Saagie©

Platform Quick Guide

— v0.10 Luna-9 —

Table of contents 1/2

Table of contents 2/2

Luna-9 : new release, new features

History:

Luna 9, internal designation Ye-6 No.13, was an unmanned space mission of the Soviet Union's Luna programme. On the 3 February 1966 the Luna 9 spacecraft became the first spacecraft to achieve a soft landing on the Moon, or any planetary body other than Earth, and to transmit photographic data to Earth from the surface of another planetary body.

New features:

- New "Revolution" UI
- Create a job from another
- New panel to display capsules connection details
- Set resources (CPU, RAM, DISK) for all kind of jobs

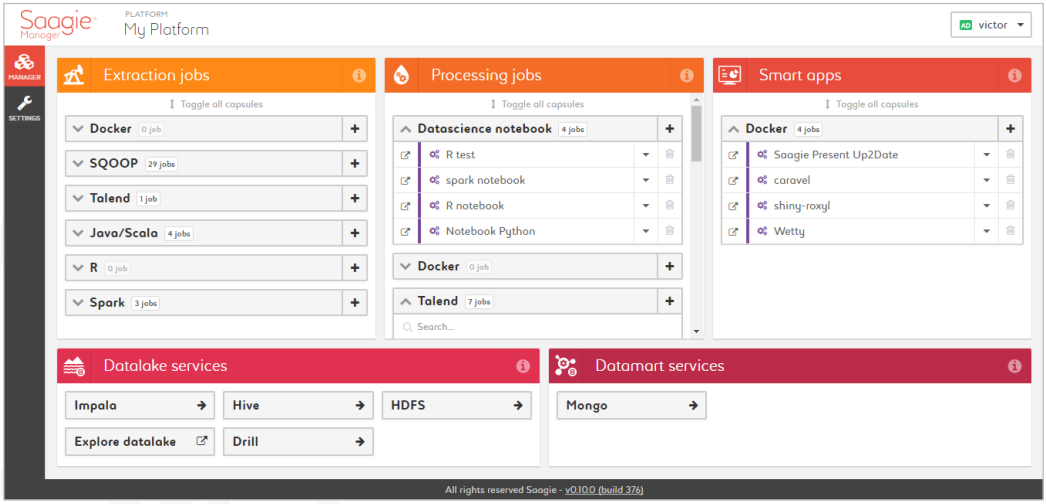
Capsules:

- Datalake: Impala - upgrade to version 2.5
- Datalake: Drill 1.7
- Datamart: MySQL 5.7

Manager 1/19

Your Platform Manager

https://manager.prod.saagie.io/



Available capsules

Extract


Talend

Datalake

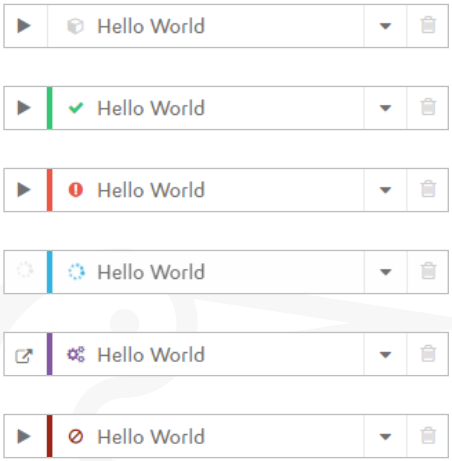
Processing

Datascience
with jupyter

Datamart

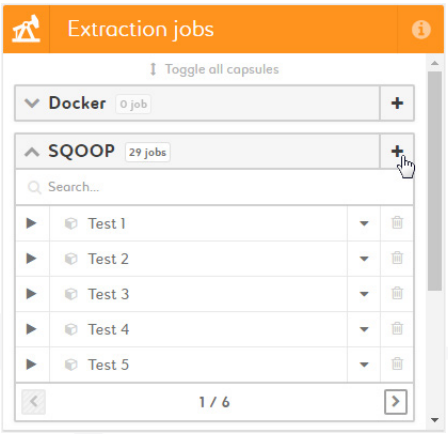
Dataviz

Job status



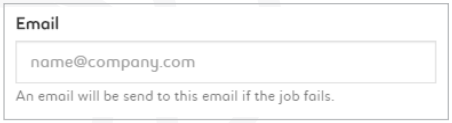
1. Job never launched
2. Job successful
3. Job failed
4. Job processing
5. Job pending
6. Job killed

Create a job

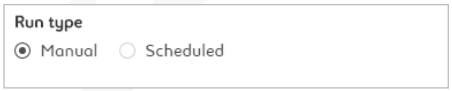


1. Click on the "+" symbol next to the capsule name

2. Name your job (required)

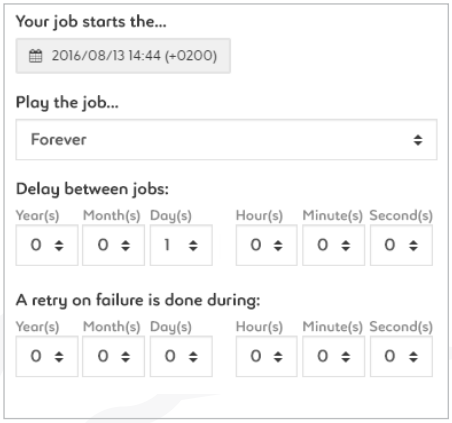


3. Enter an email address to be alerted if and when the job fails



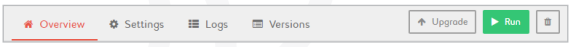
4. Run your job manually or set up a schedule

Create a job - Scheduling



1. Select the date at which the job will start
2. Select how many times the job will play
3. Select a delay between each jobs (required)
4. Set up timing for a retry in case the job fails

Access options within a job overview

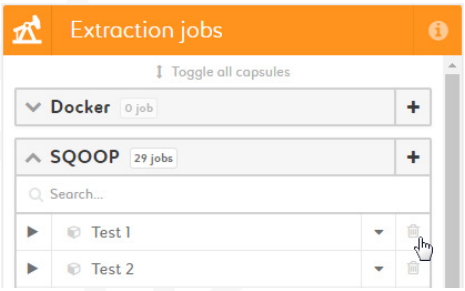


1. Select the action you want to perform
2. If in mobile view, click on the "Actions" button located in the top bar of a job overview

Search for a job

1. Click on the search box
2. Type the name of the job you are searching for

Delete a job

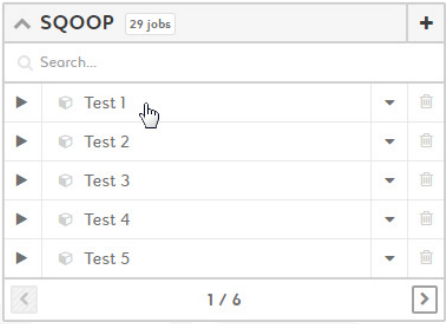


1. Click on the "bin" icon next to a job
2. Confirm your action by clicking on the “Delete” button

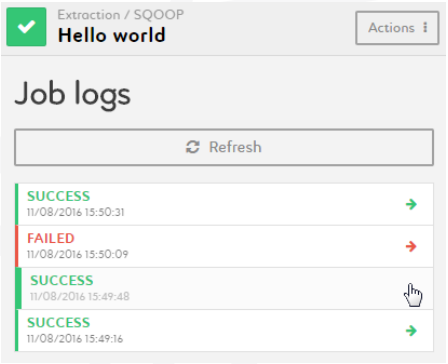
Run a job

1. Click on the "play" icon
2. To stop your job, click on the "stop" icon

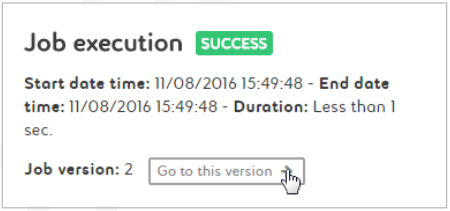
Display logs 1/2



1. Click on a job to display the logs screen

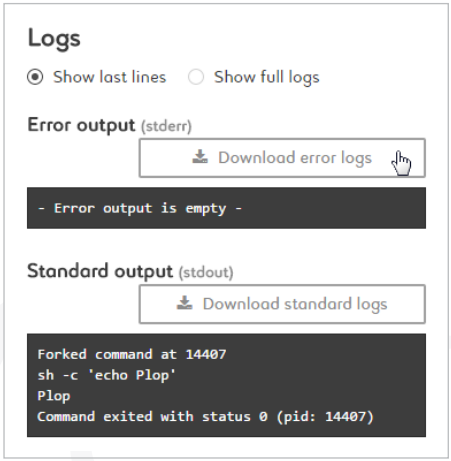


3. Select which version of the job you want to display logs for



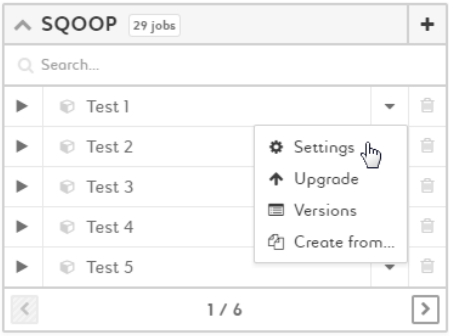
4. Click on “Go to this version” to rollback to a specific version of a job

Display logs 2/2



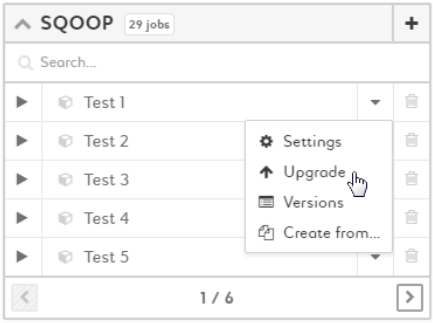
5. Select between only showing the last lines of the logs or full logs
6. Click on “Download error/standard logs” to download your logs or read them through the online viewer

Edit settings

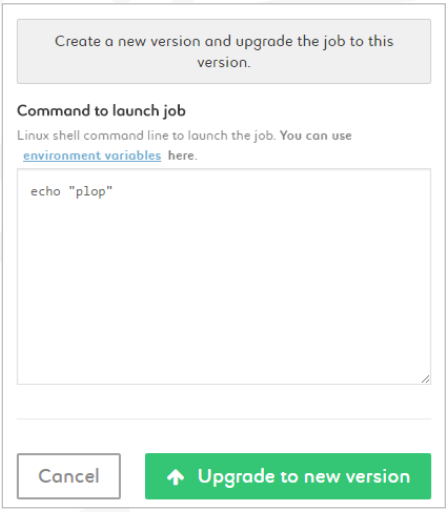


1. Click on the arrow to display the options dropdown menu
2. Click on “Settings” to edit settings for a specific job
3. Edit the email address used for alerts and the run type (manual or scheduled)

Upgrade a job

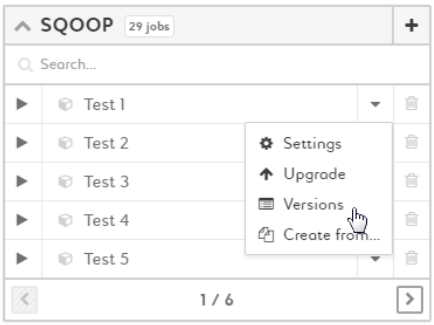


1. Click on the arrow to display the options dropdown menu
2. Click on “Upgrade” to access the job upgrade display

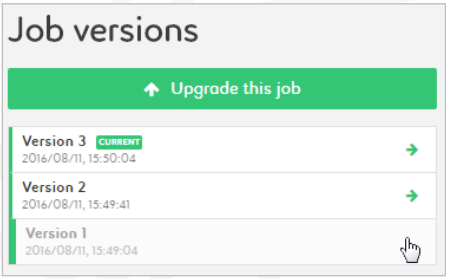


3. Edit your command
4. Click on “Upgrade to new version” to set this version as the current version of your job

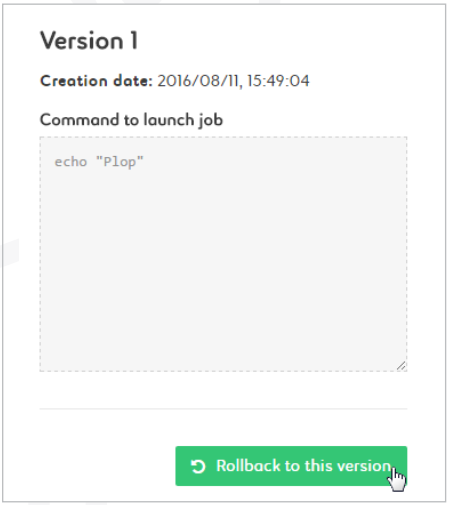
Manage versions of a job



1. Click on the arrow to display the options dropdown menu
2. Click on “Versions” to access the job versions display

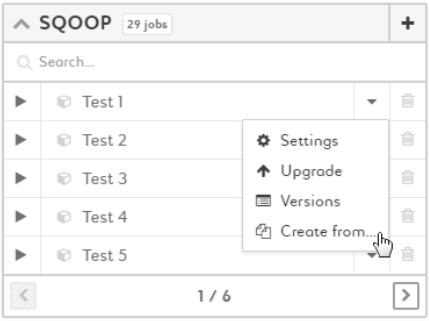


3. Select which version of the job you want to display



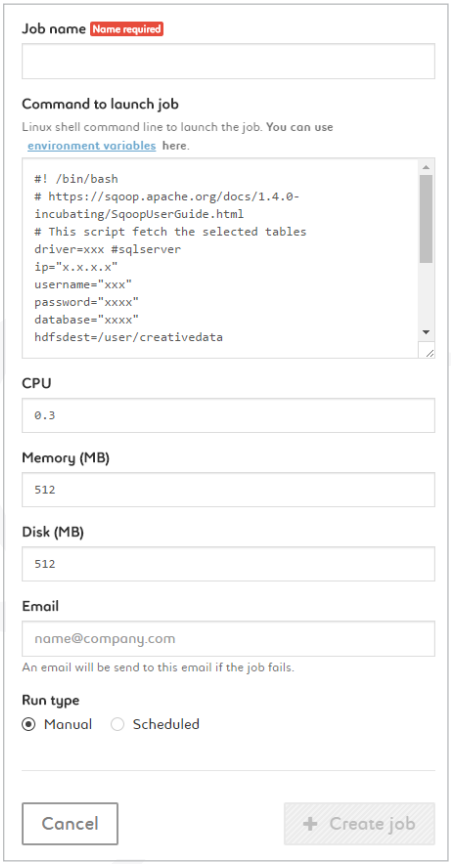
4. Check the details of the version and click on "Rollback to this version" if you want to set this version as the current version of the job
5. You can download the package of the job if one has been uploaded

Duplicate a job



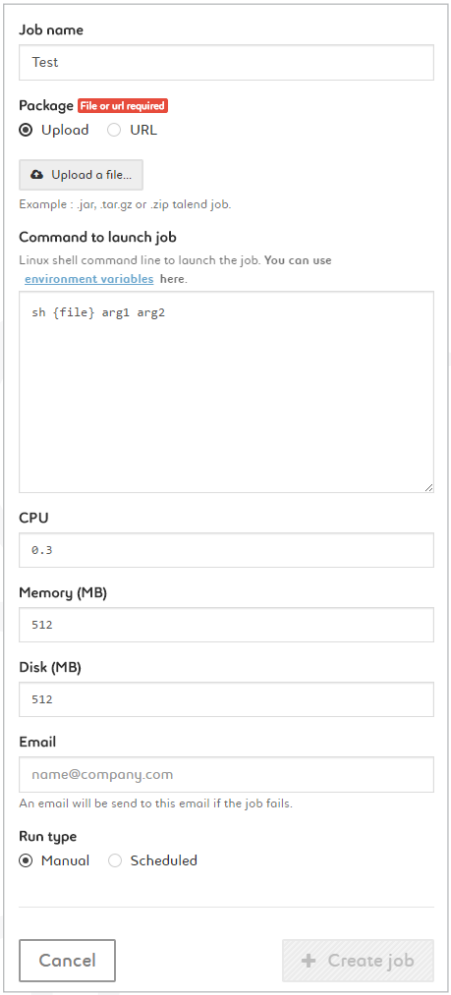
1. Click on the arrow to display the options dropdown menu
2. Click on “Create from...” to duplicate the selected job
3. Set up your job as explained in the "Create a job" section of this guide

Create a Sqoop job



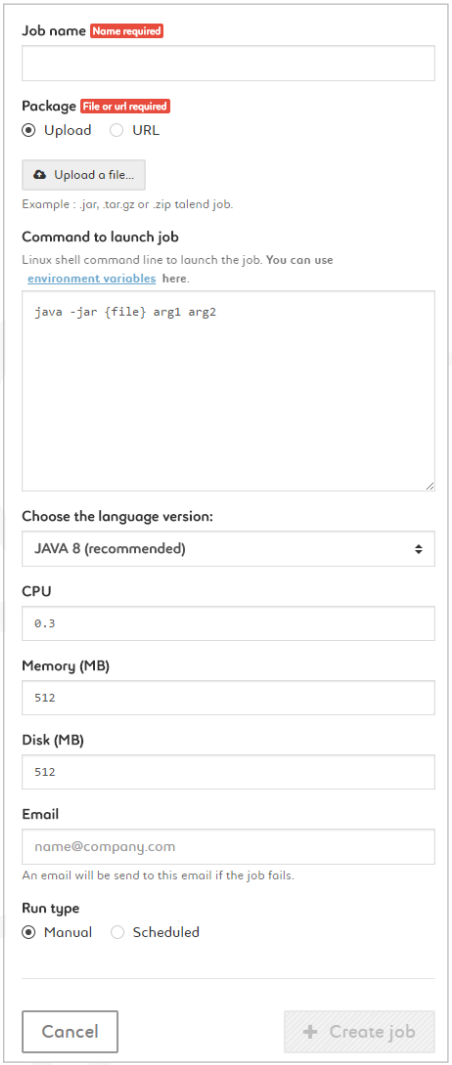
1. Name your job
2. Type your command using the ready-to-use template provided. Add environment variables as $VARIABLE or access your variable library by clicking the link
3. Set your CPU, memory and disk settings
4. Enter an email address to be alerted if and when the job fails
5. Run your job manually or set up a schedule

Create a Talend job



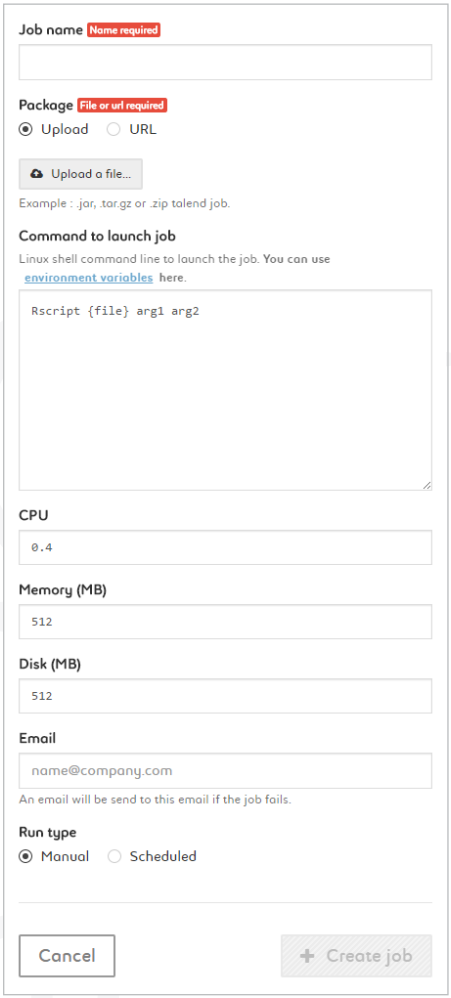
1. Name your job
2. Add a package by uploading a zip of by entering a URL
3. Type your command. {file} is a parameter (don't change it). Customize arg1, arg2, etc. Add environmnent variables as $VARIABLE or access your variable library by clicking the link
4. Set your CPU, memory and disk settings
5. Enter an email address to be alerted if and when the job fails
6. Run your job manually or set up a schedule

Create a Java/Scala job



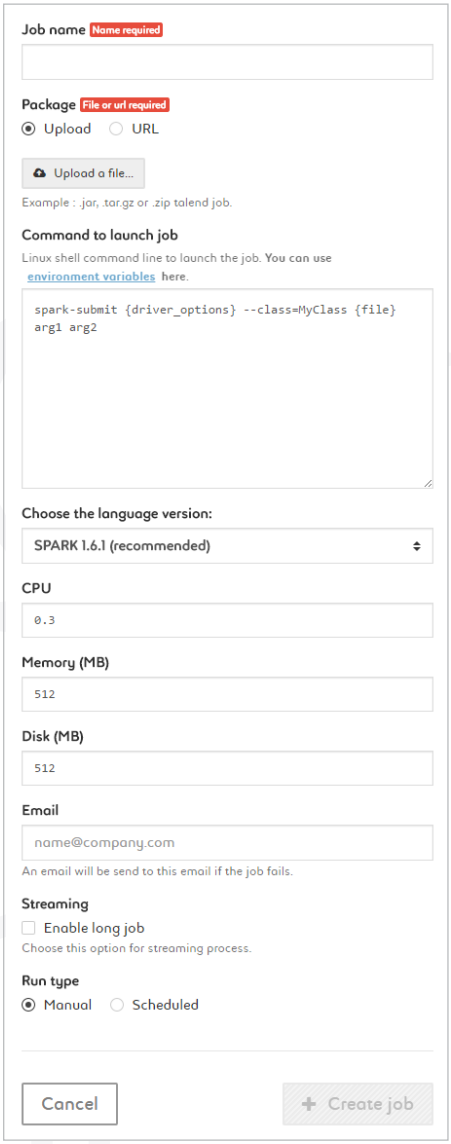
1. Name your job
2. Add a package by uploading a jar of by entering a URL
3. Type your command. {file} is a parameter (don't change it). Customize arg1, arg2, etc. Add environmnent variables as $VARIABLE or access your variable library by clicking the link
4. Choose the language version: JAVA 8 (recommended) or JAVA 7
5. Set your CPU, memory and disk settings
6. Enter an email address to be alerted if and when the job fails
7. Run your job manually or set up a schedule

Create a R job



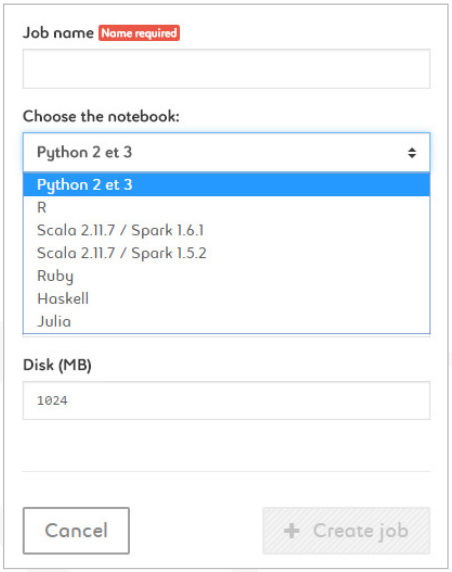
1. Name your job
2. Add a package by uploading a R file of by entering a URL
3. Type your command. {file} is a parameter (don't change it). Customize arg1, arg2, etc. Add environmnent variables as $VARIABLE or access your variable library by clicking the link
4. Set your CPU, memory and disk settings
5. Enter an email address to be alerted if and when the job fails
6. Run your job manually or set up a schedule

Create a Spark job



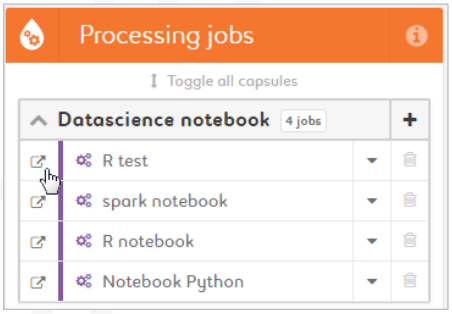
1. Name your job
2. Add a package by uploading a jar of by entering a URL
3. Type your command. {file} and {driver_options} are parameters (don't change it). Customize arg1, arg2, etc. Add environmnent variables as $VARIABLE or access your variable library by clicking the link
4. Choose the language version: Spark 1.6.1 (recommended) or Spark 1.5.2
5. Set your CPU, memory and disk settings
6. Enter an email address to be alerted if and when the job fails
7. Activate the streaming option
8. Run your job manually or set up a schedule

Create a Datascience Notebook



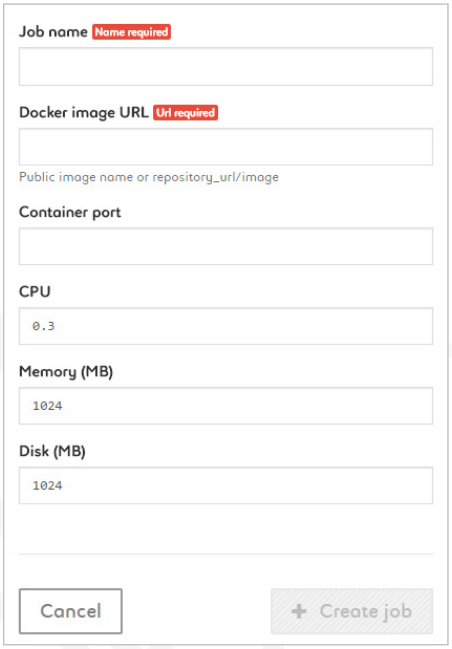
1. Name your job
2. Choose the notebook: Python 2 & 3, R, Scala 2.11.7 / Spark 1.6.1, Scala 2.11.7 / Spark 1.5.2, Ruby, Haskell, Julia
3. Set your CPU, memory and disk settings

Access a Datascience Notebook



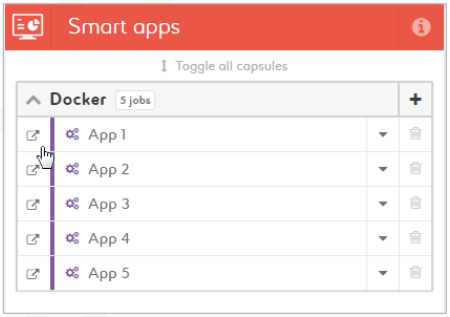
1. Click on the "Open in new window" icon next to a notebook

Create a smart app using Docker



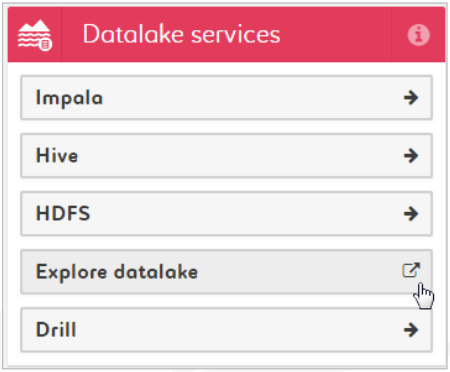
1. Name your job
2. Type the Docker image URL
3. Type the container port depending on the EXPOSE at the end of your Dockerfile
4. Set your CPU, memory and disk settings depending on your app/ Docker image

Open a smart app



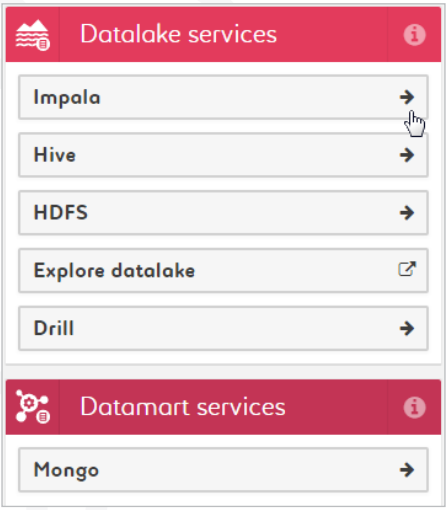
1. Click on the "Open in new window" icon next to a smart app

Explore your datalake with HUE



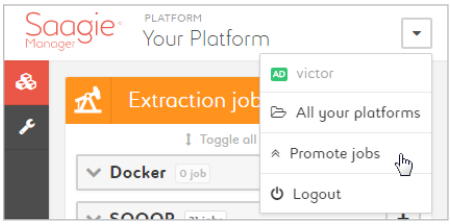
1. Click on the "Explore datalake" button in the "Datalake services" module
2. Explore your datalake in the now opened HUE interface

Show connection details

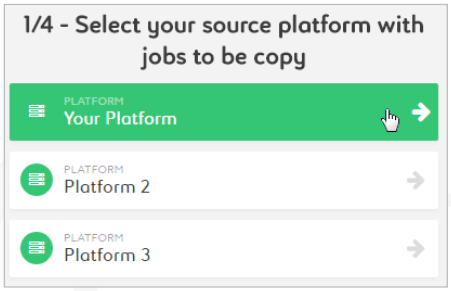


1. Click on any datalake or datamart capsule to display its connection details

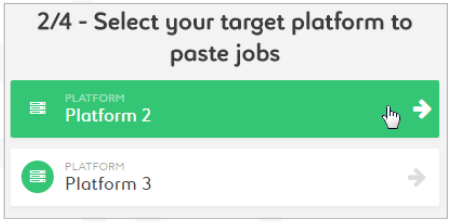
Promote jobs between platforms



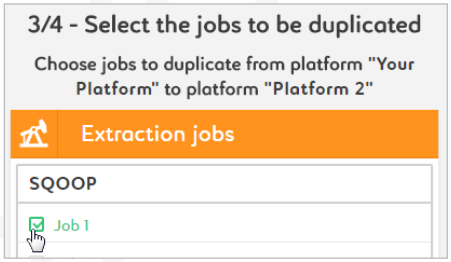
1. Click on the arrow at the right of the manager's top bar
2. Click on "Promote jobs"



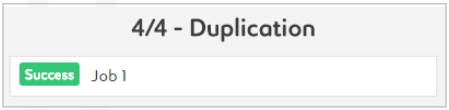
3. Select the platform you want to copy jobs from



4. Select the platform you want to copy jobs to

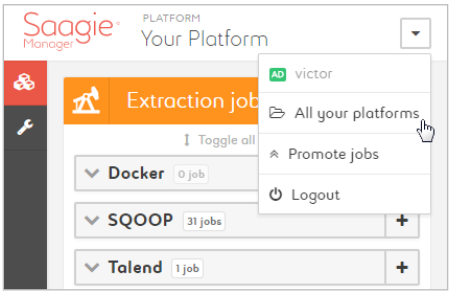


5. Select the jobs you want to copy
6. Confirm copy at the bottom of the page

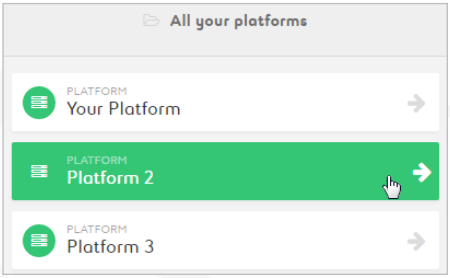


7. Check the status of your copy

Switch between platforms



1. Click on the arrow at the right of the manager's top bar
2. Click on "All your platforms"



3. Select the platform you want to switch to

ACL with Sentry

Option on demand

//create role

CREATE ROLE lob1analyst;

// assign role to group

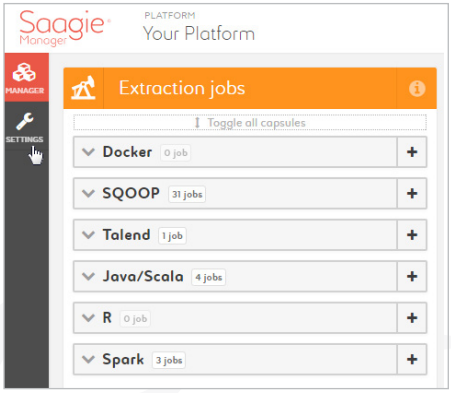
GRANT ROLE lob1administrator TO GROUP lob1adm WITH GRANT OPTION;

// manage access

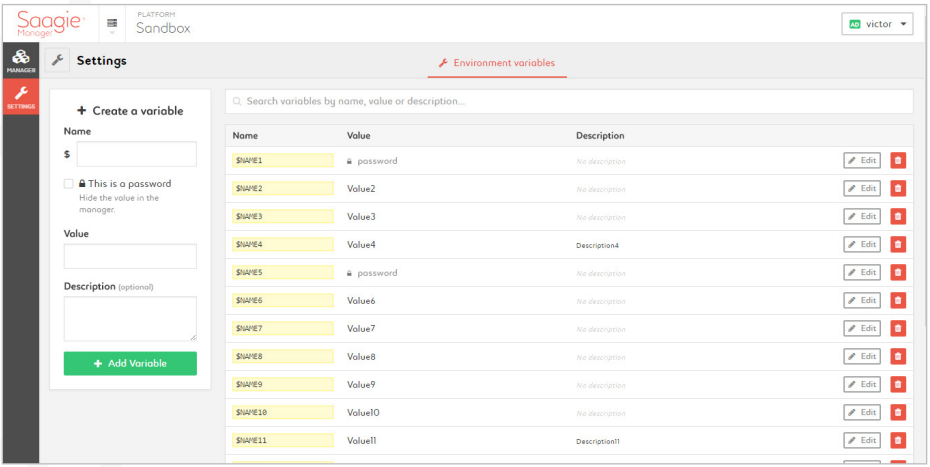
GRANT ALL ON DATABASE lob1 TO role lob1administrator;

Environment variables 1/3

Access the Environment variables display

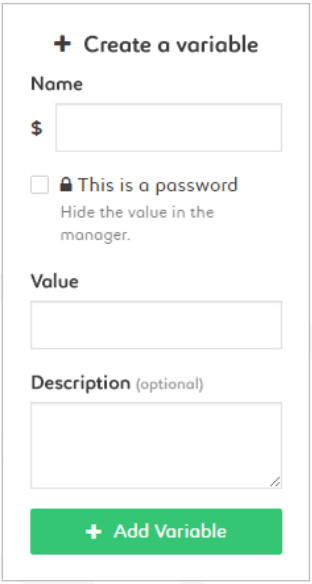


1. From the Platform Manager, click on the "Settings" button in the upper-left corner



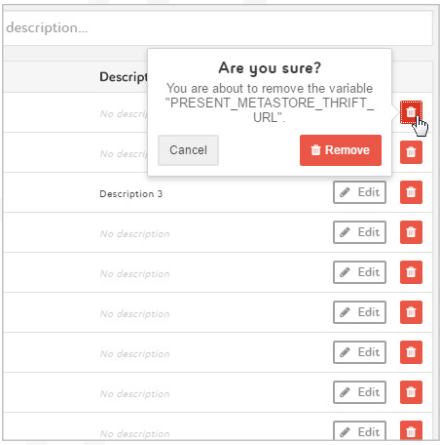
Environment variables 2/3

Create a variable



1. Name your variable
2. Select "This is a password" to hide the value of your variable in the manager
3. Enter your value
4. Describe your value for documentation purposes
5. Click the "Add variable" button when finished

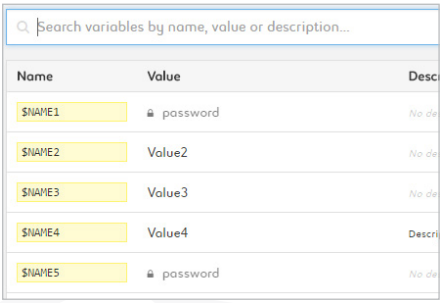
Delete a variable



1. Click on the "bin" icon next to a variable
2. Confirm your action by clicking on the "Remove" button

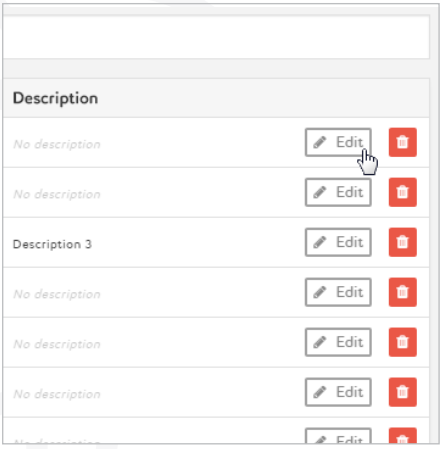
Environment variables 3/3

Search for a variable



1. Click on the search box located on top of the list of variables
2. Search variables by name, value or description

Edit a variable



1. Click on the "edit" button next to a variable
2. Edit the name, value and description fields
3. Save your edits by clicking on the "Save" button

Platform Manager API

API

Everything you can do through the UI is available through an API:

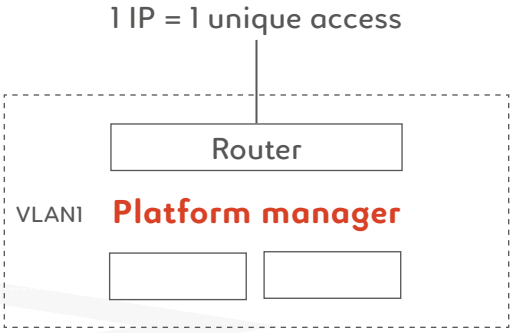
- You can automatize deployment with Jenkins
- You can gather job status in your centralized monitoring system

Documentation is available here:

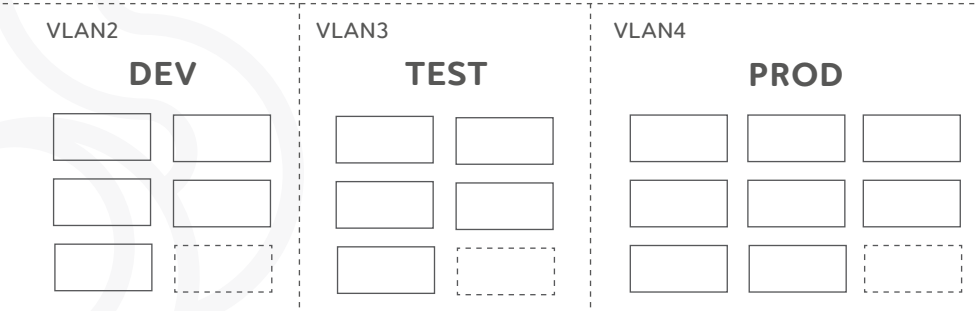
https://manager.prod.saagie.io/api/doc
https://manager.prod.saagie.io/api/doc/admin
(or with your private Cloud URL)

Network architecture

Managed version



VLAN1 can access to VLAN2, 3 & 4 but not in the other direction. VLAN2, 3 & 4 are silos



Report a bug

bug-platform@saagie.com

1. Send an email at that address explaining the bug you encountered
2. Please attach relevant files and screenshots