Data Factory CI/CD

Index:

1. Continuous Integration

Continuous Integration is the practice of testing each change done to your codebase automatically and as early as possible. Continuous Delivery follows the testing that happens during Continuous Integration and pushes changes to a staging or production system.

For Azure Data Factory, continuous integration & delivery means moving Data Factory pipelines from one environment (development, test, production) to another. To do continuous integration & delivery, you can use Data Factory UI integration with Azure Resource Manager templates.

2. Azure Repos and Branches:

We are gonna use Azure Repos for better integration with Azure DataFactory. Based on the use case, There will be three branches.

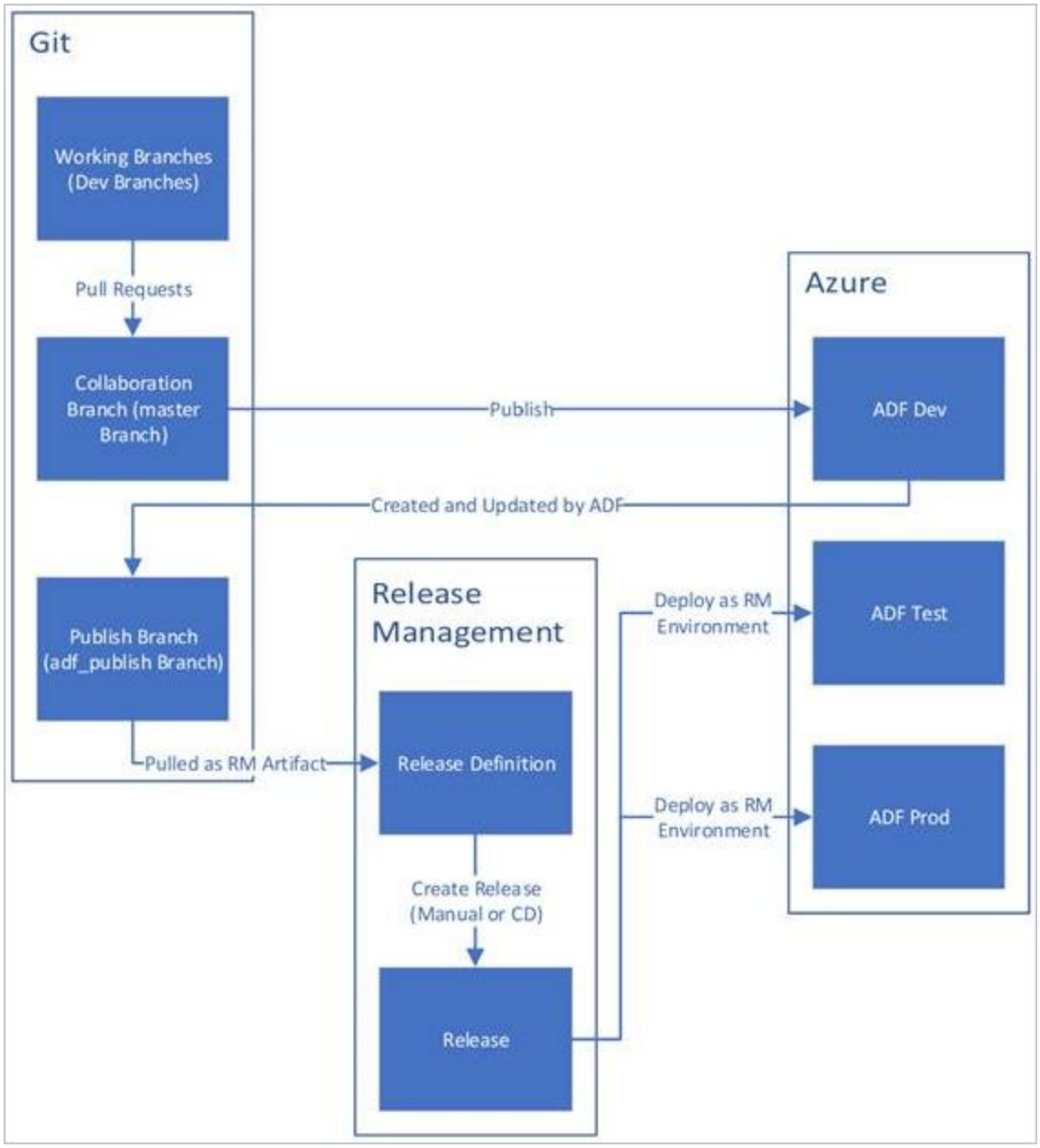
a. Development Branches
b. Master Branch (Collaboration Branch)
c. adf_publish Branch (Publish Branch)

a. **Development Branches** : Developer can create branches (Dev-1, Dev-2) and do his work like creating pipelines and commit it to respective branches inside the Repo from ADF Console using the integration.

b. **Master Branch** : Once Developer commits the code and verified by Peer, A pull request should be created and the same will be merged to Master.

c. **adf_publish Branch**: Now the Master has the latest code with the pipeline, Now we can publish the changes from ADF Console and then the ADF will create a new branch adf_publish which actually contains the ARM Template and parameters which will be deployed to other environments through Azure Pipelines.

3. Workflow of CI / CD Pipeline:

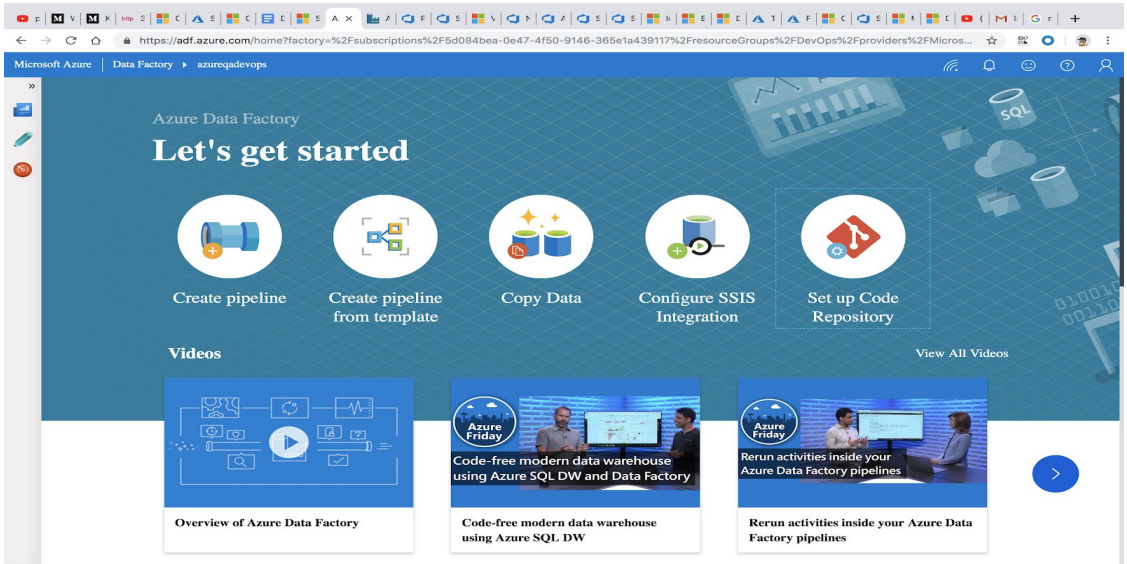


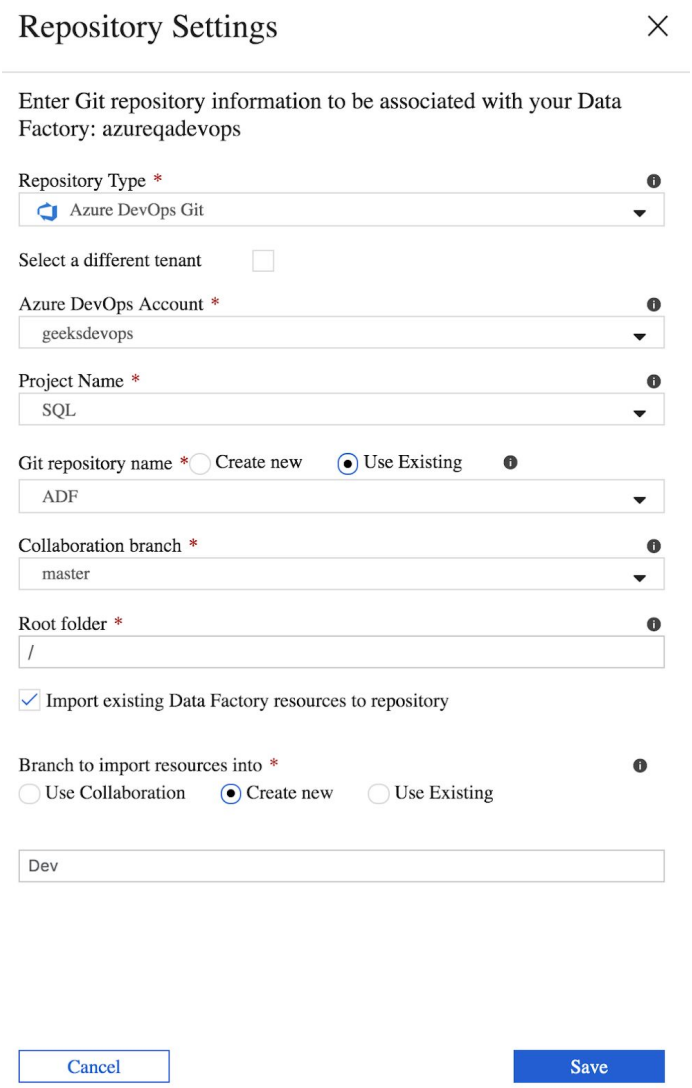
Here is the entire lifecycle for continuous integration & delivery that you can use after you enable Azure Repos Git integration in the Data Factory Console:

1. Set up a development data factory with Azure Repos in which all developers can author Data Factory resources like pipelines, datasets, and so forth.
2. Then developers can modify resources such as pipelines. As they make their modifications, they can select Debug to see how the pipeline runs with the most recent changes.
3. After developers are satisfied with their changes, they can create a pull request from their branch to the master branch (or the collaboration branch) to get their changes reviewed by peers.
4. After changes are in the master branch, they can publish to the development factory by selecting Publish.
5. When the team is ready to promote changes to the test factory and the production factory, they can export the Resource Manager template from the master branch, or from any other branch in case their master branch backs the live development Data Factory.
6. The exported Resource Manager template can be deployed with different parameter files to the test factory and the production factory.

4. Azure Data Factory Pipeline

i. Configure Git Integration with ADF Resource





ii. Create an ADF Pipeline

For Beginners,

Create a blob copy job from one folder to another in Dev-ADF Resource.
Ref: https://docs.microsoft.com/en-us/azure/data-factory/tutorial-copy-data-portal

Data Factory | Publish All | Validate All | Refresh | Discard All | Data Flow Debug | ARM Template

:ory Resources

Filter resources by name

ipelines 1

pipeline1

atasets 2

DestinationBlob

SourceBlob

ata Flows (Preview) 0

pipeline1 | DestinationBlob | SourceBlob

Activities

Search Activities

- Move & Transform
- Batch Service
- Databricks
- Data Lake Analytics
- General
- HDInsight
- Iteration & Conditionals
- Machine Learning

Save as template | Validate | Debug | Add trigger

Copy Data

Copy Data1

General | Parameters | Variables | Output

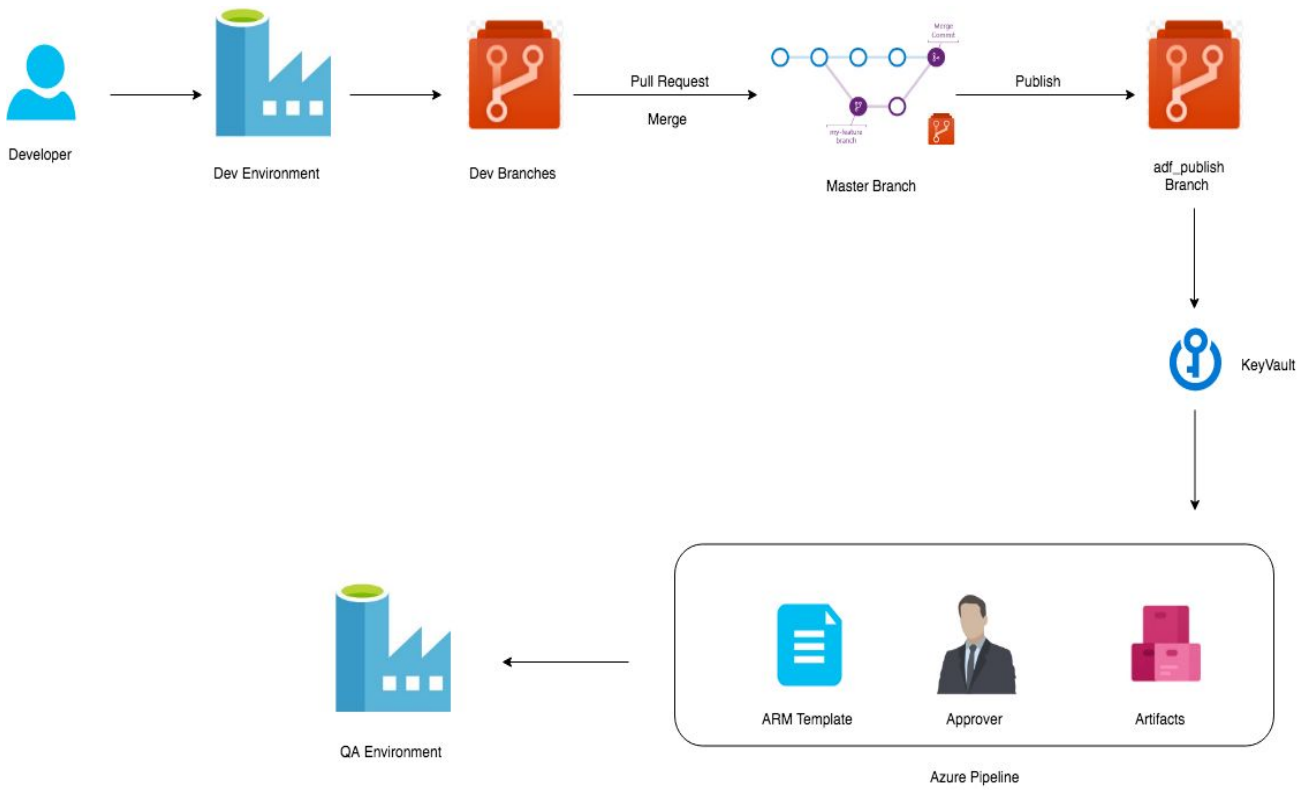
Name * pipeline1

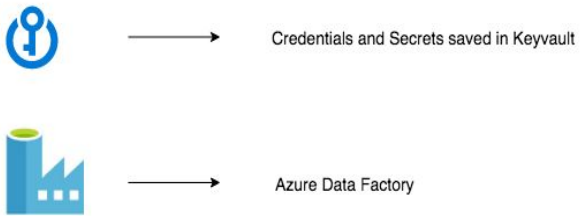
Description

Concurrency

Annotations + New

5. Azure DevOps Pipeline

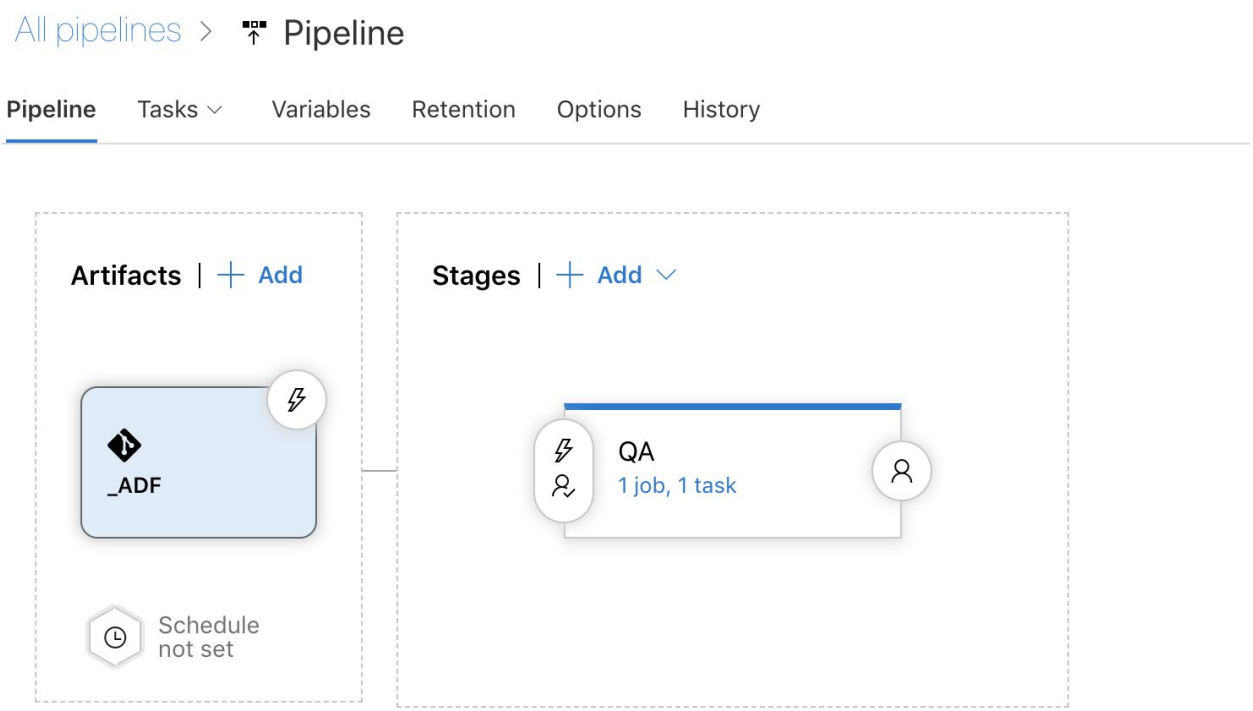




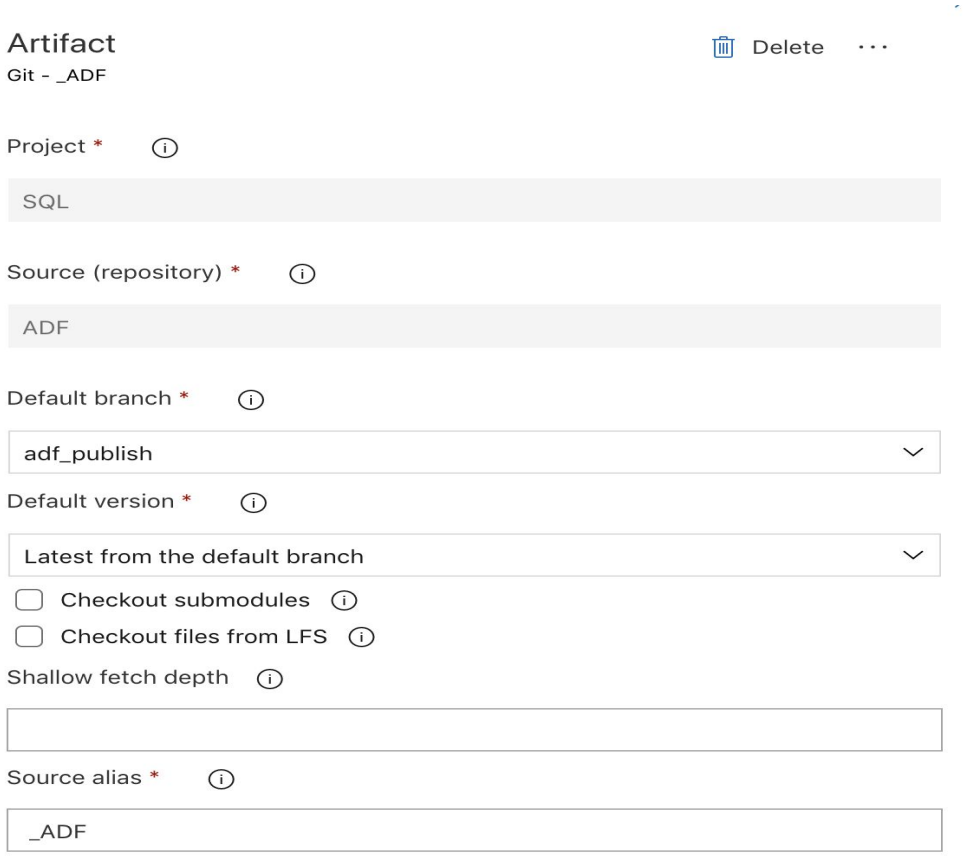
Create a Release Pipeline in Azure DevOps and configure the trigger , let's say whenever there's a change happens in adf_publish. Pipeline will be initiated.

Pipeline Stages:

Phase 1 : Trigger the Pipeline with Artifacts
Phase 2 : Manual Approval
Phase 3 : Deployment

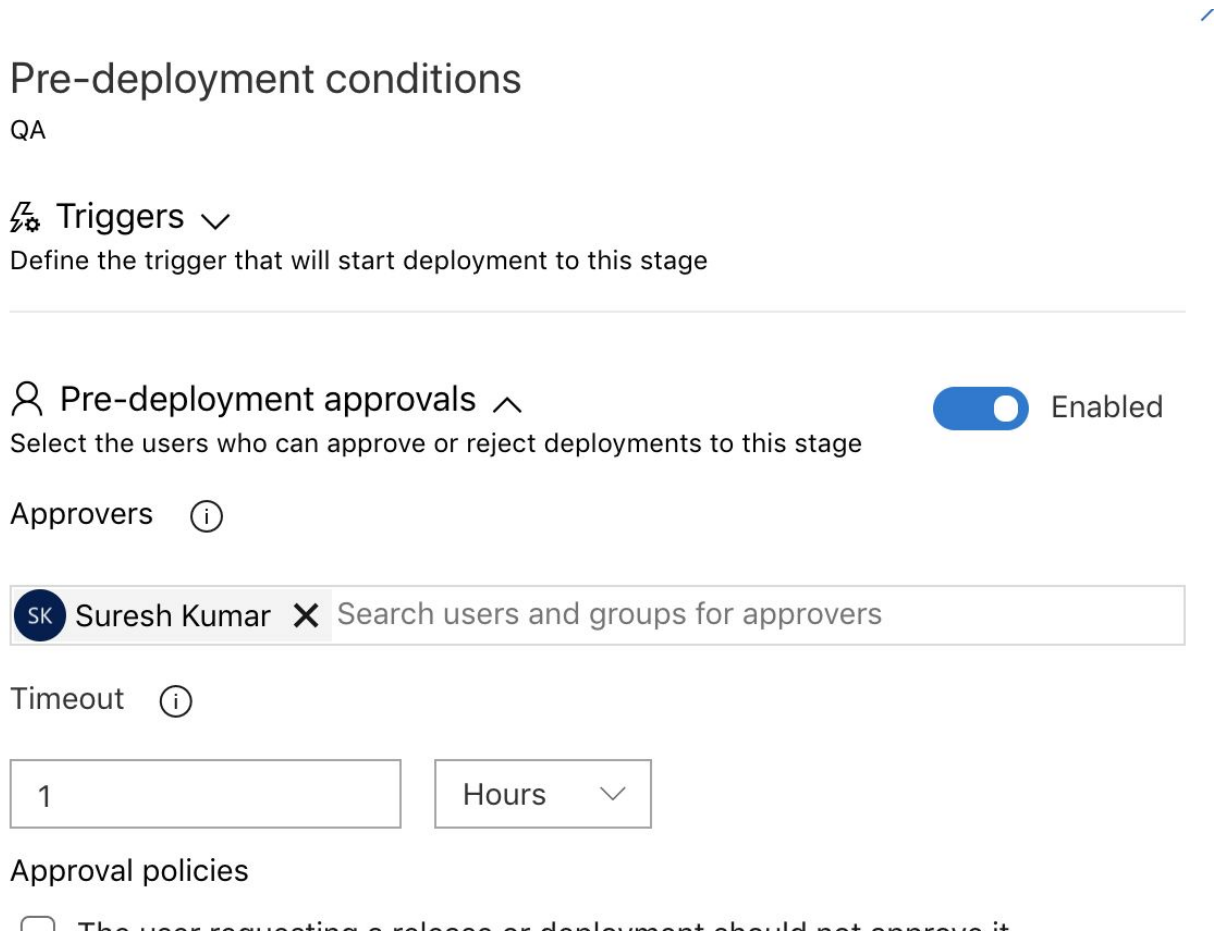


Phase 1 (Artifacts):



Phase 2 (Manual Approval) :

The Artifact will be downloaded and before deploying to QA, It will be sent for approval. The respective owners will get email for the same. Once they approve, It will go to the next stage.



Phase 3 (Deployment) :

The Artifact will be downloaded and Deployment Stage (Stage 2) will filter the ARM Template based on the configurations mentioned and it will initiate the Pipeline Deployment on QA-ADF Resource

Azure Resource Group Deployment ⓘ

View YAML | Remove

Task version 2.*

Display name *

Azure Deployment:Create Or Update Resource Group action on DevOps

Azure Details

Azure subscription * ⓘ | Manage

CloudGeeks (5d084bea-0e47-4f50-9146-365e1a439117)

Scoped to subscription 'CloudGeeks'

Action * ⓘ

Create or update resource group

Resource group * ⓘ

DevOps

Location * ⓘ

East US

Template

Template location *

Linked artifact

Template * ⓘ

$(System.DefaultWorkingDirectory)/_ADF/azuredfdevops/ARMTemplateForFactory.json

Template parameters ⓘ

$(System.DefaultWorkingDirectory)/_ADF/azuredfdevops/ARMTemplateParametersForFactory.json

Override template parameters ⓘ

-factoryName "azureqadevops" -SourceBlob_properties_typeProperties_fileName "text" -SourceBlob_properties_typeProperties_folderPath "adftutorial/input" -AzureKeyVault1_properties_typeProperties_baseUrl "https://bwaccesskey.vault.azure.net/" -DestinationBlob_properties_typeProperties_fileName "qa" -

Deployment mode * ⓘ

Incremental

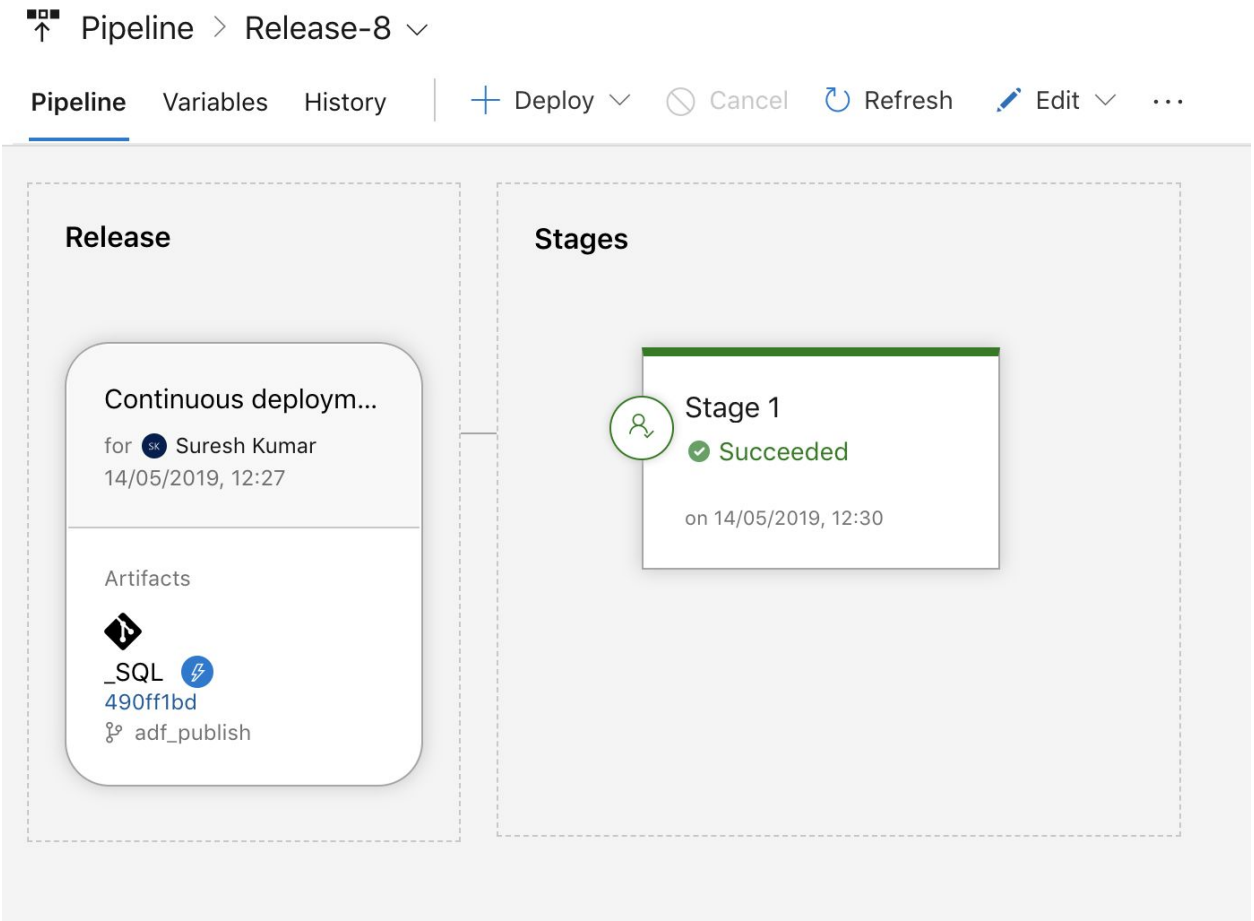
Note:

1. Keyvault : Azure DataFactory needs access to do modifications inside Azure Blob Storage for copy activity. So we need to provide the Access Key and it should not be visible , So we are using Azure KeyVault here.
2. Configuration : There will be small configuration changes based on environment, let's say,

   QA will have QA-ADF and

   Prod will be having PROD-ADF

We need to update the configuration inside the Stage 3, before deploying, You can find the same in above image **"Override template parameters"**


Pipeline
Release-8
Pipeline
Variables
History
Deploy
Cancel
Refresh
Edit
Release
Continuous deploym...
for Suresh Kumar
14/05/2019, 12:27
Artifacts
_SQL
490ff1bd
adf_publish
Stages
Stage 1
Succeeded
on 14/05/2019, 12:30