



# Pipelines in Azure Data Factory

This post is part 5 of 25 in the series [Beginner's Guide to Azure Data Factory](#)



In the previous post, we used the [Copy Data Wizard](#) to copy a file from our demo dataset to our storage account. The Copy Data Wizard created all the factory resources for us: pipelines, activities, datasets, and linked services.

In this post, we will go through pipelines in more detail. How do we create and organize them? What are their main properties? Can we edit them without using the graphical user interface?

## Pipelines: The Basics

When I was new to Azure Data Factory, I had *many* questions, but I didn't always have someone to ask and learn from. When I did work in a team, I didn't always dare ask my team members for help, because I felt silly for asking about things that I felt I should probably know.

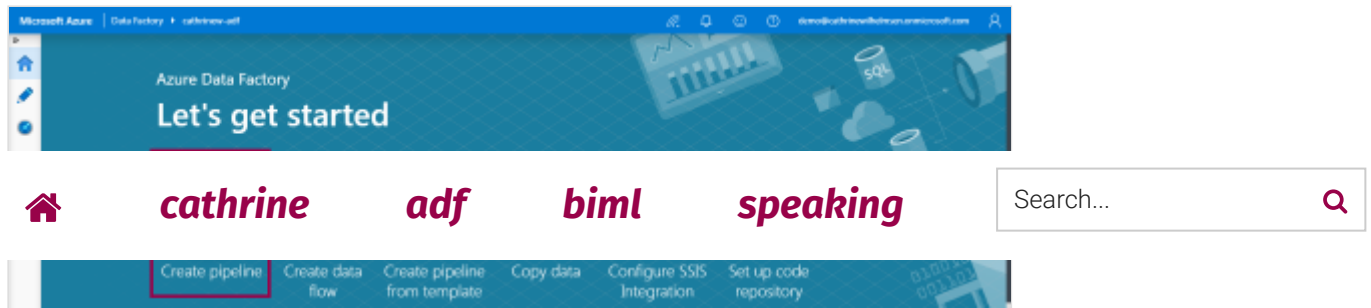
Yeah, I know... It's easy to tell *others* that there are no silly questions, but I don't always listen to myself :)

I don't want you to feel the same way! So. Let's start from the beginning. These are the questions that I had when I was new to Azure Data Factory. Or, these are the questions that I realized I should have asked when I discovered something by accident and went "*Oh! So that's what that is! I wish I knew that last week!*"

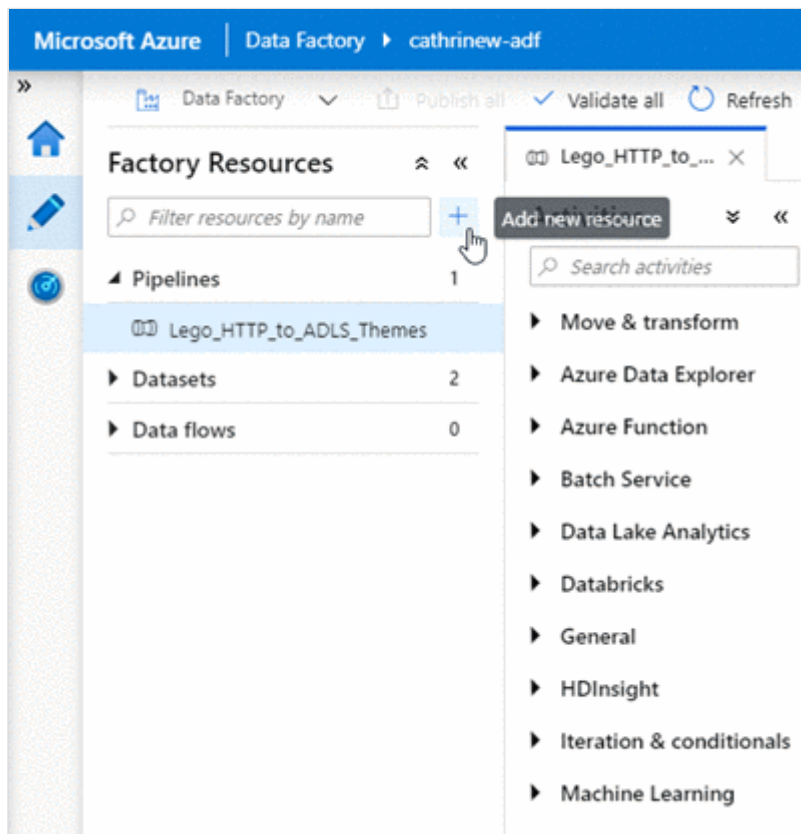
### How do I create pipelines?

So far, we have created a pipeline by using the Copy Data Wizard. There are several other ways to create a pipeline. Some of these were not immediately obvious to me :)

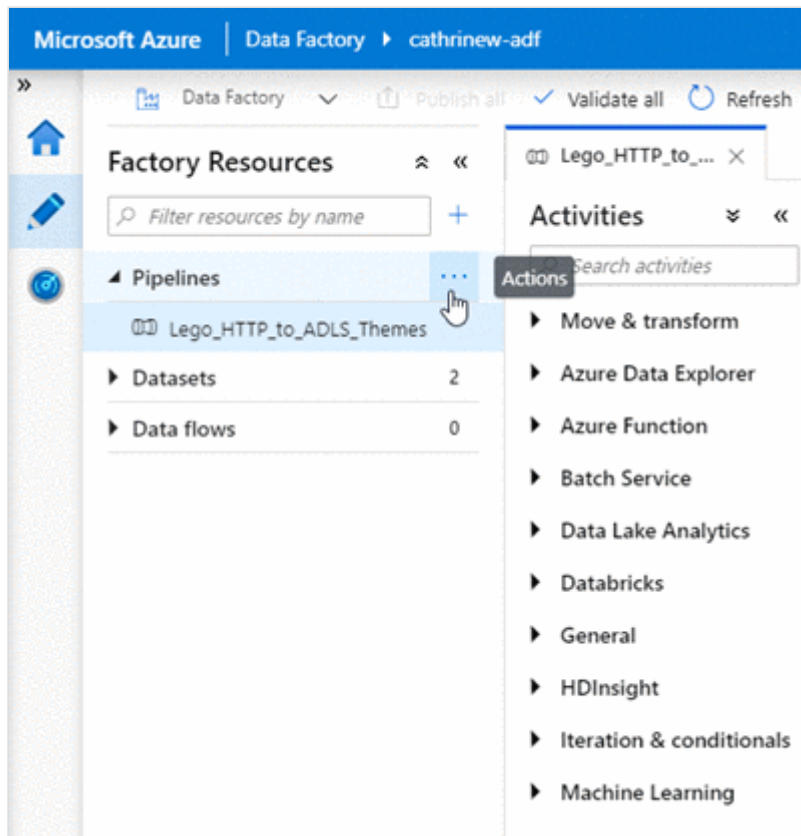
On the *Home* page, click on **create pipeline**:



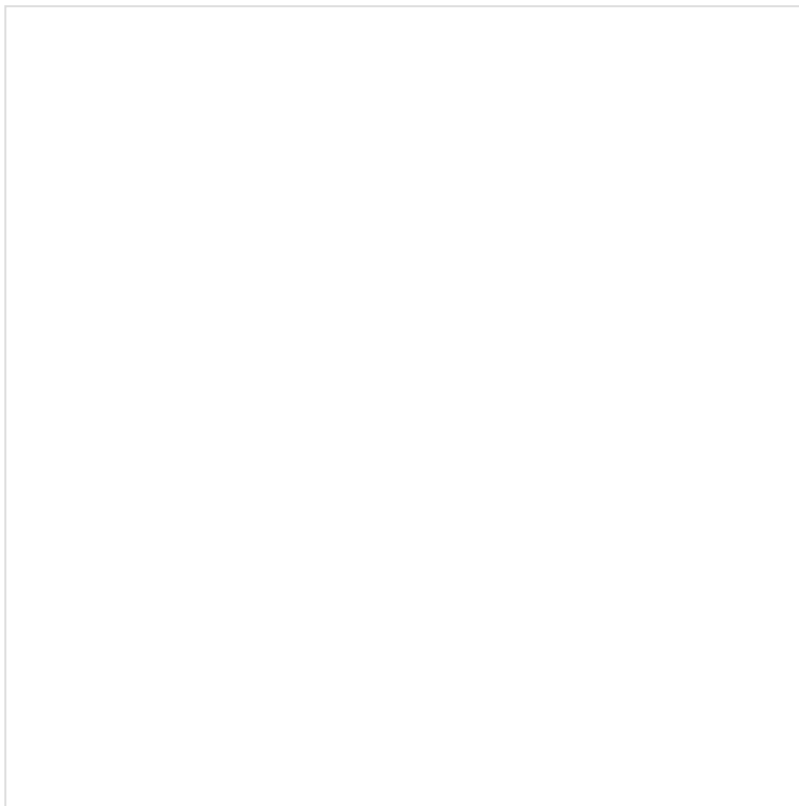
On the *Author* page, click **+** (add new resource) under factory resources and then click **pipeline**:



Hover over the number next to the pipelines group. The number will change into the actions ellipsis (...). Click the ellipsis, then click **new pipeline**:



If you already have a pipeline, hover over the name to show the actions ellipsis (...). Click the ellipsis, then click **clone**:

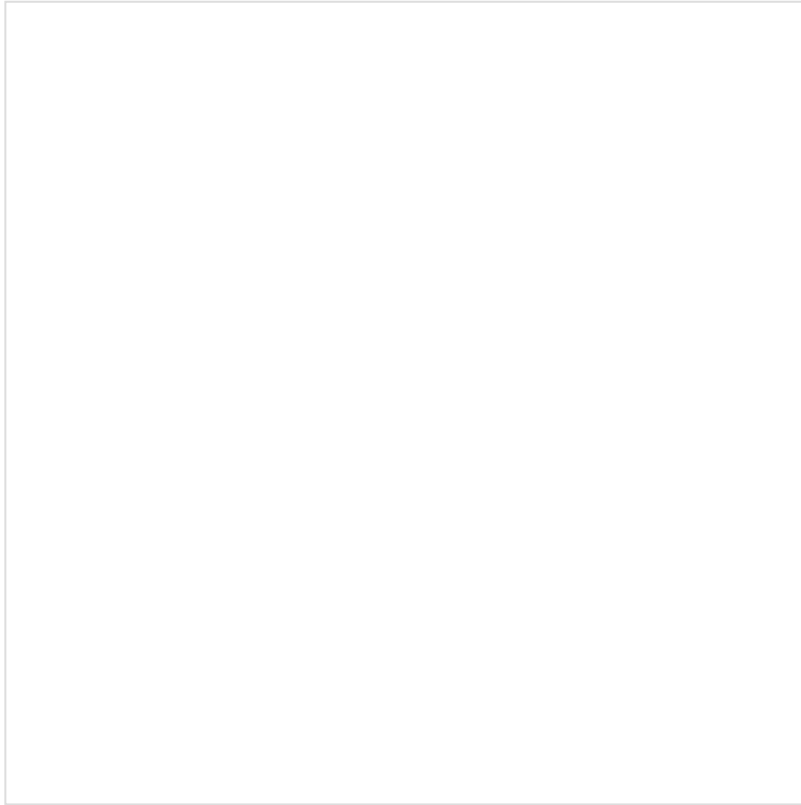


*(These actions also work for datasets and data flows.)*

## ***How do I organize pipelines?***

Pipelines are sorted by name, so I recommend that you decide on a naming convention early in your project. And yeah, I keep saying this to everyone else, but then I can never decide on how to name my own pipelines, haha :) Don't worry if you end up renaming your pipelines several times while you work on your project. It happens, and that's completely fine, but try to stick to *some* kind of naming convention throughout your project.

In addition to naming conventions, you can create folders to organize your pipelines. Click the actions ellipsis next to the pipelines group, then click **new folder**. Folders can be nested:

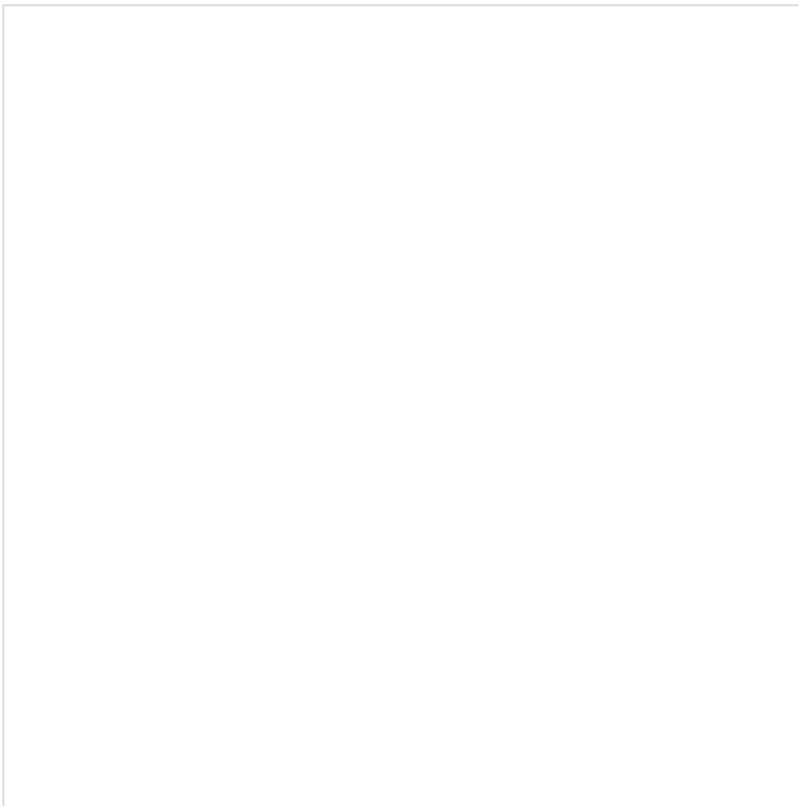


(You can create *at least* 8 levels of folders, apparently! Do you want a challenge? Figure out how many levels are supported, and if they are limited by the length of the level names. I'm curious, but I only had time to try 8 levels for these screenshots 😊)

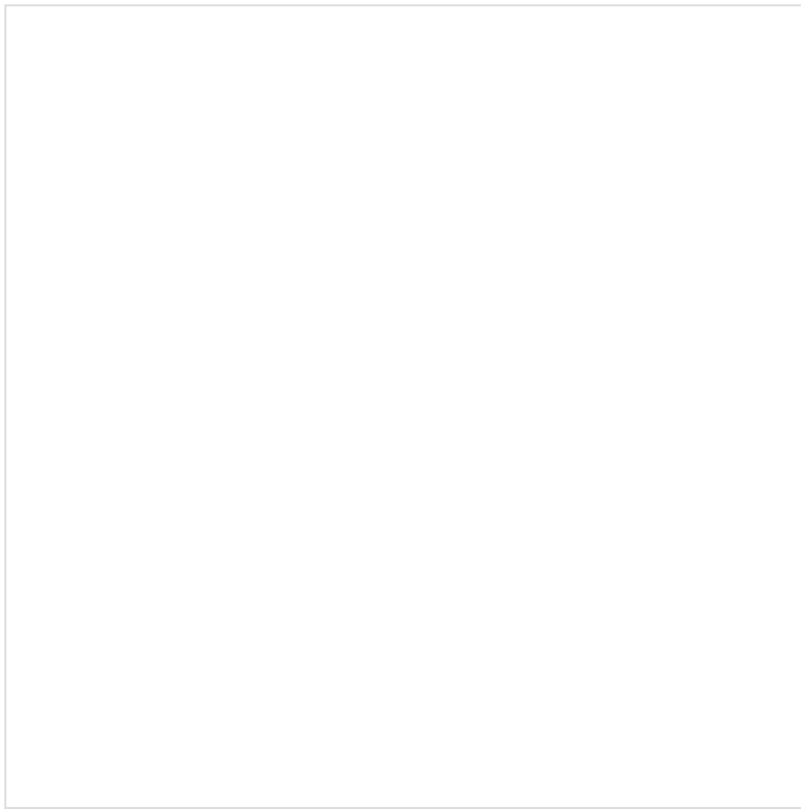
You can also hover over the number next to the folder. The number will change into the actions ellipsis (...). From here, you can rename and delete the folder, or create a sub-folder:



It is currently not possible to create a pipeline directly in a folder. You have to first create the pipeline, and then move it into a folder. You can either move it by dragging and dropping:



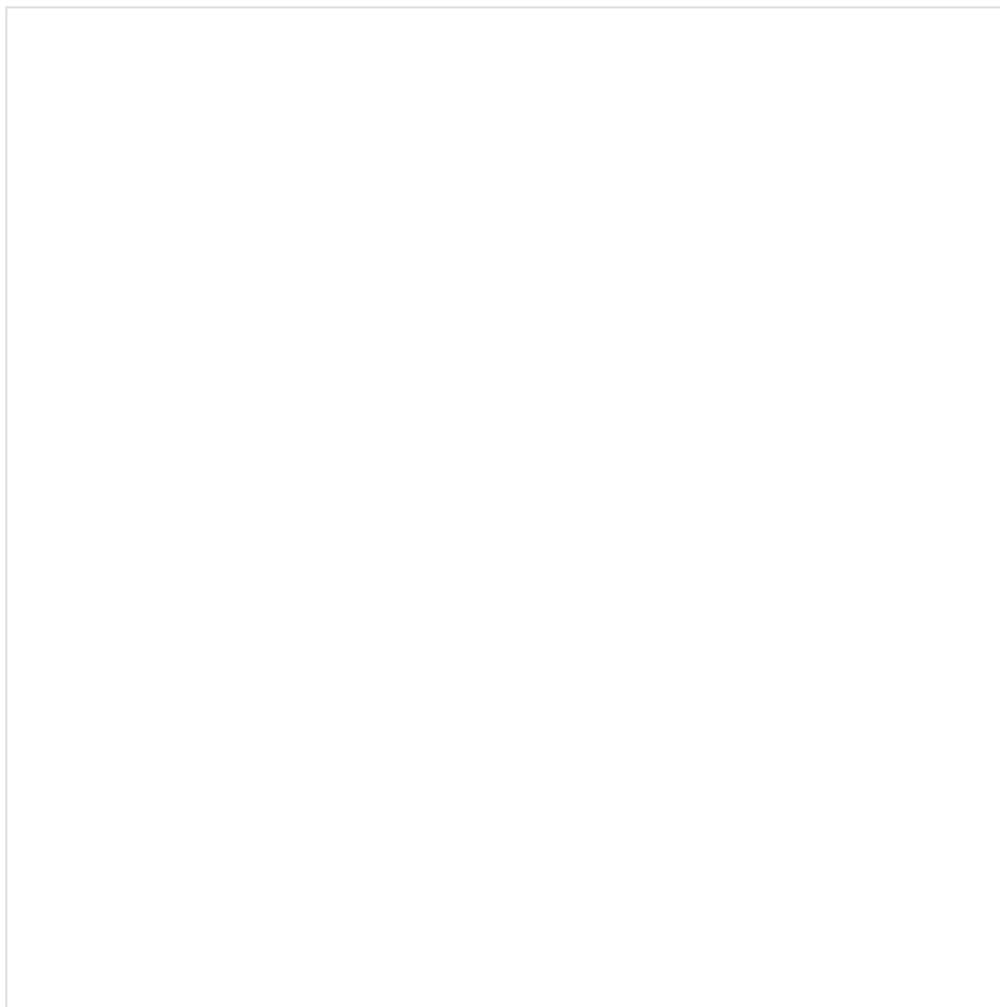
Or move it by clicking the actions ellipsis (...) and then **move to**:



I prefer dragging and dropping when I have a small number of folders and pipelines. Once the list extends below the visible part of the screen, it can be easier to use the *move to* feature.

### ***How do I build pipelines?***

You build pipelines by adding activities to it. To add an activity, expand the activity group and drag an activity onto the design canvas:



You can move the activity by selecting it, then dragging it. If you click on the blank design canvas and drag, you can move the entire pipeline around.

To **chain activities**, click and hold the little, green square on the right side of the first activity, then drag the arrow onto the second activity. The activities will now be executed sequentially instead of in parallel:



You can copy, paste, and delete activities in three ways. Select and use the keyboard shortcuts CTRL+C, CTRL+V, and DELETE. You can also **right-click** and use the menu:





Instead of copying and pasting, It might be easier to click the **clone** button:



And instead of using the keyboard or menu, you can click the **delete** button:



### ***How do I adjust the layout of a pipeline?***

If you are like me and cringe when you see messy layouts, don't worry! There's a special button just for us :) Click **auto align** to clean up the layout. Magic! That looks better:



You can zoom in and out using the **+** and **–** buttons. You can also **zoom to fit** the entire pipeline. If you have many activities, it will zoom *out* for you. If you have few activities, it will zoom *in* for you. Like, zoom in a *lot*:



If you want to reset the design canvas, click **reset zoom level**:



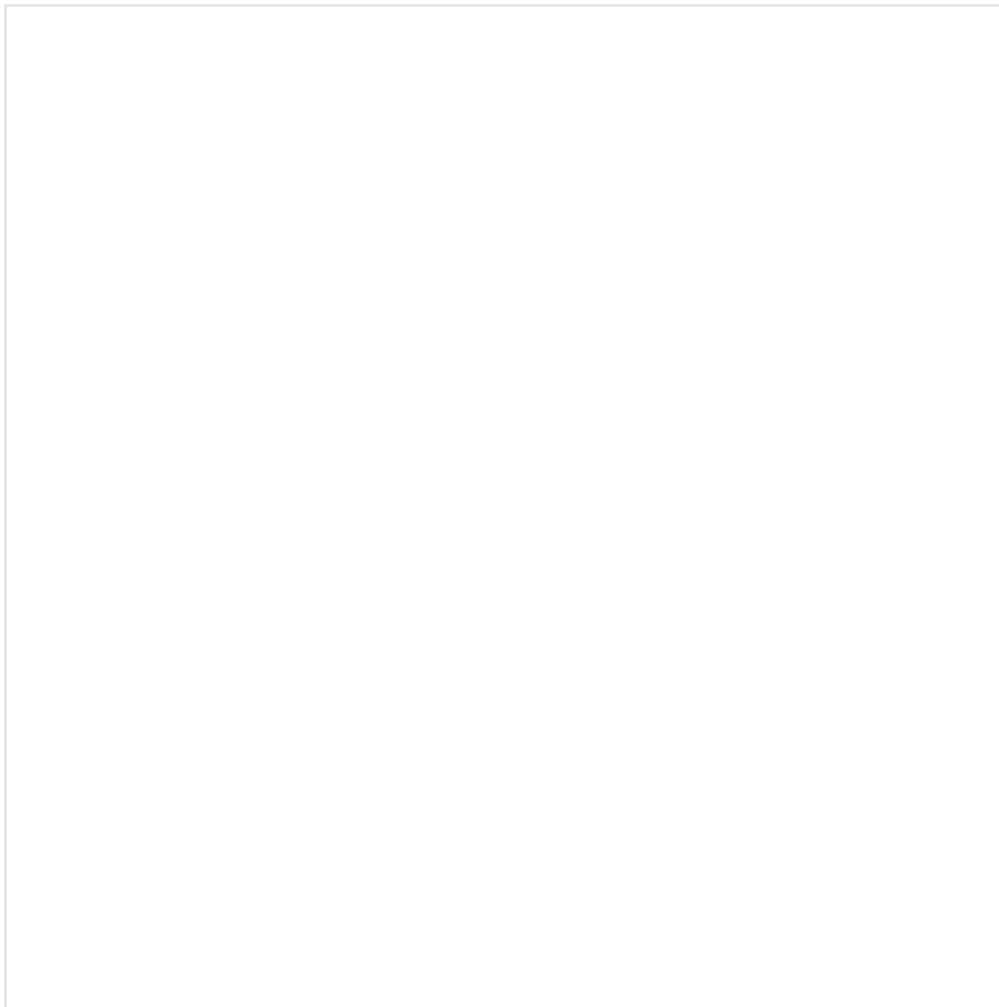
And! If you accidentally scroll or drag too much and all the activities disappear off the design canvas... *(It happens more often than I want to admit...)* Click zoom to fit and then reset zoom level. Tadaaaa! This centers the pipeline on your screen. It's a handy trick :)

Now, if you prefer to align your activities in a specific way, you can absolutely get creative and move them around as you like. Once you have finished your artwork, you can click **lock canvas** so you don't accidentally move anything. Just remember to *not* click auto align afterwards :)

### ***What are the general pipeline properties?***



You always have to specify the pipeline **name**. I also recommend adding good **descriptions**. You can even search for pipelines based on their descriptions!



In Azure Data Factory, you can execute the same pipeline many times – at the same time. Sometimes, this is a great way of improving performance. Other times, this is a bad idea. For example, you may not want to truncate and load a single table many times – at the same time. Or maybe you want to limit the number of times a single pipeline can connect to your source or destination? Whatever the reason, you can control this by changing the **concurrency** setting. The default number of concurrent pipeline runs is *unlimited*. Change this to 1 to ensure that the pipeline *has* to finish before it can be run again.

You can also add **annotations**, or *tags*, to your pipelines. We'll cover this more in a later post :)

### ***Do I have to use the graphical user interface?***

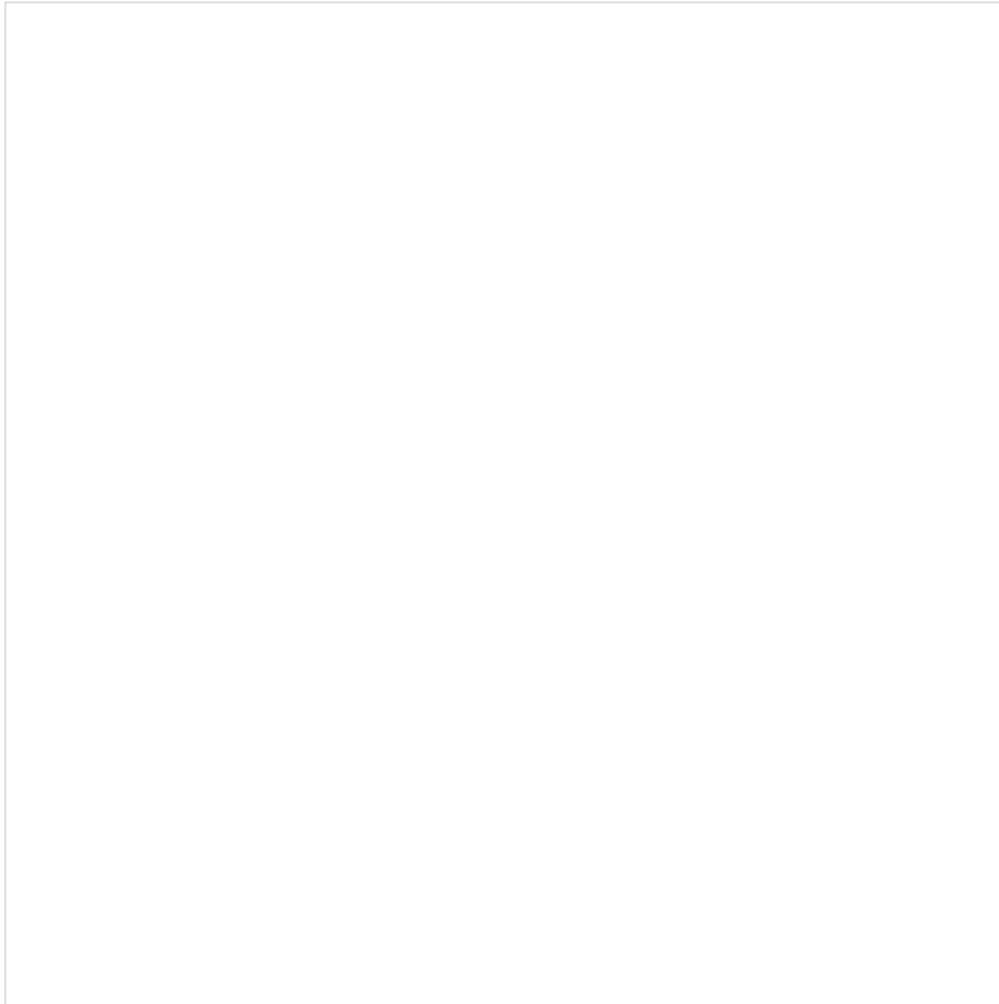
Nope! :)

If you want to, you can create *everything* in Azure Data Factory by writing JSON code. Click on **code** in the top right corner:





Edit the JSON code, and click **finish**:



I mostly use the graphical user interface when creating pipelines. But when I need to rename something in multiple activities, I often find it easier to edit this in the JSON code. You can use **CTRL+F** to find text, and **CTRL+H** to replace text:



And! This is something that I only recently discovered. While this code view looks fairly simple, there is a full code editor behind the scenes. Whaaat! :D You can discover this magic by right-clicking in the editor:



Click **F1** to open the **command palette**:



There are a lot of hidden goodies in the command palette. Some are useful, some might be more for the curious. Try it out! :)

### ***How do I check to see if my pipeline is valid?***

If you are building more complex pipelines, I recommend validating them once in a while. It's really annoying building something that you think is awesome, just to be told "sorry, that's not supported" :(

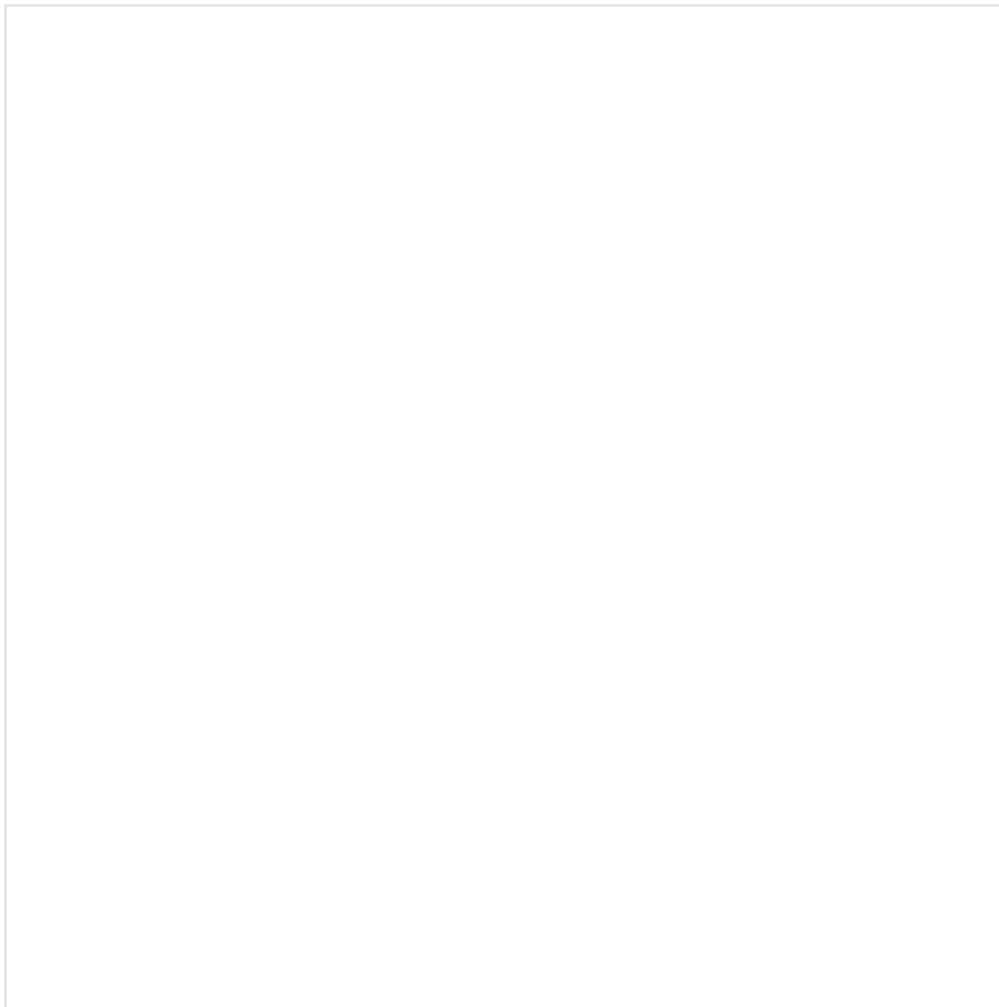
So! Click the little **validate** buttons once in a while:



It will tell you what isn't working, and you can click each error message to go to where you need to fix it:



Hopefully, most of the time you will see this friendly checkmark:



Now, here's the gotcha. It only checks what it can check inside Azure Data Factory. It doesn't go to external sources and validate whether or not the file you are trying to load actually exists, for example.

### ***How do I save my pipeline?***

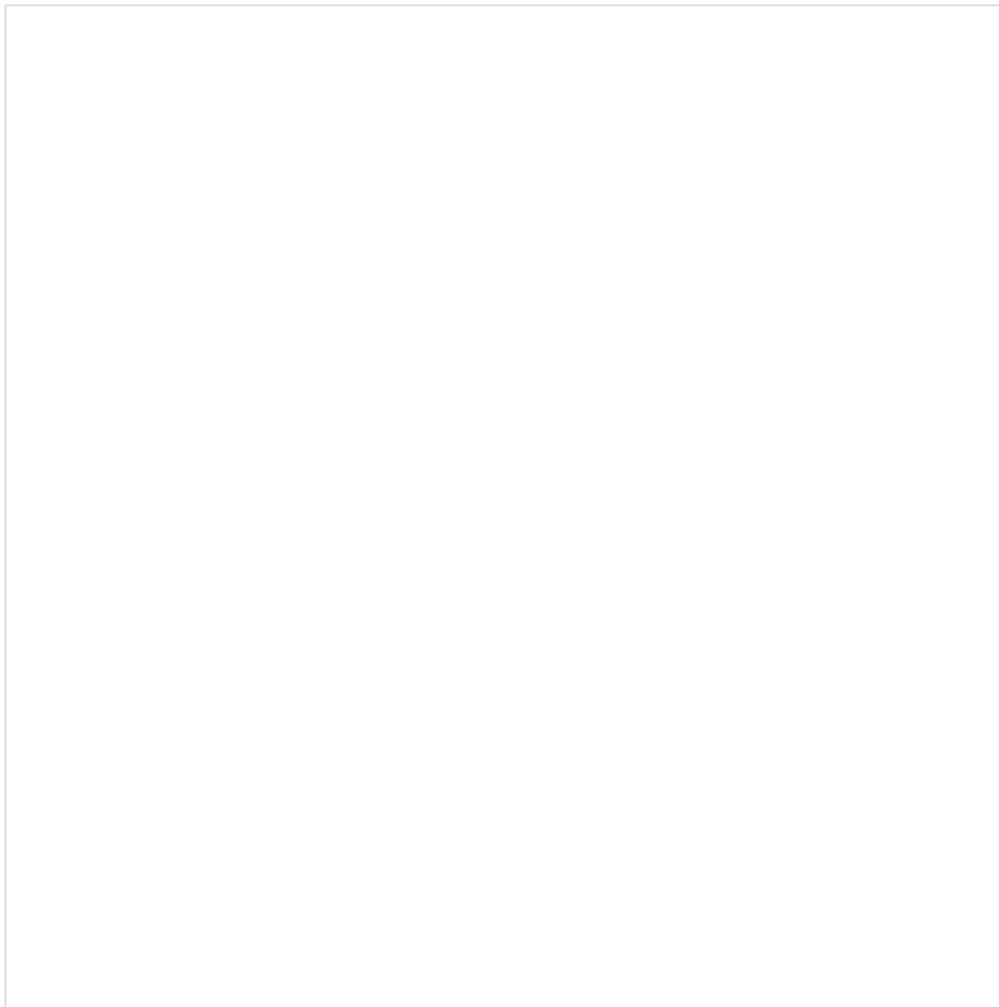
To save your pipeline, you need to make sure it validates first. That's another reason to validate once in a while. You definitely don't want to be in the middle of building something complex, realize you didn't get something quite right, notice it's the end of your workday, leave your browser running because you can't save the pipeline, come back the next day, and find that your computer restarted in the middle of the night. Trust me! It's not fun.

But as soon as it validates, click that **publish** button:





This deploys the change from the user interface to the Azure Data Factory service:



(Now you can leave work without worrying!)

## Summary

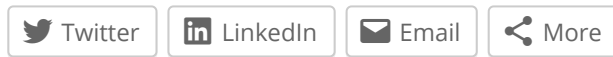
In this post, we went through Azure Data Factory pipelines in more detail. We looked at how to create and organize them, how to navigate the design canvas, and how to edit the JSON code behind the scenes.

But...

Does the “Copy\_blc” activity in the screenshots above bother you just as much as me? You have no idea how difficult it was to screenshot all the things, leaving that as is 😓 I did it for a reason, I promise! We originally built this pipeline using the Copy Data Wizard, which gives activities default names using a random suffix.

In the next post, we’re going to fix this. Let’s dig into the [copy data activity](#)!



**Share?****Related**[Copy Data Wizard in Azure Data Factory](#)[Parameters in Azure Data Factory](#)[Data Flows in Azure Data Factory](#)[🕒 Dec 5, 2019](#) [📁 Data Platform](#) [🔖 Azure Data Factory](#)

## About the Author



Cathrine Wilhelmsen is a Microsoft Data Platform MVP, BimlHero Certified Expert, Microsoft Certified Solutions Expert, international speaker, author, blogger, and chronic volunteer who loves teaching and sharing knowledge. She works as a Senior Business Intelligence Consultant at Inmeta, focusing on Azure Data and the Microsoft Data Platform. She loves sci-fi, chocolate, coffee, craft beers, ciders, cat gifs and smilies :)

**Find me!****Subscribe to new posts?****Recent Posts**[Sneaking back in... \(2020 edition\)](#)[Keyboard shortcuts for moving text lines and windows \(T-SQL Tuesday #123\)](#)[Speaking at NIC 2020](#)[Azure Data Factory Training Day at SQLBits 2020](#)[Personal Highlights from 2019](#)

## Popular Posts

---

[Table Partitioning in SQL Server - The Basics](#)

[Preparing for and Taking Microsoft Exam DP-200 \(Implementing an Azure Data Solution\)](#)

[Parameters in Azure Data Factory](#)

[Table Partitioning in SQL Server - Partition Switching](#)

[Custom Power BI Themes: Page Background Images](#)

## Top Tags

---

[Azure Data Factory Biml](#) [Certifications](#) [Don't Repeat Yourself](#) [Microsoft Ignite](#) [Notepad++](#) [PASS Summit](#) [Personal Precon](#) [Speaking](#) [SQLBits](#) [SQLFamily](#) [SQLHangout](#) [SQLSatOslo](#) [SQLSaturday](#) [SQL Server](#) [SSIS](#) [T-SQL](#) [Tuesday](#) [Volunteering](#) [Webinar](#)

## All Categories

---

▼

## Full Archive

---

▼

© *cathrine wilhelmsen* 2012-2020