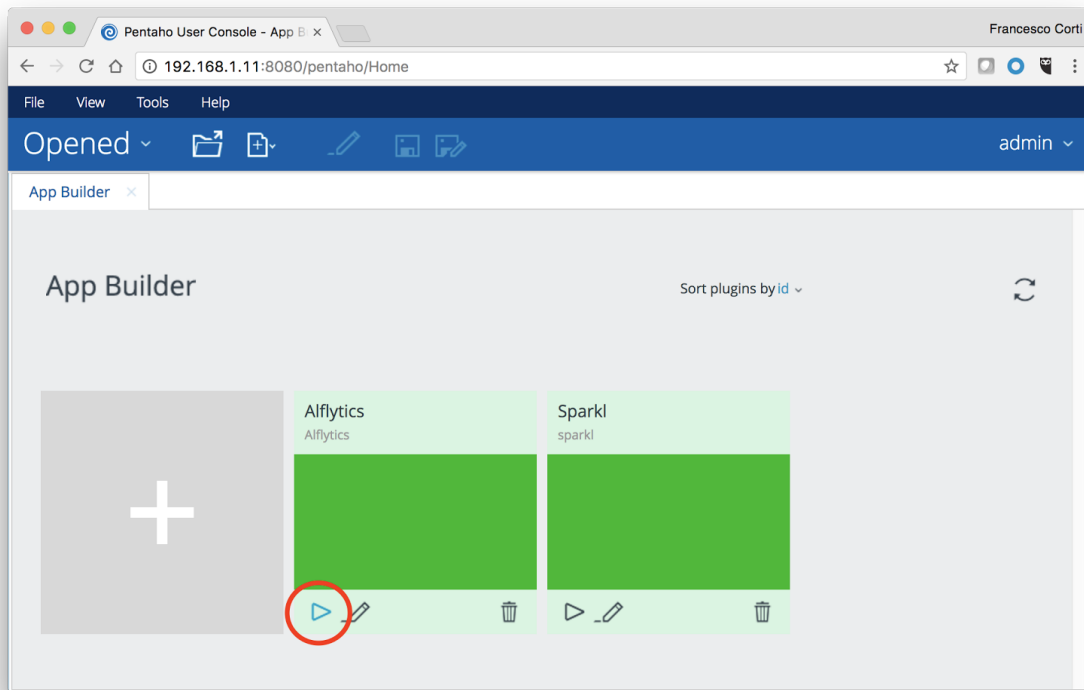
```
http://<server>/pentaho/plugin/Alflytics/api/main
```

Alflytics v5.0 Early Access

The Open Source Business Intelligence solution over Alfresco



As alternative, click on the `App Builder` item of the menu shown in the previous image (in the `Checking Alflytics` paragraph). Once the `App Builder` item is clicked, a panel will appear as described in the screenshot below.



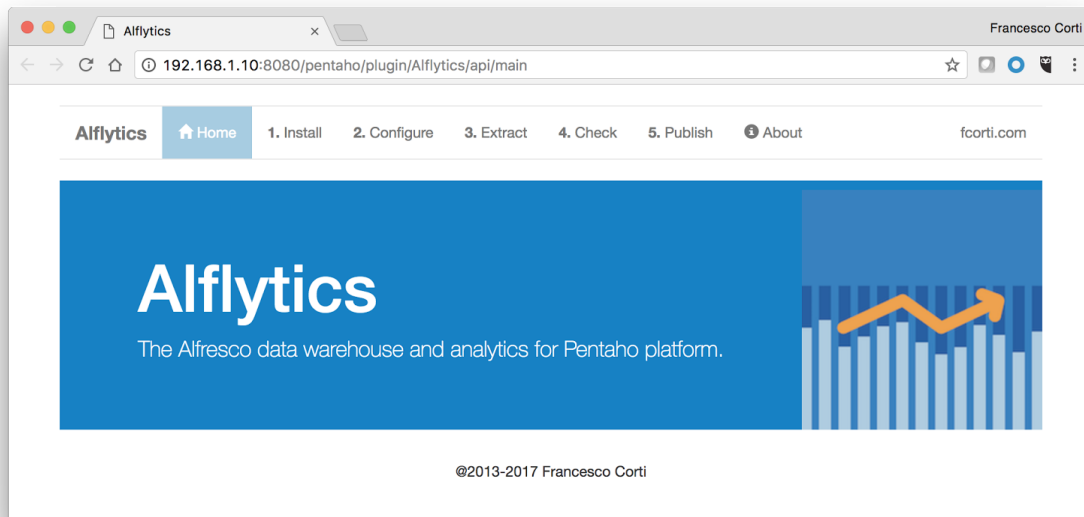
To access to the Alflytics wizard, click on the arrow, as highlighted in the screenshot.

3.3.1 Installing Alflytics using the wizard

The Alflytics Administration Web pages can easily be used, starting from the upper menu shown in the screenshot below.

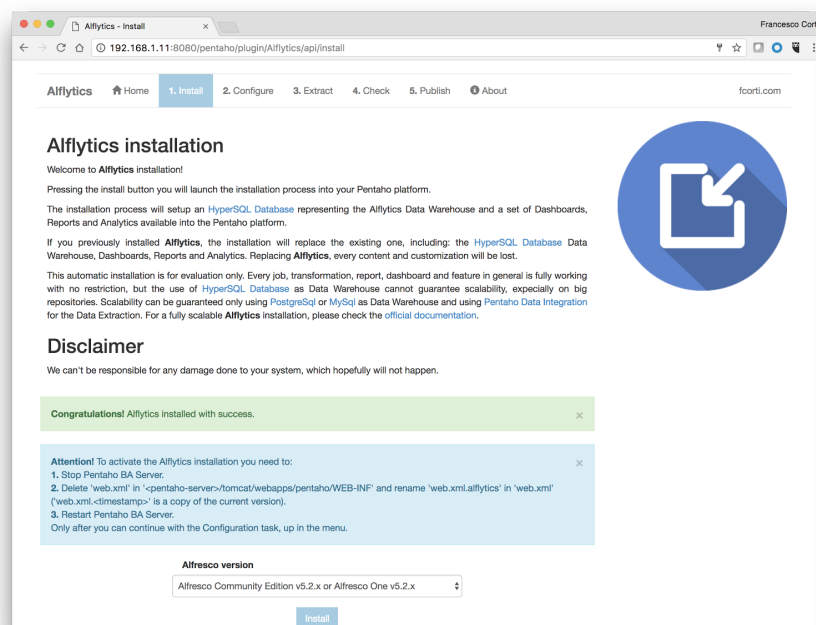
Alflytics v5.0 Early Access

The Open Source Business Intelligence solution over Alfresco



Each task is numbered to define the right order in terms of execution. The first task to follow, on the right of the `Home`, is about installing Alflytics. Click on the `1. Install` item to access to the installation page. As you can see directly in the page, the installation is very straightforward and does not require any difficult setup or question. Simply select the Alfresco version you are going to use as source of your data, and click the `Install` button.

Below a screenshot showing the result, if everything is successfully installed. As usual, if something goes wrong or you want to be sure about the details of the task, check the `catalina.out` of the Pentaho Analytics platform.





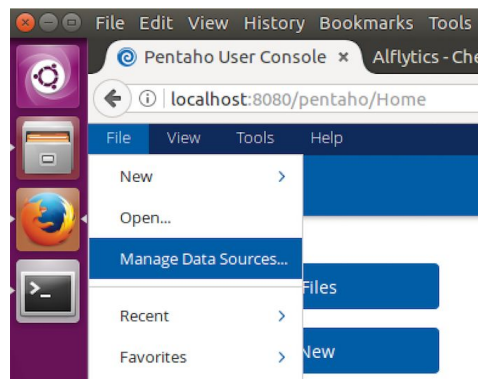
To activate the Alflytics installation, there are few tasks more to do. Below the list to follow.

First of all stop the Pentaho Analytics platform, executing the `stop-pentaho` script stored in the `pentaho-server` folder. Then delete the `web.xml` file in `<pentaho-server>/tomcat/webapps/pentaho/WEB-INF` and rename `web.xml.alflytics` as `web.xml` (`web.xml.<timestamp>` is a copy of the current version of the file). Restart Pentaho Analytics platform again, executing the `start-pentaho` script stored in the `pentaho-server` folder.



There is currently a not solved issue on the datasource creation. The good news is that it affects only the installation and can be easily solved with an extra-task. Below the step-by-step description of the extra-task mentioned above.

Access to the Pentaho User Console and click on the `Manages Data Sources` item in the upper menu.



Select the `Alflytics` data source of the `JDBC` type and edit it (using the gear icon). Once in the data source edit window, change all the fields according to the image below. The click on the `test` button to check everything is working fine, and then click on `Ok`.



Database Connection

General
Advanced
Options
Pooling

Connection Name:
Alflytics

Database Type:
SparkSQL
Cloudera Impala
Impala
Hadoop Hive 2
Generic database
H2
Hypersonic
MonetDB
MySQL

Access:
Native (JDBC)
ODBC
JNDI

Settings
Host Name:
localhost
Database Name:
alflytics
Port Number:
9001
User Name:
admin
Password:

Adding Databases

Test

OK Cancel

Congratulations, your Alflytics environment is now installed into your Pentaho Analytics platform. Now it's time to configure the Alfresco source, clicking the **2.Configuration** item, up in the menu.

3.3.2 Configuring the access to Alfresco

The access to Alfresco is relevant to extract the data into the Alflytics Data Warehouse (a database schema into the RDBMS). To complete the task, click on the **2.Configuration** item, up in the menu, and complete the requested fields. Once correctly filled, press the **Save** button and a message will tell you the result.



The configuration wizard updates the `DWH_ALF_DIM_INSTANCES` table into the Alflytics Data Warehouse (a database schema into the RDBMS). You can manually setup the `DWH_ALF_DIM_INSTANCES`, if you want to change the settings and have a full control.



In some cases is experienced that the fields are not correctly updated in the form, after the **Save** button has been pressed. This is caused by the database caching done by the Pentaho platform. If you are sure you filled the fields with the correct values and pressed the **Save** button, feel free to ignore this behaviour. As alternative, to be sure, use your preferred database client and access directly to the `DWH_ALF_DIM_INSTANCES` table, to check (and eventually modify) the fields.



After the configuration has been completed, it's time to extract the data from Alfresco and only them using dashboards, reports and free analysis. To understand how to do it, follow the chapter dedicated to "How to use Alflytics".

3.3.3 Troubleshooting

If you have any kind of issue during the installation and/or configuration tasks, the suggestion is to check the `catalina.out` file of the Pentaho Analytics platform. The log file is very detailed and always reports a lot of useful information. Only after a deep study of the `catalina.out` file, if you don't know how to solve the issue, try to ask for support as described into the "Contacts and support" chapter.

3.4 Manual installation of Alflytics (advanced)

Why you should do manual installation: scalability (postgresql and mysql and PDI), customizations.

3.4.1 Installing Alflytics on PostgreSQL

ATTENTION: Only data quality dashboard will work in the wizard.

- 5.1.1 Installing PostgreSQL and pgadmin3.
- 5.1.2 Creating the Alflytics schema.

<https://github.com/fcorti/alfresco-audit-analysis-reporting/tree/master/Pentaho/Alflytics-App-Pentaho-v7/endpoints/kettle/src/5.2/data/postgresql>

- 5.1.3 Updating the data quality dashboard.
- 5.1.4 Copying the repository files.
- 5.1.5 How to setup the default db in the jobs?
- ...how to extract data from Alfresco.
- ...how to publish data into Alfresco.

3.4.2 Installing Alflytics on MySQL

ATTENTION: Only data quality dashboard will work in the wizard.

TODO

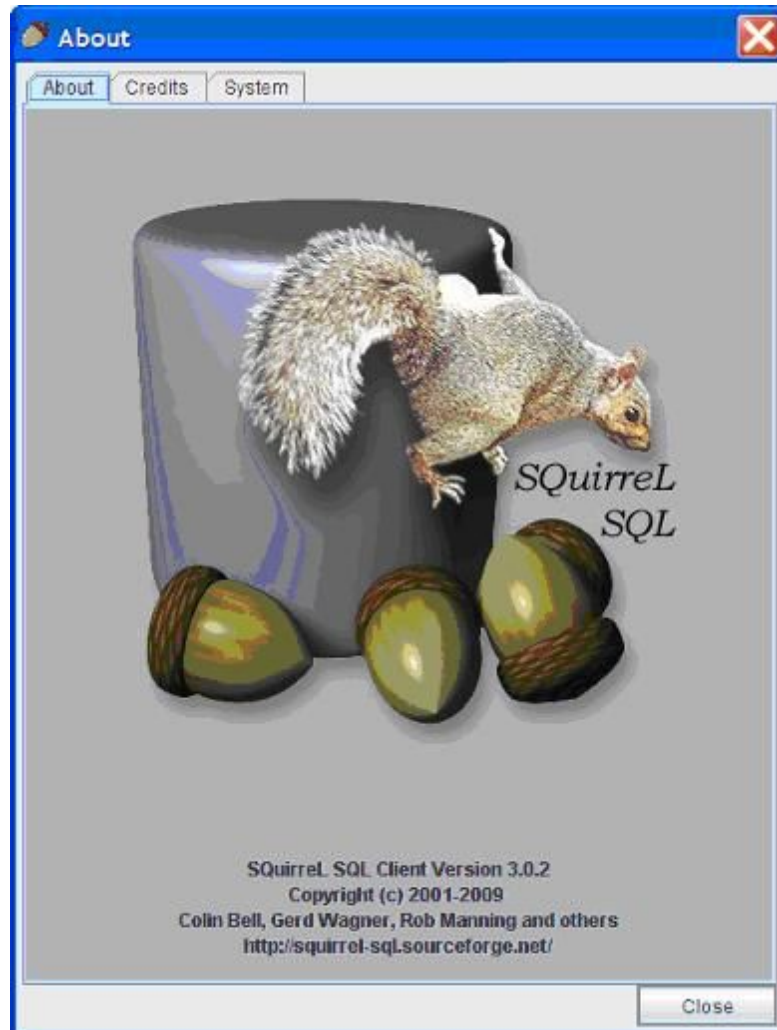
3.4.3 Installing Alflytics on HyperSql

You could decide to manually install alflytics on HSql.

TODO

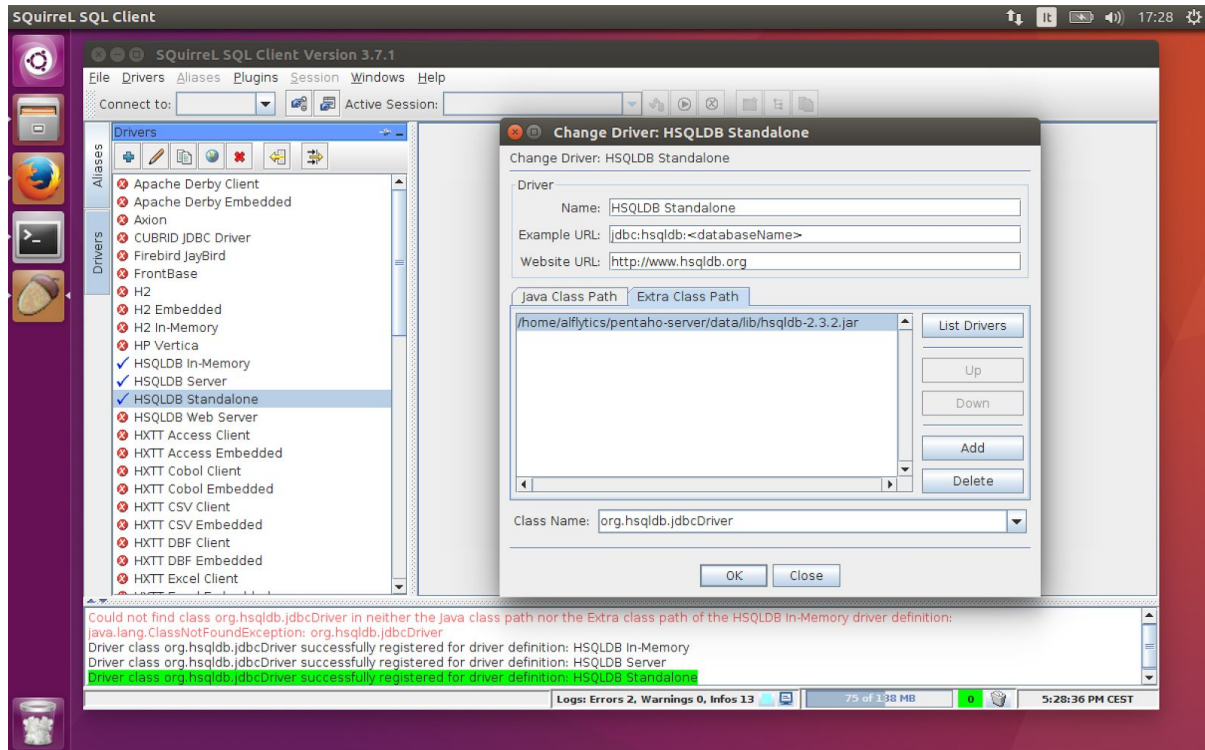
3.4.3.X Installing Squirrel Universal SQL Client

To have a full access to the Alflytics Data Warehouse, you need a database client. There are several useful database clients, and honestly you can decide to use your preferred one, but here we are going to explain how to install (and use) [Squirrel Universal SQL Client](#).



The official website and documentation is extremely clear and straightforward. Below are shared some further details on specific configurations, useful to our use case.

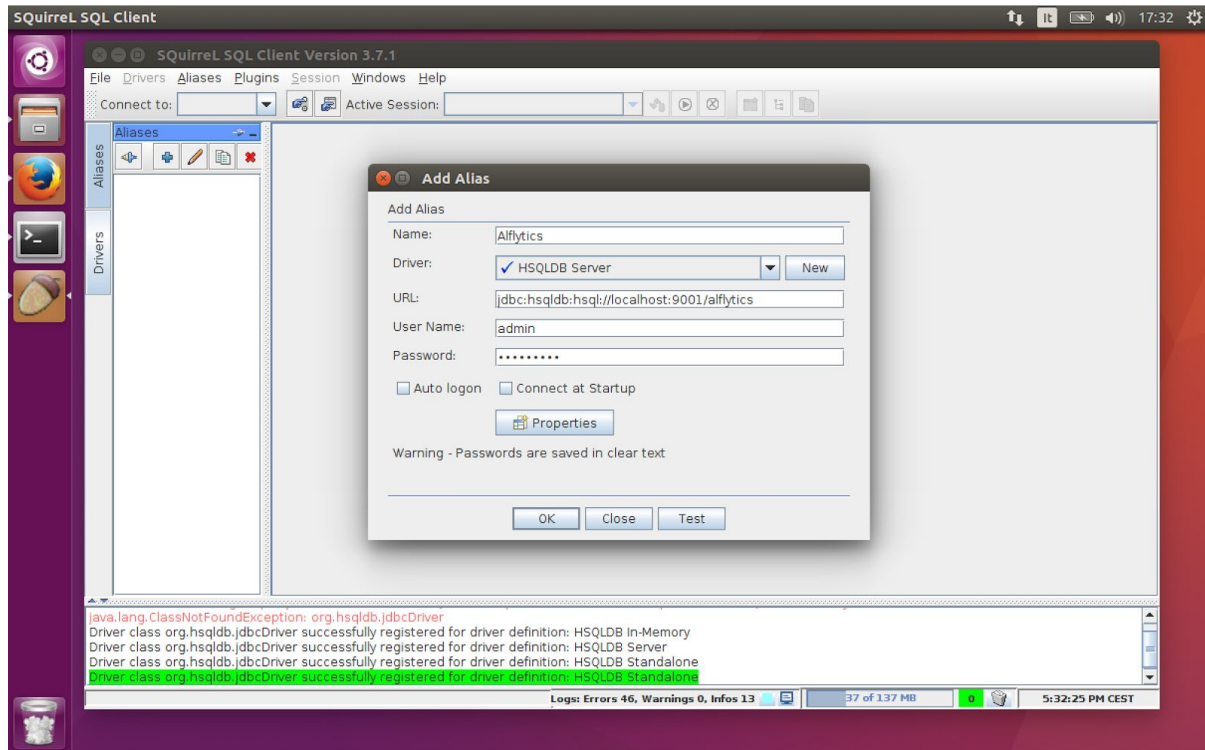
- The installation is a standard installation, as documented in the official [Squirrel SQL Client](http://squirrel-sql.sourceforge.net/) website.
- To have access to the Alflytics Data Warehouse on HyperSQL (or HSQLDB), you need to add few new drivers, using the Squirrel User Interface. From the driver list, identify the HSQLDB In-Memory, HSQLDB Server and HSQLDB Standalone. For each driver, modify it adding an extra class-path. The class-path to add is the JAR file in the `<pentaho-server>/data/lib/hsqldb-<version>.jar` path. Below a screenshot showing the user interface.



- To have access to the Alflytics Data Warehouse on HyperSQL (or HSQLDB), you need to add an alias, using the Squirrel User Interface. From the alias list, add a new one, with the settings below.

Name: Alflytics
Driver: HSQLDB Server
URL: jdbc:hsqldb:hsq://<server>:9001/alflytics
User Name: admin
Password: password

Clicking on the OK button and you will have access to the whole Alflytics Data Warehouse with a full control. Below a screenshot showing the user interface.



3.5 About scheduling extractions and publications

CURL sotto CRON.

4. Customizing Alflytics

What customizing means

4.1 Customizing reports

What customizing reports means

Installing Pentaho reporting

The available reports

Basis of Pentaho reporting

4.2 Customizing dashboards

TODO

4.3 Customizing the warehouse and the ETL

TODO

How to add custom properties, types and aspects.



How to customise audit trail.

How to integrate external sources (ERP, CRM, etc.)

5. Scalability

Alflytics is fully scalable, thanks to the use of a pure (and Enterprise) Business Intelligence solution. The Alflytics scalability is possible following different paths. Each path gives to the architecture a “boost” on a specific topic or components. In the following paragraphs we are going to describe all the available paths, with dedicated descriptions and setup.

5.1 The Alflytics Data Warehouse on PostgreSQL or MySQL

This scalability is given by the use of a scalable database, storing the data coming from the sources. In the previous chapters we saw how to install Alflytics using a PostgreSQL database or a MySQL database. By their nature, those two databases are highly scalable and guarantee great performances on huge repositories.

We won't treat more than this here, the scalability topic using PostgreSQL or MySQL, but it is highly recommended to adopt them, especially in production environments or huge repositories (also in POC or evaluation tasks).

5.2 Tuning the Pentaho Analytics platform

JAVA_OPTS...

5.3 Using the Pentaho Data Integration for ETL

Why a PDI is useful

Introducing the ETL server using PDI.

5.3.1 Installing Pentaho Data Integration

Reference: <http://fcorti.com/2016/12/05/installing-pentaho-data-integration-7/>

Unzip the distribution package.

`sudo apt-get install libwebkitgtk-1.0-0`

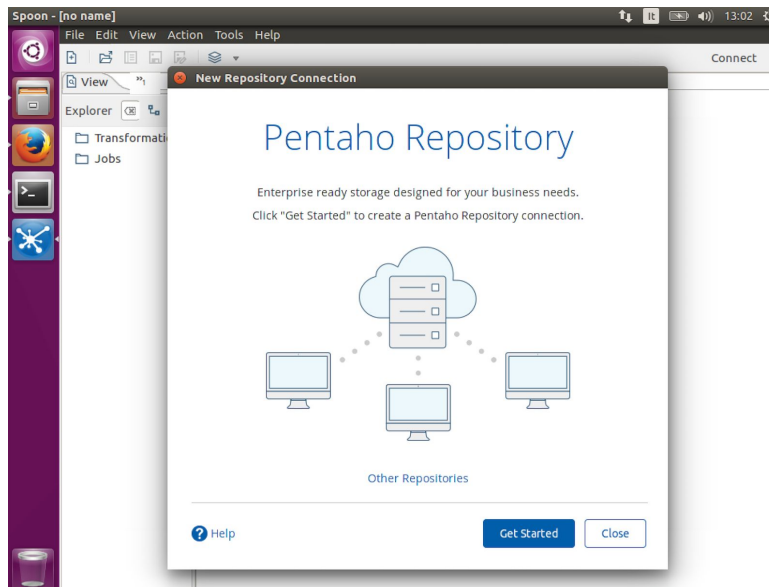
Run spoon script

5.3.2 Creating the connection to the Alflytics repository

Before using the Pentaho Data Integration for the first time, you have to setup the connection to the Alflytics repository. The task is extremely straightforward and can be done using the Spoon User interface.



From the main window, on the top right you should see a `Connect` button. Clicking on it, you will see a window appear, similar to the one in the following screenshot.



Click on `Get started` and fill the `Display name` (digit `Alflytics` or choose your preferred name) and URL. Be sure you digit the correct Pentaho Analytics Server URL. Click on the `Finish` button to complete the task. Don't connect immediately, but click on the `Finish` button to go to the main window of the Pentaho Spoon User interface.

Now click again on the `Connect` button on the top right of the main window, and select the new connection (in our case we named `Alflytics`). The login and password requested are the Pentaho Analytics ones (by default `admin` as user Name and `password` as Password).

Once connected, click on the `File` item in the upper menu and then `Open`. Under the following path you can find the root of all the jobs and transformations of the solution.

```
/public/Alflytics/ETL
```

`Get all` is the main job used to extract the whole Alfresco data. You can studying this, for a better comprehension of how Alflytics works.

5.3.3 Creating and scheduling the scripts to run the ETLs

Introducing kitchen

Creating the scripts with kitchen

Extract and Publish

How to schedule the scripts.



6. How to contribute to the project

TODO

7. Support and contacts

TODO

7.1 Disclaimer

The software is given as is, no support or responsibilities

7.2 Preparing the request of support

Dump your alflytics database instance.

Zip the /tmp/Alflytics folder.

Zip the catalina.out file (or the part containing the error).

7.3 How to get support from the community

Write into the community.alfresco.com

Write an issue on the github repository

Write to Francesco.

7.4 How to get professional support

Write to Francesco.

Appendix - FAQs

TODO