

Create Draft Flow

Date published: 2021-04-06

Date modified: 2024-01-09

CLOUDERA

Legal Notice

© Cloudera Inc. 2024. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 (“ASLv2”), the Affero General Public License version 3 (AGPLv3), or other license terms. Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER’S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

Contents

Creating a new draft.....	4
Creating a draft from a flow definition in the Catalog.....	5
Creating a draft from a ReadyFlow.....	6
Adding a parameter to your draft.....	7
Adding a parameter from Flow Options.....	7
Creating a parameter from the Configuration pane.....	9
Configuring advanced usage of the UpdateAttribute processor.....	10
Add a Rule.....	12
Reorder Rules.....	13
Testing JoltTrasformJSON processor configuration.....	13
Adding a service to your draft.....	15
Adding a service from Flow Options.....	15
Creating a service from Configuration.....	15
Adding inbound connection support to a draft.....	17
Copying and pasting processors.....	18

Creating a new draft


Learn about creating a draft in CDF Flow Designer.

Before you begin

- Make sure that you have DFDeveloper permission to perform this task. For information on account and resource roles, see DataFlow Authorization.

Procedure

1. Open Cloudera DataFlow by selecting the DataFlow tile in the CDP sidebar.

2. Select  Flow Design in the left navigation pane.

You are redirected to the Flow Design page, where previously created drafts are displayed, one per row.

3. Select  Create Draft.

4. Select the Target Workspace where you want to create the draft.

5. Select the Target Project to which you want to assign the draft.

6. Provide a valid Draft Name.

Draft names must be unique within a workspace. If a draft with the provided name already exists, you need to provide a different name.

7. Click Create.



Flow Designer creates a draft and a default Process Group with the Draft Name you provided and you are redirected to the **Flow Design** canvas. The **Configuration** pane on the right displays configuration options for the default Process Group.

8. Start building your draft by dragging components to the canvas from the Components sidebar on the left.

The **Configuration** sidebar always displays configuration options for the highlighted component. To select a component for configuration, click on it.

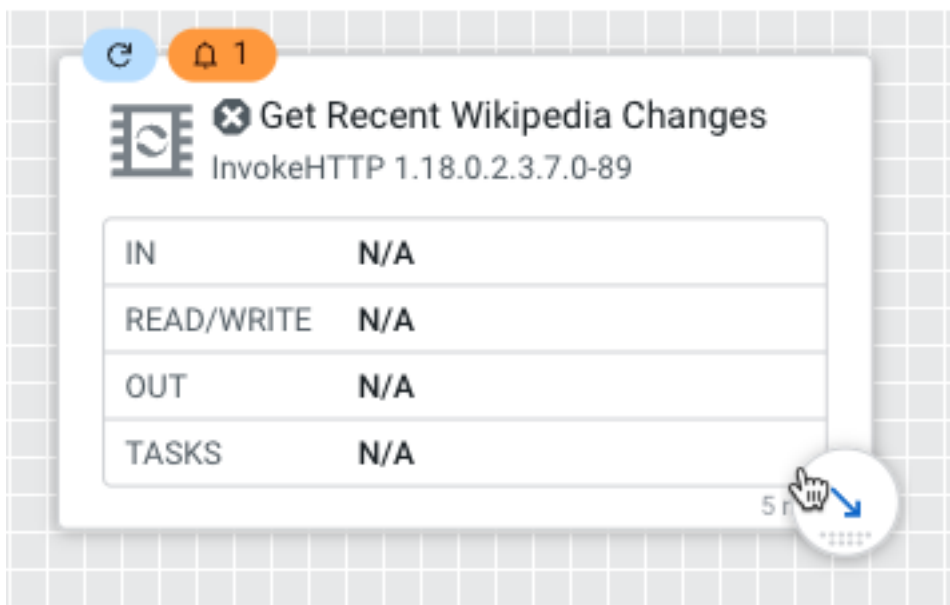


Tip:

- To create a Controller Service, click Flow Options  Services Add Service .
- To create a Parameter, click Flow Options  Parameters Add Parameter .

Unlike in NiFi, there is one default parameter group auto-created for each draft, where you can add parameters. This parameter group is then automatically bound to each Process Group you create within your draft.

9. You can connect two processors by hovering the cursor over the lower-right corner of the source processor, clicking the arrow that appears and dragging it to the target processor.



10. To test your draft, start a Test Session by clicking **Flow Options Test Session Start Start Test Session**.

This provisions a NiFi sandbox and allows you to validate your draft and work with live data by starting and stopping components.



Tip: You can check the status of your Test Session in the upper right corner of your workspace. That is where you can also deactivate the Test Session.

11. Once you are confident that your flow design is ready for production, click **Flow Options Publish To Catalog Publish**.

This step turns your draft to a flow definition and adds it to the Catalog. From there, you can deploy your flow to any environment where you have DFFlowAdmin authorization.

Related Tasks

[Adding a parameter to your draft](#)

[Adding a service to your draft](#)

Related Information

[DataFlow authorization](#)

Creating a draft from a flow definition in the Catalog

You can open flow definitions in the Catalog as drafts in Flow Designer. Depending on whether the flow definition has associated flow drafts or not, you may have more than one option.

Before you begin


- You have an enabled and healthy CDF environment.
- You have been assigned the DFDeveloper role granting you access to the Flow Designer.
- You have been assigned the DFCatalogAdmin or DFCatalogViewer role granting you access to the Catalog. You will need this authorization to view and open flow definitions, and to publish your draft as a flow definition to the Catalog.
- You have been assigned the DFFlowAdmin role for the environment to which you want to deploy the flow definition.

About this task

If you have sufficient rights, you can create a new draft from existing flow definitions in the Catalog. This gives you two options:

- You can update the existing flow definition by creating a draft and then publishing your work as a new version of the source flow definition.
- You can create a draft using the source flow definition as a template and then publish your work as a new flow definition to the Catalog.

Procedure

1. Open Cloudera DataFlow by clicking the DataFlow tile in the CDP sidebar.
2. Select  Catalog in the left navigation pane.
3. Select the flow that you want to open as a draft.
4. Depending on whether there already is one or more associated draft for the selected flow, you may have two options to select from:
 - If there is one or more ASSOCIATED DRAFT, you can select that and the draft opens for editing in the Flow Design view.
 - If there are no associated drafts or you want to create a new flow based on the existing one, select Create New Draft.



Note: If there is more than one version available, select the version that you want to use as a template for your draft.

Proceed according to *Creating a draft*.

Related Tasks

[Creating a new draft](#)

Creating a draft from a ReadyFlow

You can use ReadyFlows as the starting point of your flow design. Select an available ReadyFlow from the ReadyFlow Gallery, open it for editing on the Canvas, and publish it to the Catalog as your own custom flow definition.


Before you begin

- You have an enabled and healthy CDF environment.
- You have been assigned the DFDeveloper role granting you access to the Flow Designer.
- You have been assigned the DFCatalogAdmin or DFCatalogViewer role granting you access to the Catalog. You will need this authorization to publish your draft as a flow definition to the Catalog.
- You have been assigned the DFFlowAdmin role for the environment to which you want to deploy the flow definition.

About this task

ReadyFlows are read-only, therefore you cannot add a new version to them. When you are done with creating your customized version of the given ReadyFlow, you can only publish your work to the Catalog as a new flow definition.

Procedure

1. Open Cloudera DataFlow by selecting the DataFlow tile in the CDP sidebar.
2. Select  ReadyFlow Gallery in the left navigation pane.

3. Select the ReadyFlow you want to use as a template for your flow design.
4. Click Create New Draft.
5. Select the Target Workspace where you want to create the draft.
6. Select the Target Project to which you want to assign the draft.
7. Provide a valid Draft Name for the draft.

Flow names must be unique within a workspace. If a draft with the provided name already exists, you need to provide a different name.

8. Click Create.

The selected ReadyFlow opens as a draft in the designated Flow Designer workspace with the Draft Name you provided, on the **Flow Design** canvas.

Adding a parameter to your draft

The values of properties in the flow, including sensitive properties, can be parameterized. Learn how to add a parameter to your draft.

About this task



Note:

Parameters are created within Parameter Contexts. In CDP Flow Designer, one default Parameter Context is auto-created when you create a new draft. You can then add parameters to this one context, you cannot create additional ones. This default context is automatically bound to each Process Group you create within your draft, making all parameters available to be used in any process group.

Before you begin

- Make sure that you have DFDeveloper permission to perform this task. For information on account and resource roles, see DataFlow authorization.
- You have created a draft in CDF Flow Designer.

Related Information

[DataFlow authorization](#)

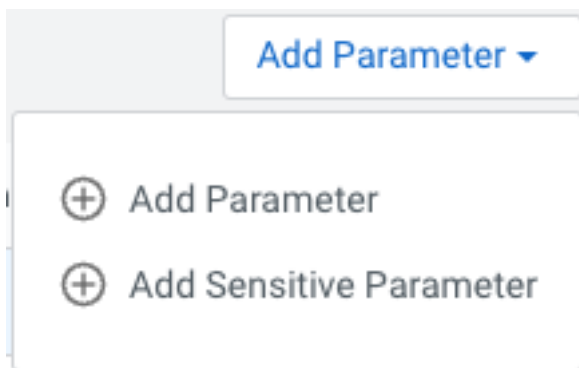
[For more information on Parameters, see the NiFi User Guide.](#)

Adding a parameter from Flow Options

Learn about creating parameters from the Flow Options menu.

Procedure

1. Click Flow Options Parameters Add Parameter Add Parameter or Add Sensitive Parameter.



Note: A sensitive parameter cannot be later changed to non-sensitive and this is also true the other way around. Keep this in mind when creating parameters.

2. Provide the parameter Name, define a Value (or select Set empty string) and optionally provide a parameter Description.

Add Parameter

Name

Create a unique name

Value

Enter parameter value

☐ Set empty string

Description

Add description

Cancel

Apply

3. When you are done, click Apply.

4. For your parameter changes to take effect, click **Apply Changes**.

Flow Design / cdf-priv / cdev-test / Parameters

To fully validate your flow and develop with live data, start a test session.

You must first apply or discard the changes that have been made.

Back To Flow Designer

Add Parameter

Name ↑	Value	Changed
another_para	Empty string set	<div>New</div>
generic_para	42	
sensitive_para	<div>Sensitive value set</div>	

Apply Changes

Discard Changes

Flow Options

»

another_para

Not Sensitive

A test session must be running in order to add files.

Value

Set empty string

Description

Description

Referencing Components

No referencing Processors to display.

No referencing Services to display.

What to do next

Click Back to Flow Designer to return to the canvas when you are done adding parameters.

Creating a parameter from the Configuration pane

Learn about converting a property of a component into a parameter from the Configuration pane.

Procedure

1. Select the component with the Property that you want to convert to a parameter.
Click the component on the Canvas to access its properties in the Configuration pane.

2. Locate the property that you want to turn into a parameter and click  More  More Convert To Parameter .

Properties


Property	Value
Record Reader 	 JSON_Reader_Recent_Changes 
Record Writer 	 AvroWriter_Recent_Changes 
Include Zero Record FlowFiles 	true 

Convert To Parameter




Relationships

3. Edit the parameter Name, Value, and optionally add a description.
4. Click Add.
The Value of the Property is automatically updated to a reference to the newly created parameter.
5. Click Apply to save the change you made to your component.



Note: Even if you click  Discard Changes at this stage, the parameter is created, but the Property is reset to its previous Value.

To delete an unwanted parameter:

- a. Go to Flow Options  Parameters .
- b. Click  in the row of the parameter to be deleted.
- c. Click  Delete.

Configuring advanced usage of the UpdateAttribute processor

The UpdateAttribute processor may be used to make conditional changes, based on user defined rules.

About this task

Rules are a set of Conditions and associated Actions. Each rule basically says, "If these conditions are met, then do this action." One or more conditions may be used in a given rule, and they all must be met in order for the designated action(s) to be taken.

In this topic we use two rules as an example to demonstrate how adding rules work: CheckForLargeFiles and CheckForGiantFiles.

The CheckForLargeFiles rule has the following conditions:

- `${filename:equals('fileOfInterest')}`
- `${fileSize:toNumber():ge(1048576)}`
- `${fileSize:toNumber():lt(1073741824)}`


The associated action for the filename attribute:

- `${filename}.meg`

Taken together, this rule says:

- If the value of the filename attribute is fileOfInterest, and
- If the fileSize is greater than or equal to (ge) one megabyte (1,048,576 bytes), and
- If the fileSize is less than (lt) one gigabyte (1,073,741,824 bytes)
- Then change the value of the filename attribute by appending ".meg" to the filename.

Continuing with this example, we can add another rule to check for files that are larger than one gigabyte. When

we add this second rule, we can use the previous rule as a template, so to speak, by selecting the  Expand



Duplicate option for the CheckForLargeFiles rule.

In this example, the CheckForGiantFiles rule has these conditions:

- `${filename>equals('fileOfInterest')}`
- `${fileSize:toNumber():gt(1073741824)}`

Then it has this action for the filename attribute:

- `${filename}.gig`

Taken together, this rule says:

- If the value of the filename attribute is fileOfInterest, and
- If the fileSize is greater than (gt) one gigabyte (1,073,741,824 bytes)
- Then change the value of the filename attribute by appending ".gig" to the filename.


Procedure

1. To add rules to the UpdateAttribute processor and to define behavior when multiple rules match, go to

« Configuration Pane  Define Rules .










»





UpdateAttribute
 UpdateAttribute 1.21.0.2.3.9.0-24

[More Details](#) ▾

 Configuration

 Define Rules

Rules

Define the behavior when multiple rules match. [Learn more](#) 


*FlowFile Policy

Use Clone ▾

 Search by name, expression, attribute or value

 Add Rule

*Rule Order List

 No rules to display.

2. Select a Flow File Policy:

Use Clone


Matching rules are executed with a copy of the original flow file. This is the default behavior.

Use Original

Matching rules are executed with the original flow file, in the order that you specify under Rule Order List.

Add a Rule

Procedure

1. You can add rules by clicking Add Rule.
2. Specify a Rule Name.
Specify CheckForLargeFiles
3. Add Conditions by clicking  Add Condition.




Note: All conditions you add here must be met for the rule to match.

In the **Add Condition** modal window you can specify a condition using regular expressions.

Add the `${filename:equals('fileOfInterest')}`

Click Add Condition in the modal window to save your changes.

4. To add Actions, click  Add Action.
In the **Add Action** modal window specify an Attribute name and a Value.
For example, specify filename as Attribute and `${filename}.meg` as Value.
Click Add Action to save your changes.
5. Once you are done with creating or updating a Rule, click Apply to save your changes.



Tip: As long as your rule configuration is invalid, the Apply button is inactive and you can only



Discard Changes.

Reorder Rules**Procedure**

1. To reorder Rules, grab one by the  Drag Handle and drag it to the required position.
2. Click Apply.

Testing JoltTransformJSON processor configuration

You can test the transformations a given Jolt Specification effects on a sample JSON file without starting a test session.

Procedure

1. To test Jolt Transform when using the JoltTransformJSON processor, go to  Configure  Test Jolt Transform .

- You may modify the Jolt Transformation DSL and Jolt Spec fields. If you modify them here, click Apply to confirm the changes.

The values for fields Jolt Transformation DSL and Jolt Spec are automatically retrieved from the processor



Configuration if already set.



Important: Any configuration changes you apply here overwrite the processor configuration.

Configuration

[Test Jolt Transform](#)

Jolt Transformation DSL

Chain

Attributes (0)

Jolt Spec



```
1 [
2   {
3     "operation": "shift",
4     "spec": {
5       "breadbox": "counterTop"
6     }
7   }
8 ]
```

Json Input



1

Output

Test Transform



1

Apply

Discard Changes

3. You may add attributes and values to reference within your specification for testing input. To do so, click

^ Attributes.



Important: Attribute values entered here are only available during testing, but not when executing the actual flow.

4. Paste the input file contents into the Json Input box.
5. Click Test Transform.
The result of the transformation is printed to the Output box.

Adding a service to your draft

Learn how you can add a service in CDF Flow Designer.

Before you begin

- Make sure that you have DFDeveloper permission to perform this task. For information on account and resource roles, see [DataFlow authorization](#).
- You have created a draft flow in CDF Flow Designer.

Related Information

[DataFlow authorization](#)

Adding a service from Flow Options

Learn about creating services from the Flow Options menu.

Procedure

1. To create a service for your flow, click Flow Options Services Add Service .
2. In the Add Service box you can select from the available services by scrolling down the list on the left or by starting to type the name of the service you are looking for.
You can optionally change the name of the service.
3. Click Add.
The Configuration pane opens.
4. Make the required settings to your service, then click Apply.

What to do next


Click Back to Flow Designer to return to the canvas when you are done adding services.

Creating a service from Configuration



Learn how you can create services directly from the Configuration pane for a property that expects a controller service as a value.

Procedure

1. Select the component with the Property for which you want to create the controller service.
Click the component on the Canvas to access its properties on the Configuration pane.

2. Locate the property for which you want to create the service and click  More Add Service .

Properties

Property	Value
Record Reader 	No value set
Record Writer 	No value set
Include Zero Record FlowFiles 	true



3. In the Add Service box you can select from the available services by scrolling down the list on the left or by starting to type the name of the service you are looking for.

You can optionally change the name of the service.

4. Click Add.

The Configuration pane opens.

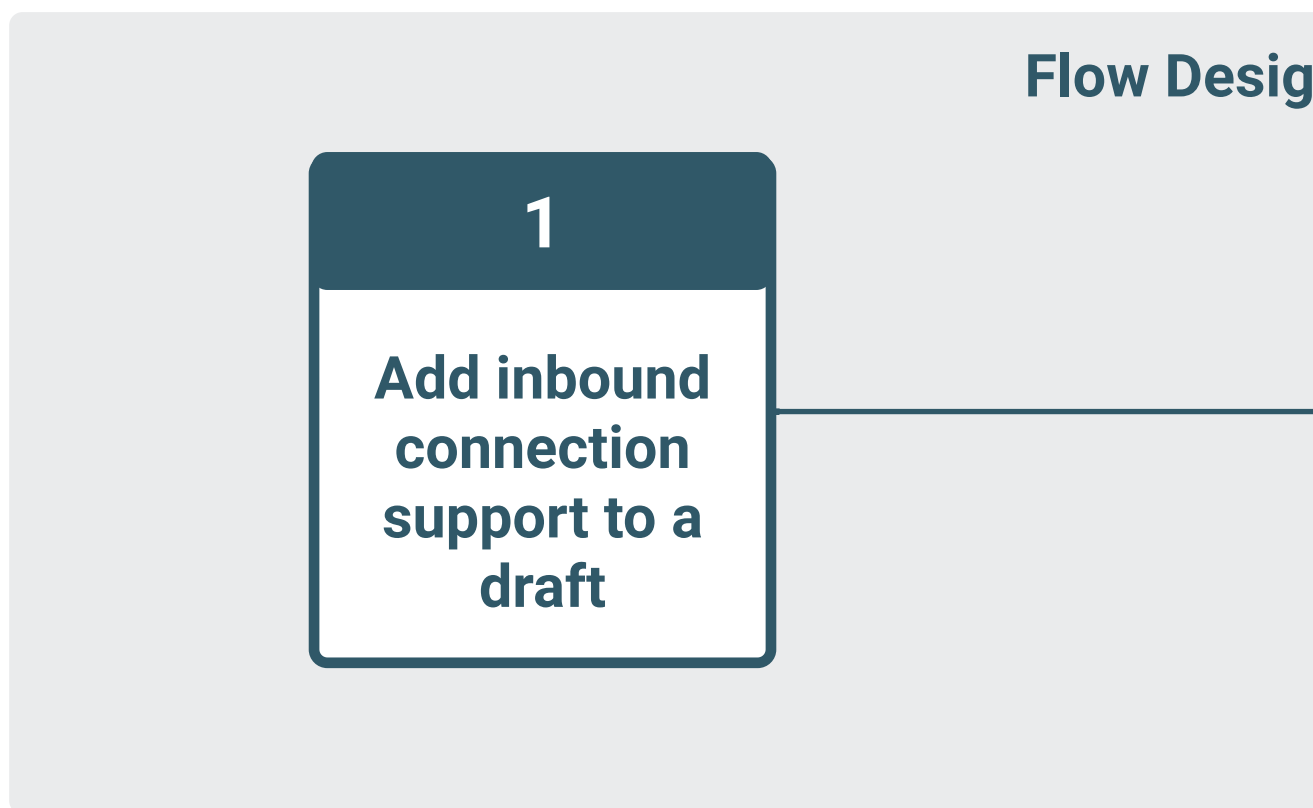
5. Make the required settings to your service, then click Apply.

What to do next

Click Back to Flow Designer to return to the canvas when you are done adding services.

Adding inbound connection support to a draft

You can enable a dataflow to listen on a public endpoint to data sources that are outside your CDP environment by adding inbound connection support to your



draft.

To enable your dataflow to use listen processors with inbound connection endpoint support, make the following addition to your draft:


1. Open the draft where you want to enable inbound connection support.
2. Add the required listen processor to your draft.

CDF supports all listen processors, including custom processors.


3. Configure the processor to enable inbound connections.

Port

Provide a port number where the flow deployment listens for incoming data. You can either add it

as a parameter by clicking  **Convert To Parameter** to make the actual port number configurable during flow deployment, or you can set an explicit port number when creating the draft.

SSL Context Service

Create an external `StandardRestrictedSSLContextService` for your processor, by clicking  **Add Service** and filtering for `StandardRestrictedSSLContextService`. You must name this context service `Inbound SSL Context Service`. No other configuration is required. The SSL context service

will be created during cluster deployment and all other properties will be populated with values generated for that NiFi cluster.

Client Auth

Set to “REQUIRED” to use mTLS.



Note: Most listen processors only support TCP. ListenSyslog supports both TCP and UDP. As dataflow deployments do not support mixed protocols for listening ports, and UDP does not support SSL Context Service and Client authentication, Cloudera advises to configure your ListenSyslog processor to use TCP protocol. This allows your CDF dataflow to use SSL Context Service for authentication and to listen to different data sources on different ports.

If this is not possible, create separate flows that listen on UDP and TCP respectively.

Copying and pasting processors

In Flow Designer you can copy processors together with their configuration and then paste them to the same or another draft, or to an external application.

About this task

Copying a processor copies the entire processor configuration to the clipboard, even sensitive properties are kept. If you paste the processor to a different draft, you may still get an error as, for example, services are not copied together with your processor and you will have to recreate those in the new draft.

Procedure

1. Select the processor you want to copy then either right-click on it and select Copy to clipboard or press **#+C**.
2. To paste the processor, press **#+V**.



Tip: You can also paste processor configurations in JSON format to external applications, such as text editors and easily share them for troubleshooting purposes.