Azure Player
Play with data & have fun!

Azure Data Factory: A deployment challenges

MVP Microsoft® Most Valuable Professional

Azure Player
Play with data & have fun!

snowflake

twoday
kapacity

ITPA
CONSULTING

Microsoft

data on
Power your data

Raffle Prizes

TIMEXTENDER

[stoltze][it]

ORANGEMAN

exmon

KPMG

Fellowwind

Kamil Nowiński

Microsoft Data Platform **MVP**

Speaker, blogger, data enthusiast

Group Manager at Avanade UK&I (www.avanade.com)

>20 yrs experience as DEV/BI/(DBA)

Member of the Data Community PL

Founder of blog AzurePlayer (www.AzurePlayer.net)

GitHub: #adftools, SCD Merge Wizard and more…

SQL Server Certificates:

MCITP, MCP, MCTS, MCSA, MCSE Data Platform,

MCSE Data Management & Analytics, DevOps Expert

Moreover: Bicycle, Running, Digital photography

@NowinskiK, @Azure_Player

Blog

- Technical posts
- Various skill level
- Cheet sheets
- Recommended books
- Many useful other links
- Interviews (Podcast)
- YouTube Channel: www.AzurePlayer.net/YouTube

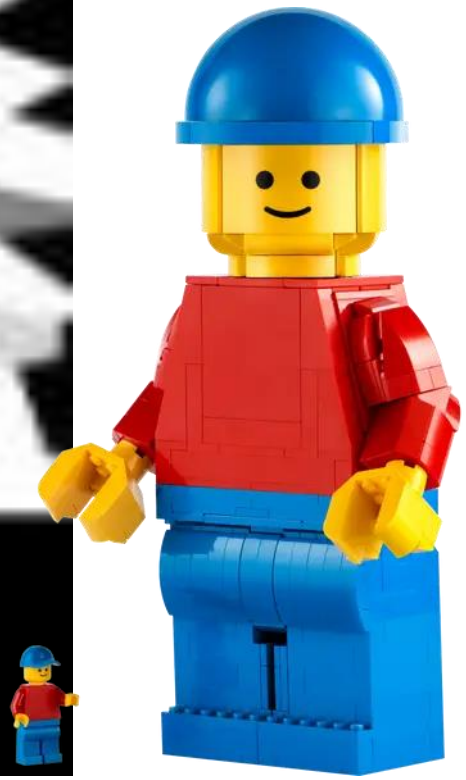
Azure Player

Play with data & have fun!

www.AzurePlayer.net

Scan me

SQLPlayer is renaming



“Ask SQL Family” #podcast

YouTube

www.AzurePlayer.net/YouTube

LEGO
https://azureplayer.net/batman-lego

Slides available:

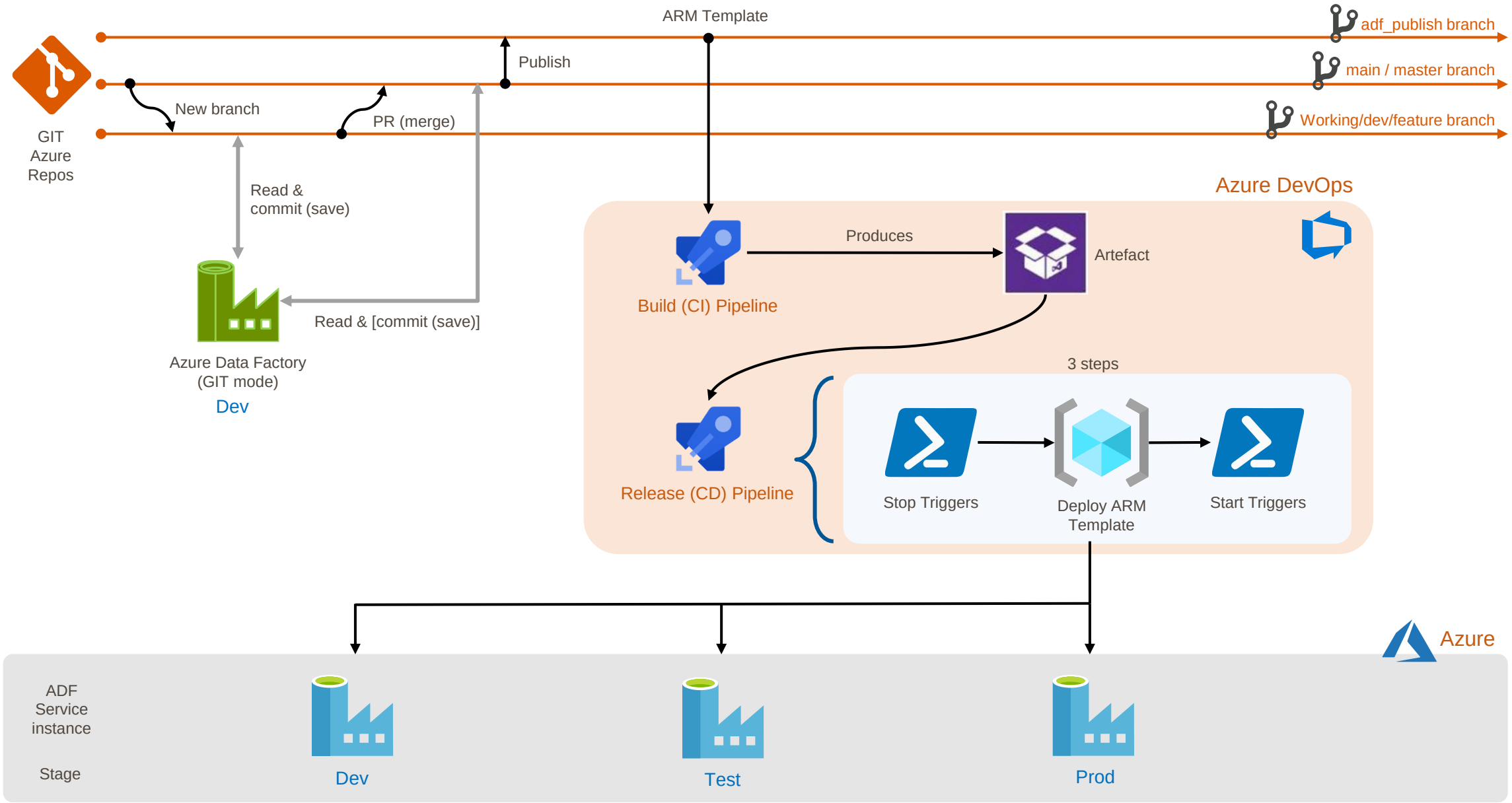
https://azureplayer.net/slides

AGENDA

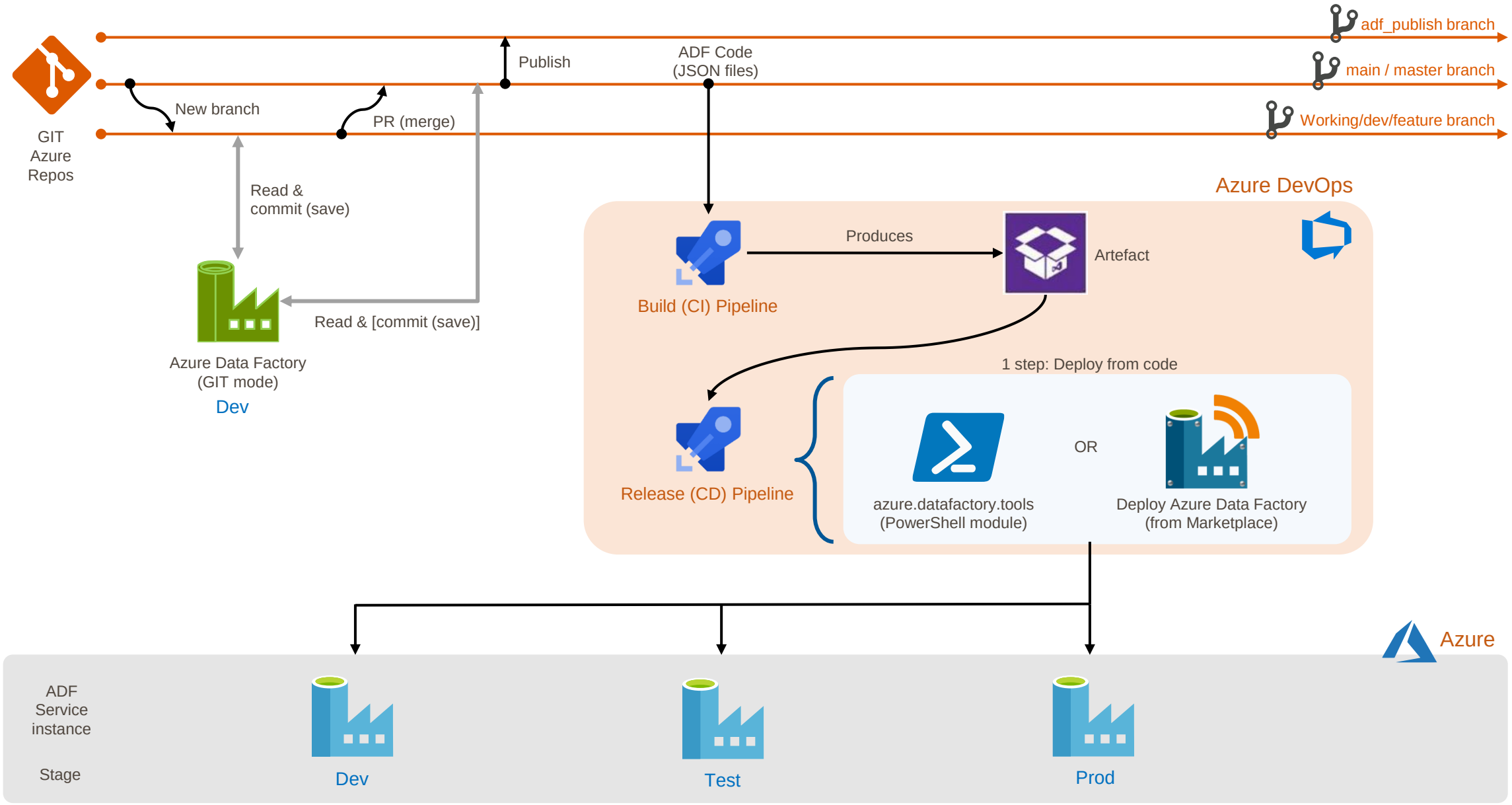
Azure Player
Play with data & have fun!

- Azure Data Factory – DEPLOYMENT only
- Two (three?) methods of ADF deployment
- How these methods work
- Differences
- npm module from Microsoft – now you can fully automate CI (build)
- #adftools – make your life easier!

Deployment (1): Microsoft's method (ARM Template)

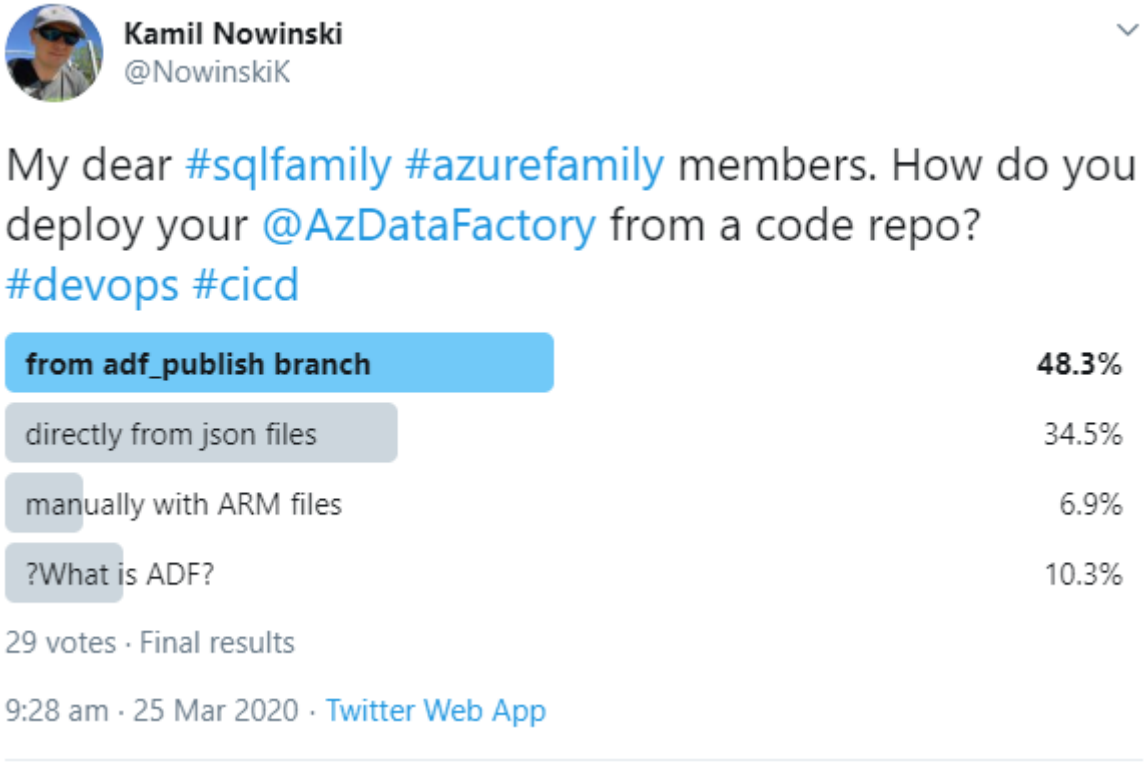


Deployment (2): Directly from code method



Who use what?

Azure Player
Play with data & have fun!



https://twitter.com/NowinskiK/status/1242745169394442240

ADF – ARM Template enhanced deployment method

Azure Player
Play with data & have fun!

ARM Template from “adf_publish” branch

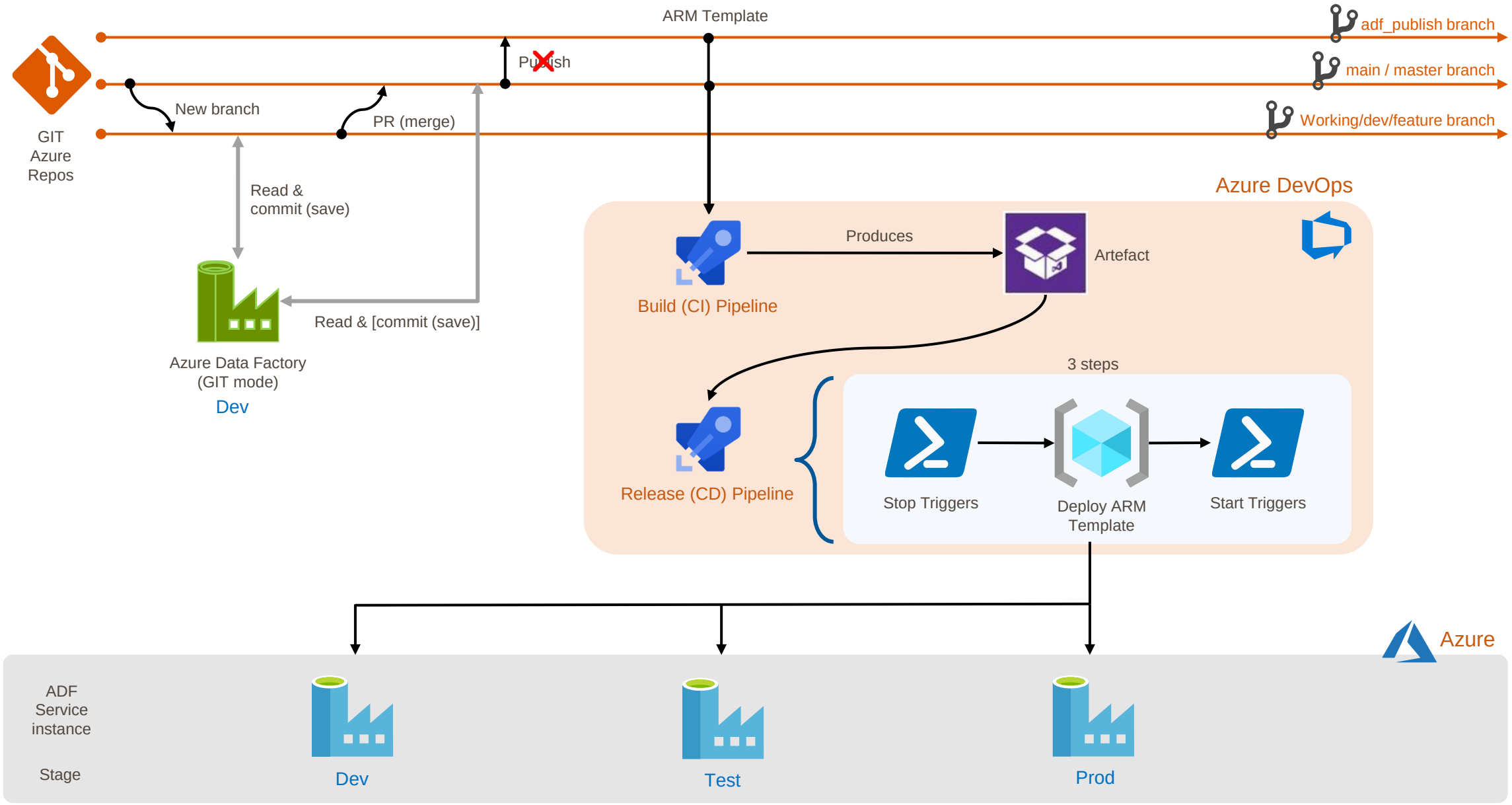
- Faster, “incremental”
- Appears in „Deployments” for Resource Group
- Parametrize elements exposed within the ARM Template Parameter *
- Full ADF (all artefacts) can be deploy only
- Restriction of 256 parameters
- ~~Limitation to one publish branch only (adf_publish)~~
- ~~Manual “Publish” step~~
- Not user-friendly npm library
- Requires adding “package” file to a repo

Since 2021

Rest-Api/PowerShell script from code (JSON objects)

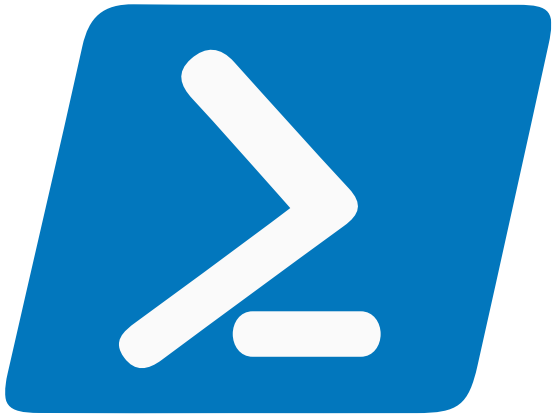
- ~~Slower~~ (no longer due to incremental deployment)
- Doesn’t appear in „Deployments” for Resource Group
- Parameterize any artefact of the Data Factory
- Selectively deploy a subset of artefacts
- Eliminates an enforcement to use only one (adf_publish) branch if company’s branches policy is much complex

Automated publishing via CI/CD within npm & ARM Template



DEMO
... in a minute;
a question first...

What do you prefer?



DEMO #1
ARM Template deployment

Azure Player
Play with data & have fun!

#adftools

#adftools

Two tools:

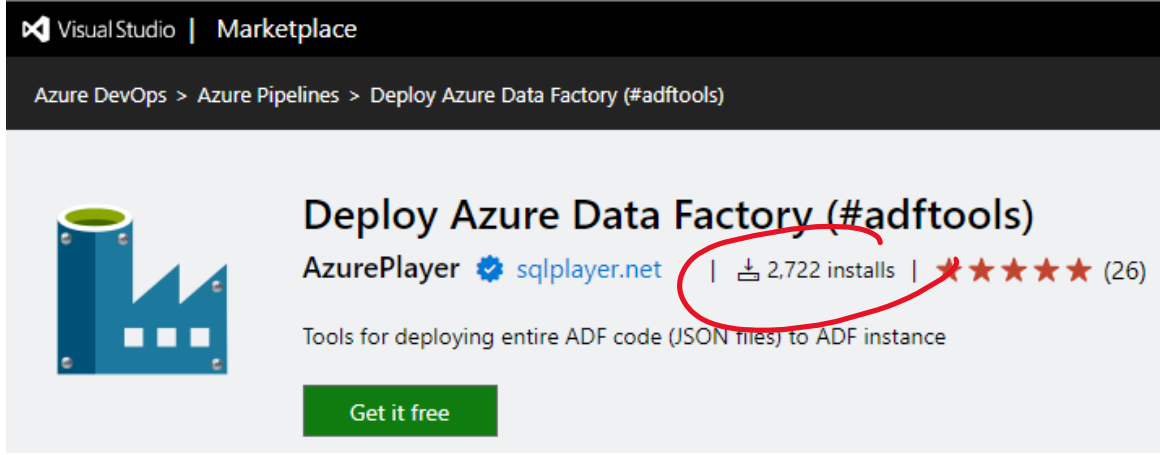
PowerShell module (azure.datafactory.tools)

Azure DevOps extension (3 tasks)

https://azureplayer.net/adftools/

DEMO #2
Deployment with PowerShell

Azure DevOps: Custom Tasks



Deploy Azure Data Factory

Key concept:

- Free & Open-Source
- One task for everything when it comes to the publishing of ADF
- Basically, it is another (UI) layer on top of „azure.datafactory.tools" module
- Public Release (GA) since 23/12/2020
- Extension contains 3 tasks that cover full deployment life-cycle for ADF

Any ideas or questions? Leave it here:
https://github.com/Azure-Player/azure.datafactory.tools/issues

PowerShell module: azure.datafactory.tools

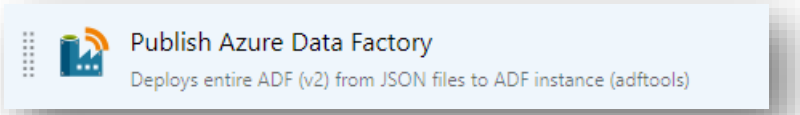
Azure Player
Play with data & have fun!

- Fully written in PowerShell, compatible with 5.1
- Uses Microsoft's PS module (Az.DataFactory) for management of ADF objects
- Available in PowerShell gallery
- `Publish-AdfV2FromJson` function - capabilities:
  - Creation of Azure Data Factory, if not exist (option)
  - Deployment of all type of objects: pipelines, datasets, linked services, data flows & power query, triggers, integration runtimes
  - Copes with dependencies (multiple levels) between objects when deploying (no more worrying about object names)
  - Build-in mechanism to replace the properties with the indicated values (CSV file)
  - Stop/start triggers (option)
  - Dropping objects when not exist in the source (code) (option)
  - Selective deployment: Filtering (include or exclude) objects to be deployed by name and/or type
  - Publish options allow you to control:
    - Whether stop and restarting triggers
    - Whether delete or not objects not in the source
    - Whether create or not a new instance of ADF if it not exist
- Free & Open-Source

https://github.com/Azure-Player/azure.datafactory.tools

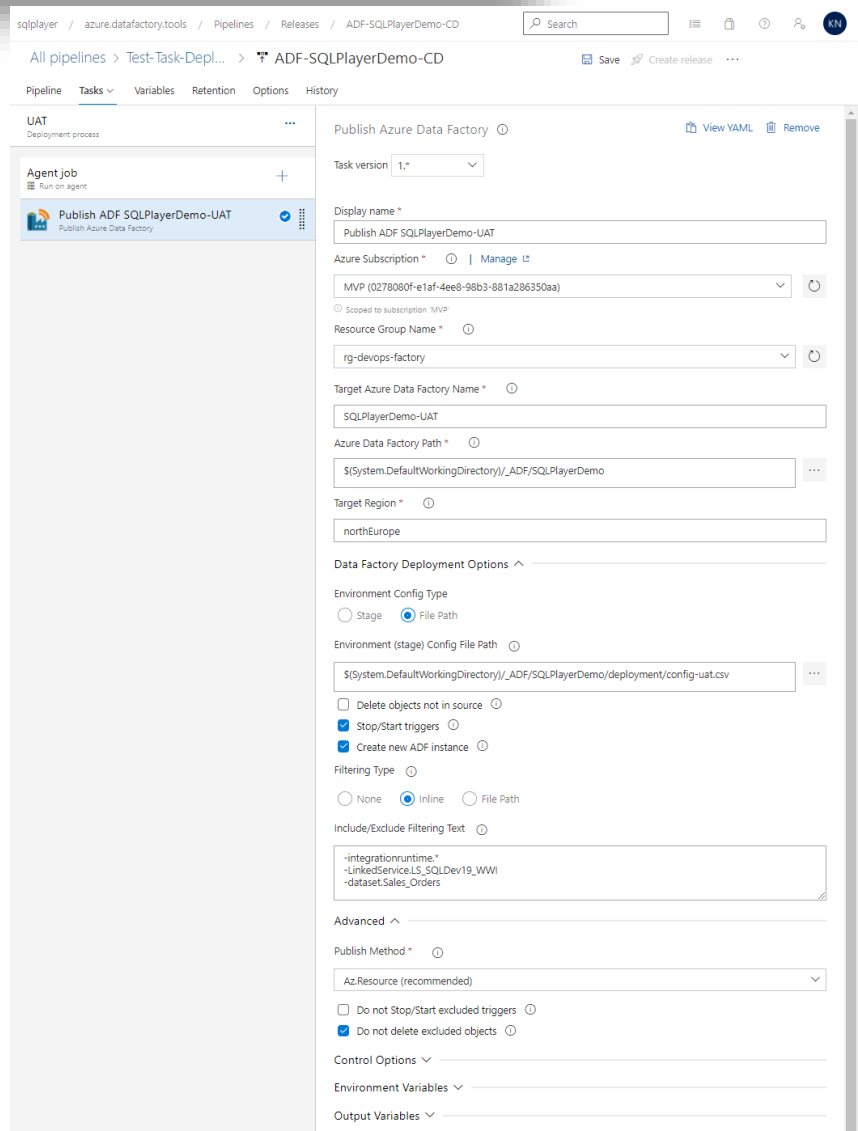
Task in Azure DevOps: Publish ADF (CD)

Azure Player
Play with data & have fun!



Key capabilities:

- Creation of Azure Data Factory, if not exist (option)
- Deployment of all type of objects: pipelines, datasets, linked services, data flows, triggers, integration runtimes, Managed Virtual Network, Managed Private Endpoint
- Copes with dependencies (multiple levels) between objects when deploying (no more worrying about object names)
- Build-in mechanism to replace the properties with the indicated values (CSV & JSON file format)
- Update, add or remove any property of ADF artefact
- Selective deployment declared in-line or by pointed file
- Stop/start triggers (option)
- Dropping objects when not exist in the source (code) (option)
- Filtering (include or exclude) objects to be deployed by name and/or type and/or type
- Filtering supports wildcards
- Publish options allow you to control:
    - Whether stop and restarting triggers
    - Whether delete or not objects not in the source
    - Whether create or not a new instance of ADF if it not exist
- Tokenisation in config file allows replace any value by Environment Variable or Variable from DevOps Pipeline
- Global Parameters



Selective deployment with PowerShell module

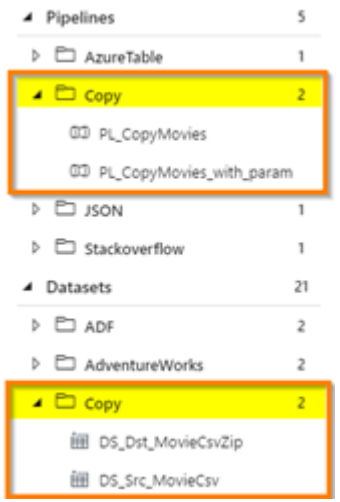
Azure Player
Play with data & have fun!

You can select objects by objects types & name using include or exclude option.

Allows to select the objects by belonging to a folder (picture)

Name can be **wildcarded**, so all such variants are possible:

```
trigger.*
dataset.DS_*
*.PL_*
linkedService.???KeyVault*
pipeline.ScdType[123]
*.*@testFolder
managedVirtualNetwork*.*
*managedPrivateEndpoint.*
```



DEMO #3
Deployment with Azure DevOps

Bonus capabilities

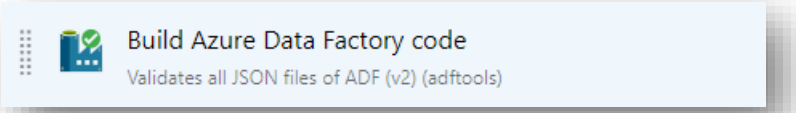
New task in Azure DevOps: Build ADF (CI)

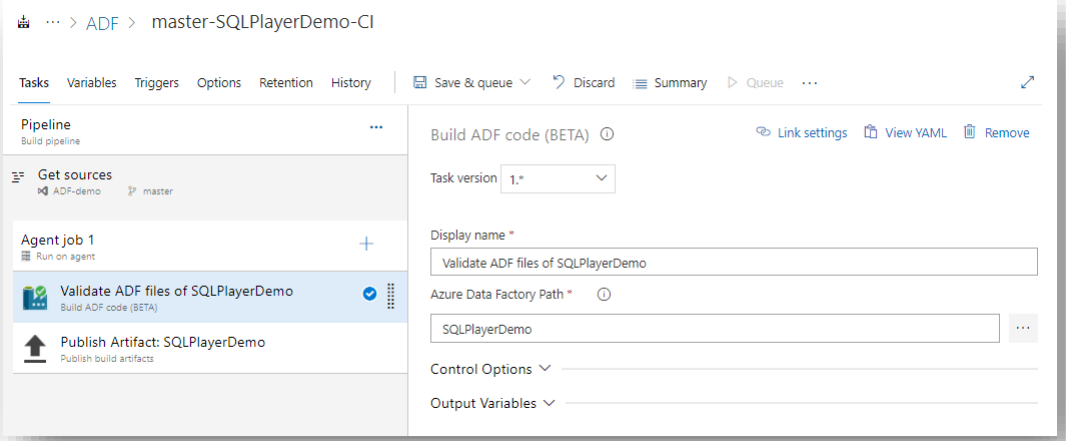
Azure Player
Play with data & have fun!

The task has 2 modes:

- **Build only**
  - Reads all files and validates its json format
  - Checks whether all dependant objects exist
  - Checks whether file name equals object name

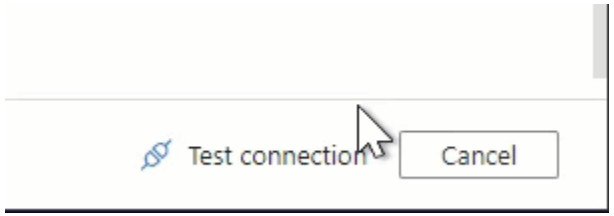
- **Validate & Export ARM Template**
  - uses ADFUtilities NPM package provided by Microsoft
  - Counterpart of Validate all and Export ARM Template in ADF UI

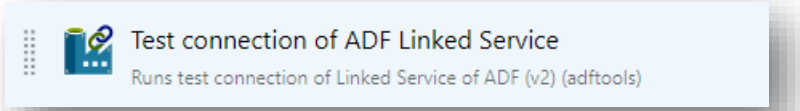


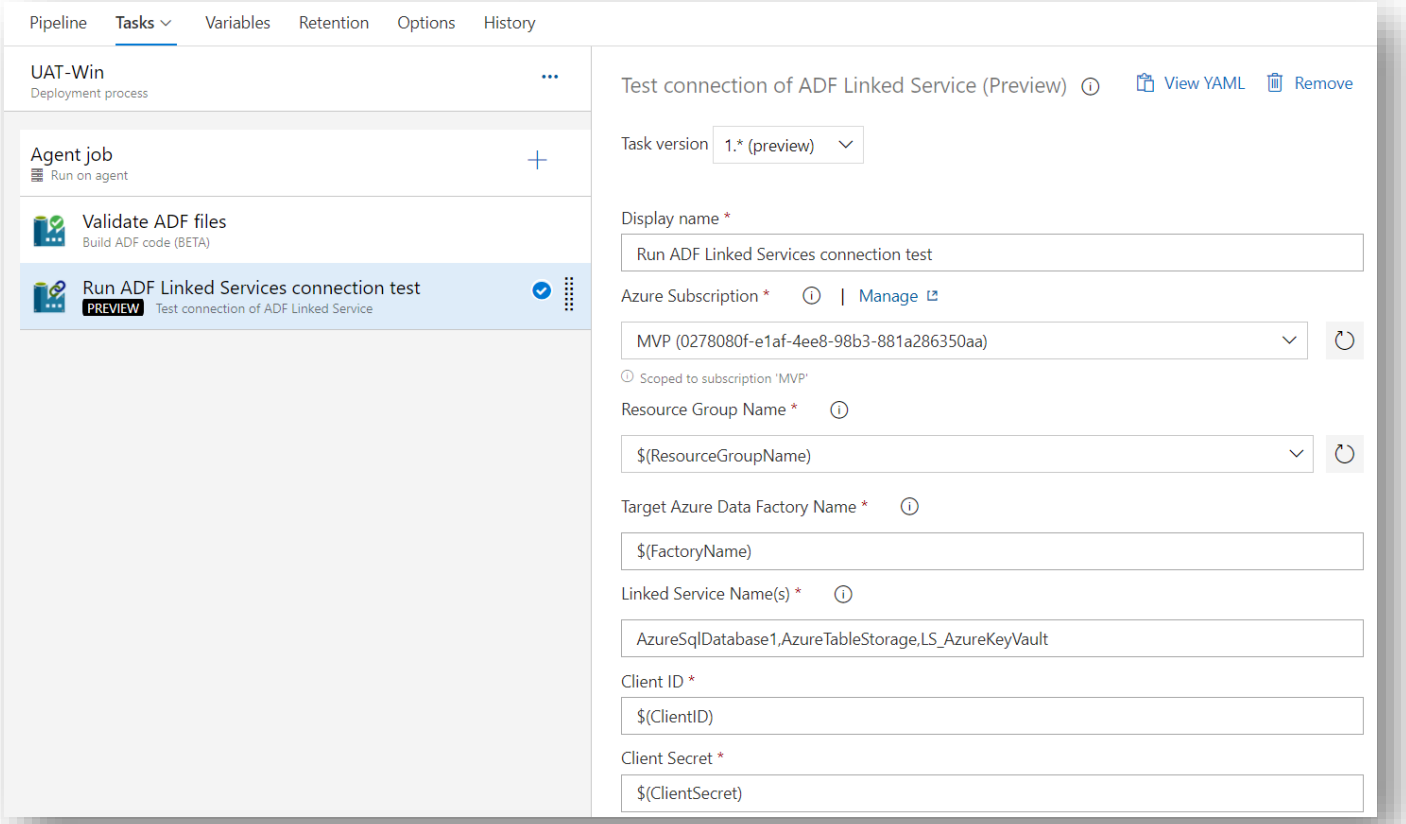


New task in Azure DevOps: Test Connection

- Run „Test connection" for a Linked Service
- Smoke tests of (some) Linked Services
- In preview

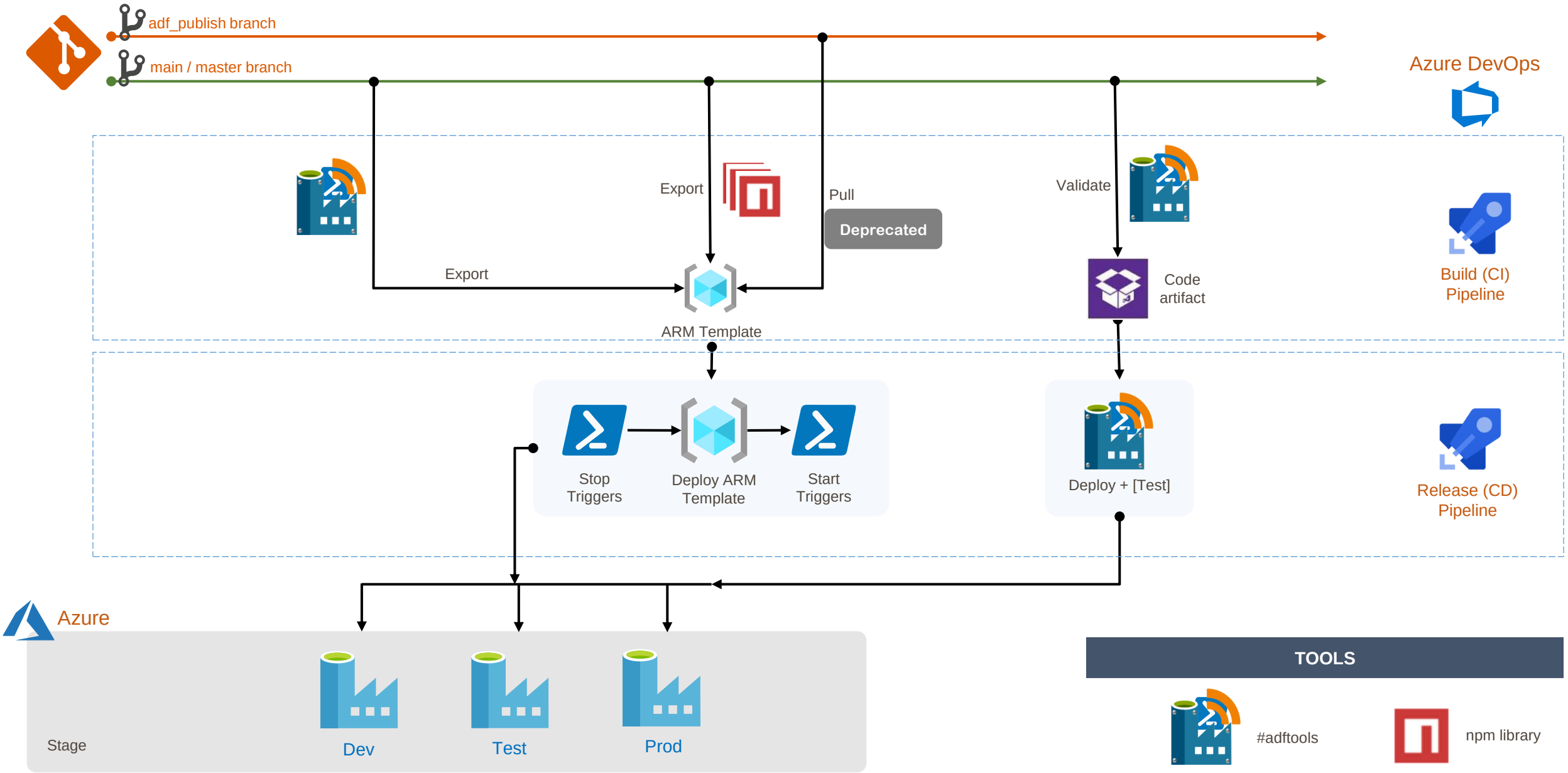






Summary

Takeaway: ADF deployment - Possible paths



#Adftools - Resources

Azure Player
Play with data & have fun!

- https://azureplayer.net/adftools
- https://github.com/Azure-Player/azure.datafactory.tools
  - Issues
  - Discussions, FAQ
- PowerShell Gallery: azure.datafactory.tools
- Marketplace: Deploy Azure Data Factory (extension for Azure DevOps)
- Microsoft ADF (npm) Utilities

- https://github.com/Azure-Player/azure.**synapse**.tools (preview!)

Questions?

Thank you!

kamil@azureplayer.net

@NowinskiK @Azure_Player

AzurePlayer.net

https://AzurePlayer.net/slides

Kamil Nowinski
Microsoft Data Platform MVP
Analytics Architect, Azure DevOps Engineer Expert

Save the dates!
ADF + Synapse Deployment workshop (2 days)

When: **November 2023**

Where: **Aarhus, Denmark**

Agenda: Azure Data Integration DevOps Training

Tickets: (link will be above)

Kamil Nowinski
Microsoft Data Platform MVP
Analytics Architect, Azure DevOps Engineer Expert