

# How to use Ambari Workflow views to schedule a Hive Query

## Workflow Manager (WFM) View Basics

Workflow Manager allows you to easily create, schedule, and monitor workflow jobs. The Apache Oozie workflow engine allows users to connect and automate the execution of big data processing tasks into a defined workflow. Workflow Manager integrates with the Hortonworks Data Platform (HDP) and supports Hadoop jobs for Hive, Sqoop, Pig, MapReduce, Spark, and more.

### 1. Log into the ambari view URL

Click on Workflow Manager Views (see Fig. 1).



Fig. 1

### 2. Creating Workflow

- Once in the WFM view, create a workflow (see Fig. 2).

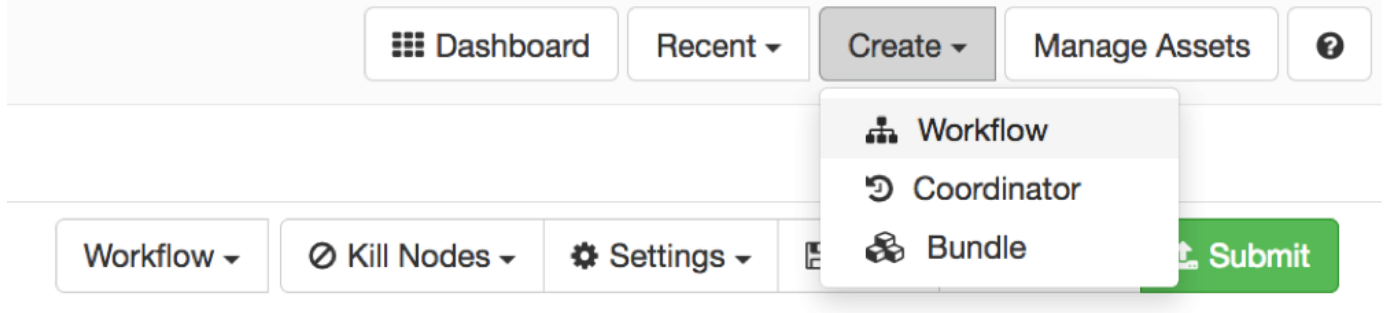


Fig. 2

- You can see in the picture below (Fig. 3) that any action in WFM has a beginning and end.

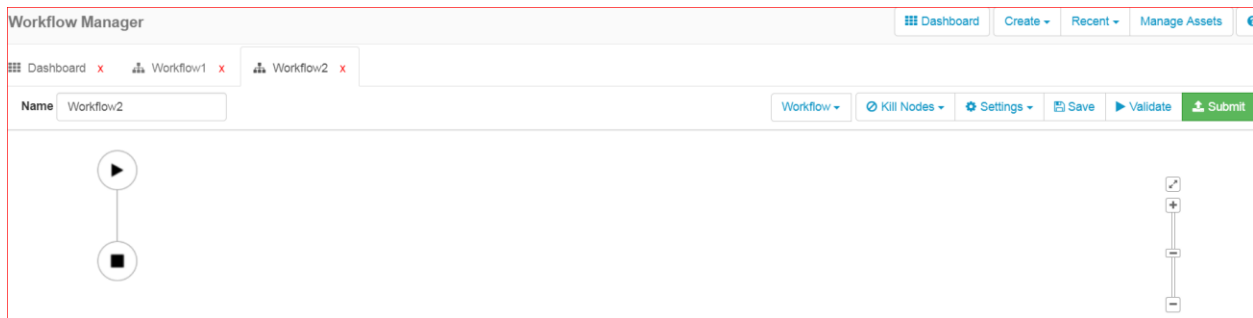


Fig. 3

- If you hover over the arrow connecting two nodes, you will notice a plus sign (Fig. 4) -- this is what you're going to use to add new actions.

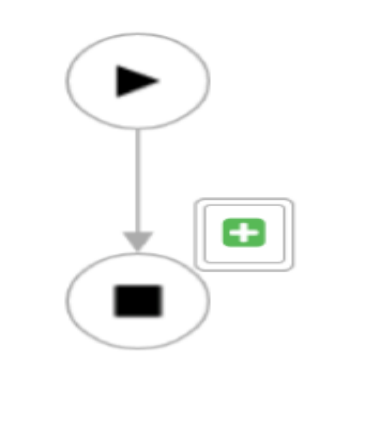


Fig. 4

- A list of all built-in actions will be visible once you click the plus sign (see Fig. 5).

### 3. Hive Action

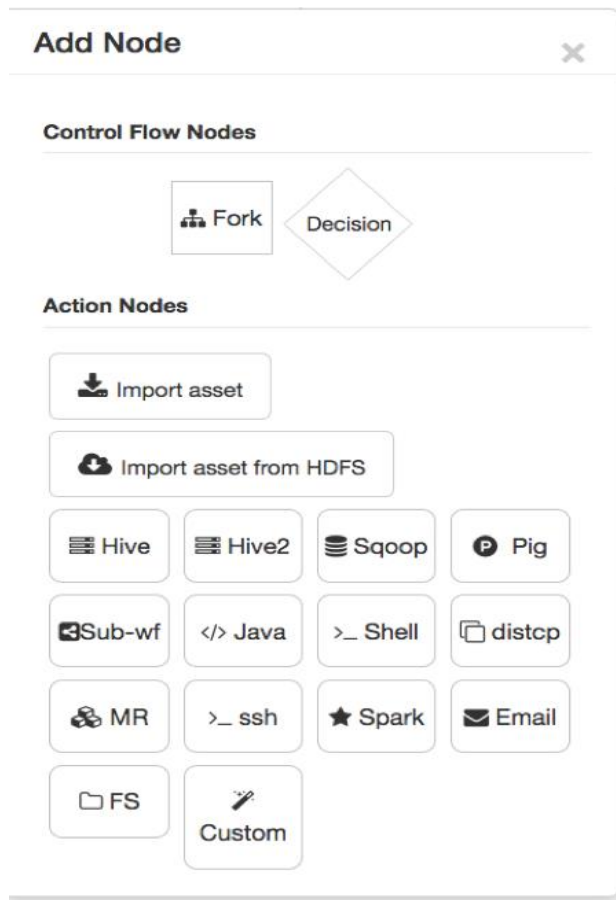


Fig. 5

- Select Hive action as that's what we're going to use in this scenario. Once clicked, you can edit the hive action node to rename it or configure it for your specific needs (See Fig. 6).



Fig. 6

## 4. Update Global settings

- Click Global settings to add Namenode HA details (see Fig. 7).

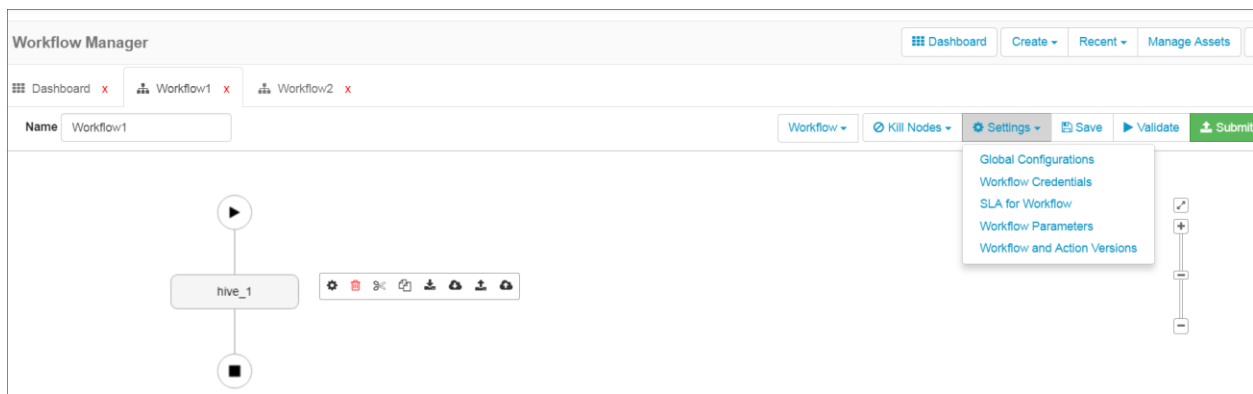
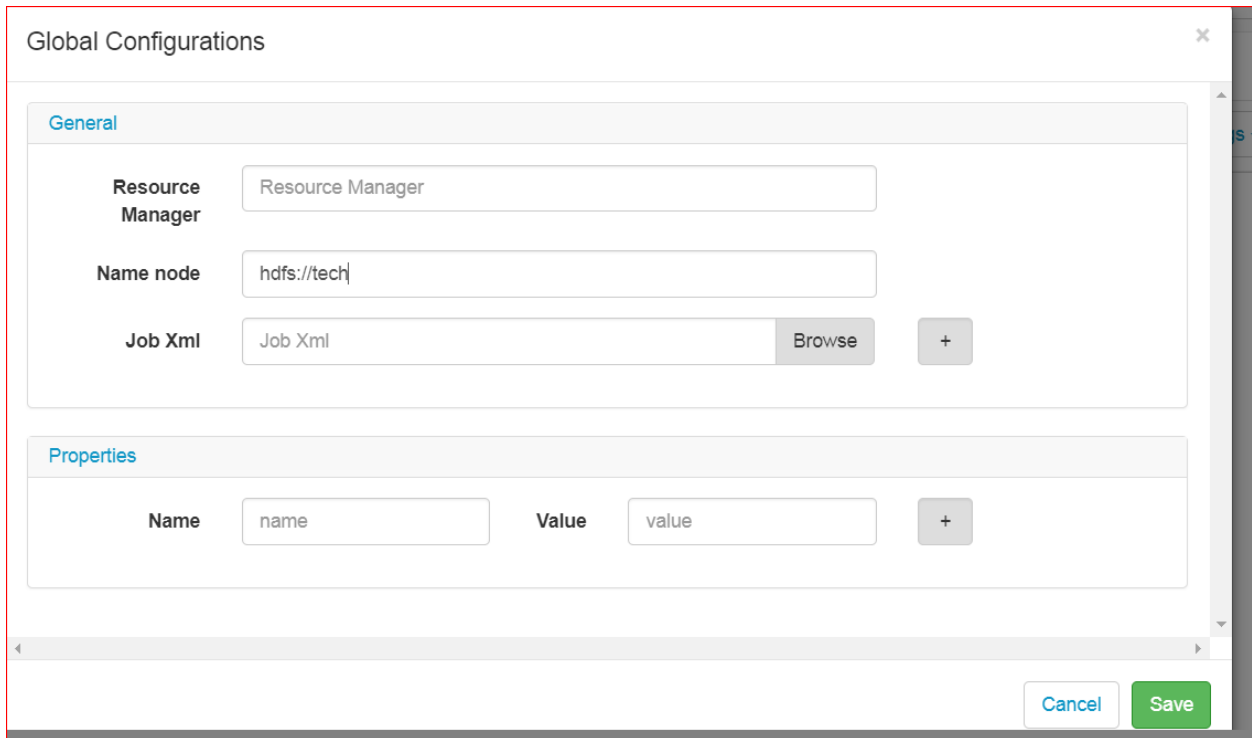


Fig. 7

- Next to "Name node" write: `hdfs://tech (clustername)` (See Fig. 8).



The image shows a 'Global Configurations' dialog box with a close button (X) in the top right corner. It is divided into two sections: 'General' and 'Properties'.  
The 'General' section contains three rows of configuration fields:

- 'Resource Manager' with a text input field containing 'Resource Manager'.
- 'Name node' with a text input field containing 'hdfs://tech'.
- 'Job Xml' with a text input field containing 'Job Xml', a 'Browse' button, and a '+' button.

The 'Properties' section contains a table with two columns: 'Name' and 'Value'.

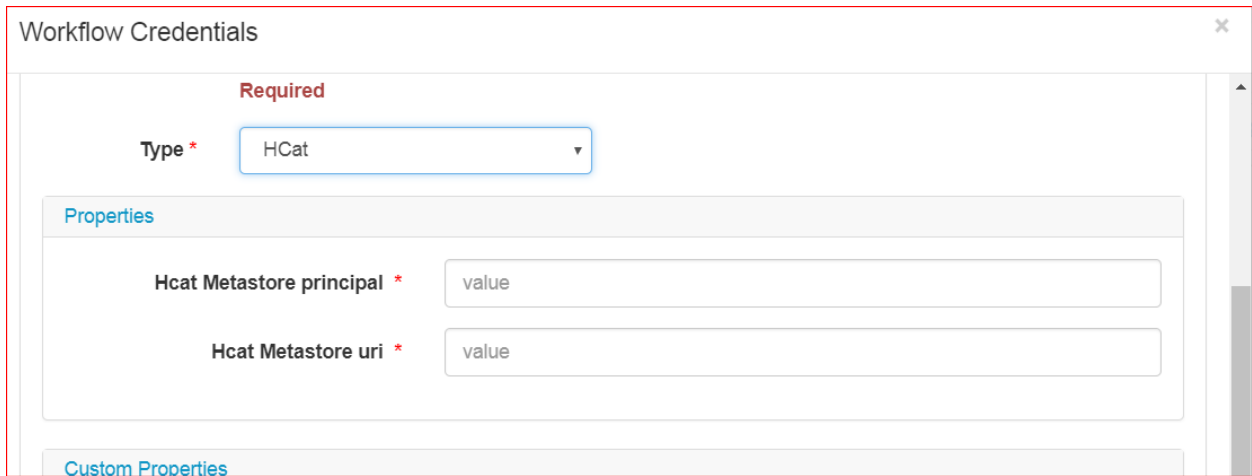
- The first row has 'name' in the 'Name' column and 'value' in the 'Value' column, followed by a '+' button.

At the bottom right of the dialog, there are 'Cancel' and 'Save' buttons.

Fig. 8

## 5. Update Workflow Credentials

- Update Workflow Credentials to **hcat** (see Fig. 9).



The image shows a 'Workflow Credentials' dialog box with a close button (X) in the top right corner. It features a 'Required' label in red above a 'Type \*' dropdown menu set to 'HCat'. Below this is a 'Properties' section with two rows of configuration fields:

- 'Hcat Metastore principal \*' with a text input field containing 'value'.
- 'Hcat Metastore uri \*' with a text input field containing 'value'.

At the bottom, there is a 'Custom Properties' section which is currently empty.

Fig. 9

Fig. 10

- For the following, fill in:
  - Name \* = hcat
  - Type \* = Hcat
  - Hcat Metastore principal \* : hive/\_HOST@TECH.HORTONWORKS.COM
  - Hcat Metastore Url \* : thrift://node.tech.hdp.hortonworks.com:9083
- Click Add and Save

## 6. Update Hive Query

- Click the gear icon to modify the configuration (see Fig. 11).

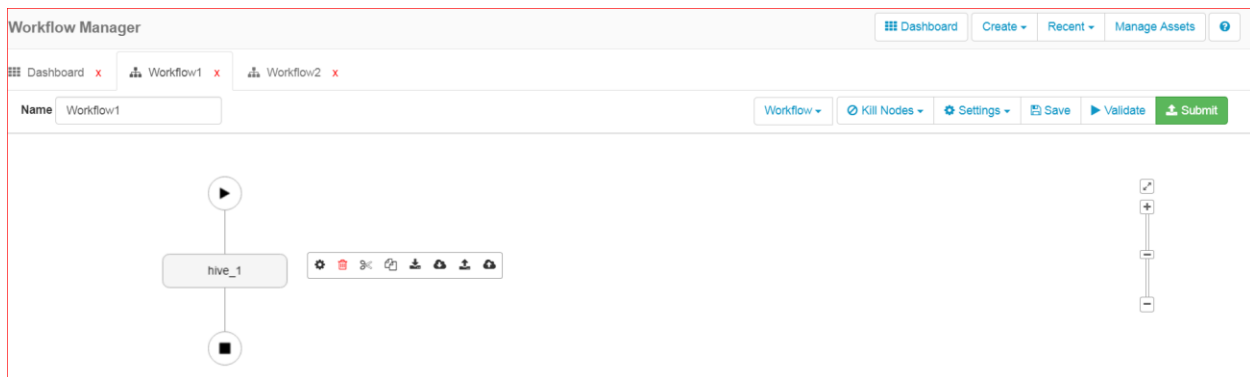


Fig. 11

- Next, write in a Query:
  - For example, Query \*: INSERT OVERWRITE DIRECTORY '/user/test/temp100' select \* from default.test\_100 limit 100;

**Note: Update any hive query here (See Fig. 12).**

(The above query run a simple hive command and capture its output as simple csv file. You can also enter name of a script file you want to execute. Then, you can scroll down into the advanced properties section to modify parameters and arguments to the command).

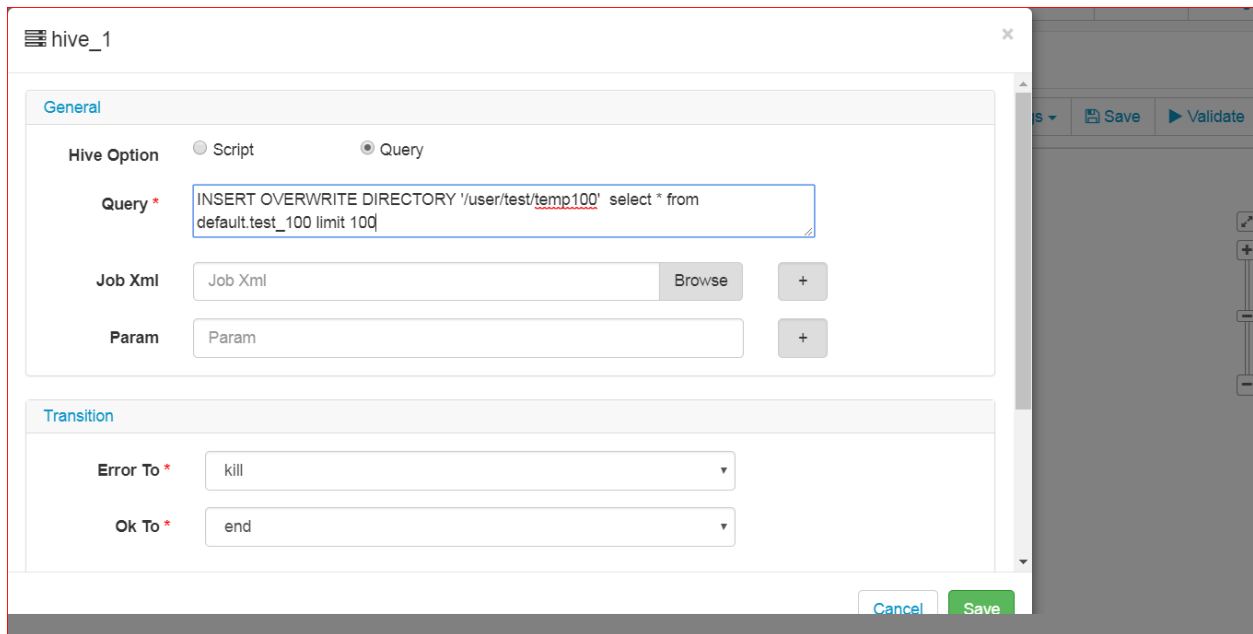


Fig. 12

## 7. Update the Credentials

- Update the credentials to Hcat (See Fig. 13).

Workflow Credentials

Edit - hcat (hcat)

Name \*

Type \*

Properties

Hcat Metastore principal \*

Hcat Metastore uri \*

Custom Properties

Name	Value	
<input type="text" value="name"/>	<input type="text" value="value"/>	<input type="button" value="+"/>

Fig. 13

## 8. Update Other properties

- Click Advanced Properties and write in (See Fig. 14):
  - ResourceManager: `${resourceManager}`
  - Name node: `hdfs://tech` (Note: for Tech cluster)
- Click Save



Credentials

Advanced Properties

Resource Manager:

Name node:

File:

Archive:

Prepare:

arg:

Fig. 14

- At this point my workflow is ready to submit. Click on that next and you will see a few more options (see Fig. 15).
- You need to specify the location of the workflow – the path must exist on hdfs and not xml.
  - If xml exists, you need to overwrite it. Then you can select run on submit.
  - Another option in the dialog is to validate your workflow. Clicking that will tell you whether your workflow is valid or not.

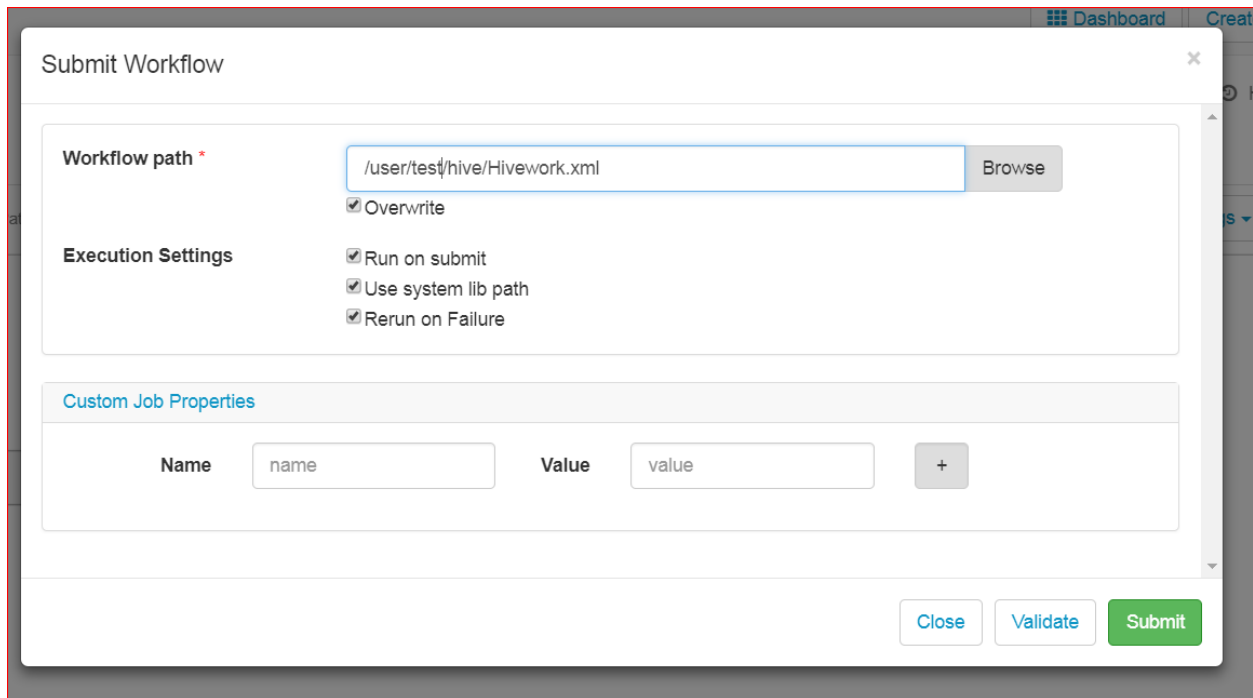


Fig. 15

- Write in the following for Workflow Path: /user/t93kq9j/hive1.xml (Note: use your home directory and name to save the workflow)
- Click Submit (See Fig. 16).

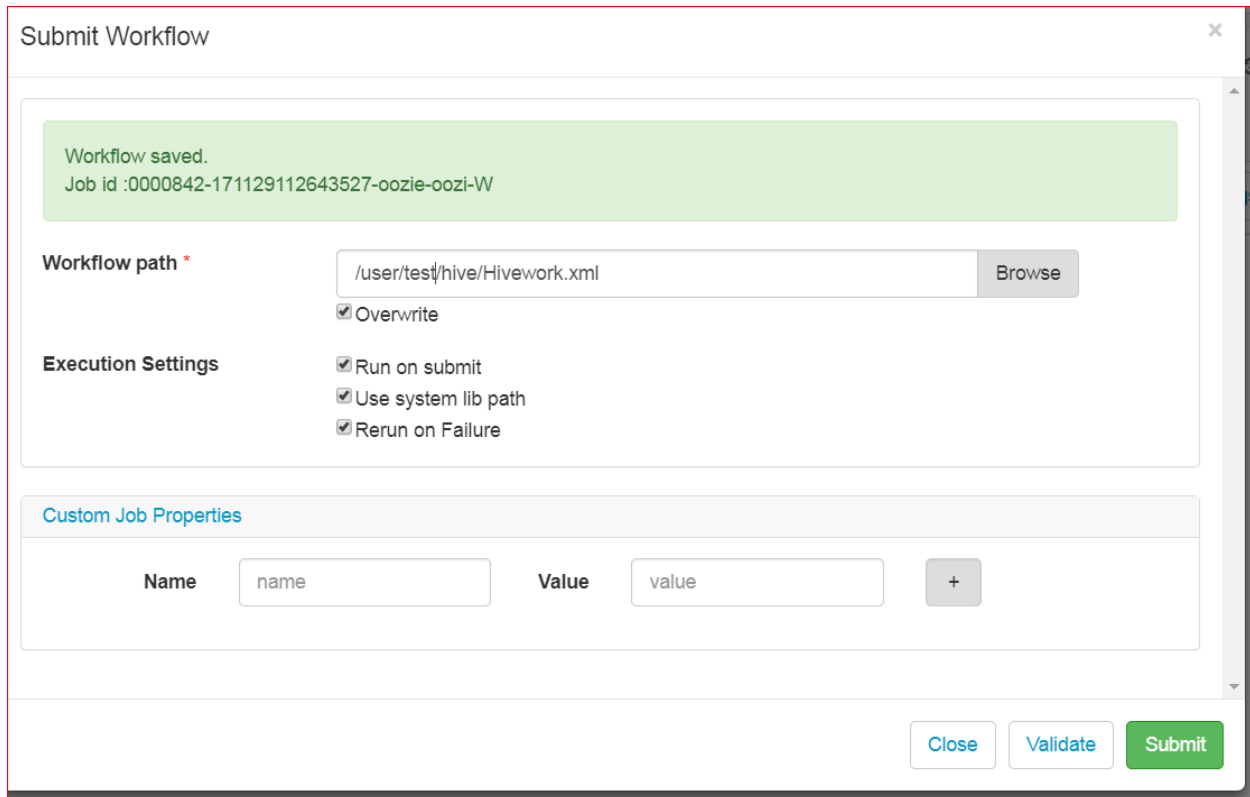


Fig. 16

## 9. Review status

- Click the dashboard tab to see the status of your workflow.
- You can now monitor your job status in this UI! Once job completes, you can navigate to the output and view the result of the workflow.

## 10. Schedule a Job to run in specified window (Hourly, Daily, Weekly, etc...)

1. Before scheduling a job, you must complete steps 1-9 as a pre-requisite.
2. Click the Coordinator under Create menu and submit.