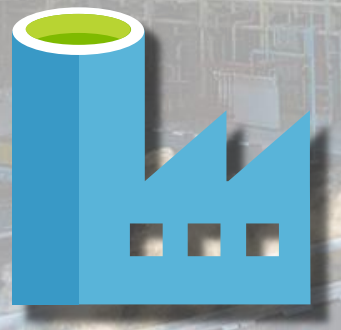
Azure Player
Play with data & have fun!

Azure Data Factory: A deployment challenges



Microsoft® MVP Most Valuable Professional

SQL Day

16 edycja konferencji SQLDay

Data Community

13-15 maja 2024, WROCŁAW + ONLINE

partner platynowy

lingaro

partner złoty

DXC TECHNOLOGY

VOLVO

software one

partner srebrny

accenture

Nordcloud an IBM Company

Capgemini

TECHNOLOGY INNOVATION DATA KNOWLEDGE hdk

elitmind think bright

C&F

Kamil Nowiński

Microsoft Data Platform **MVP, Databricks Champion**
Speaker, blogger, **YouTuber**, data enthusiast

Group Manager at Avanade UK&I
>20 years of experience in IT

Founder of blog AzurePlayer

GitHub: **#adftools**, SCD Merge Wizard and more…
Member of the Data Community PL
Former co-organiser of SQLDay (PL), volunteer at SQLBits (UK)

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

Azure Player
Play with data & have fun!

Use Data Factory pipelines in Microsoft Fabric | Lab 04

Kamil Data Geek - Azure explained • 127 views • 2 weeks ago

Work with Delta Lake tables in Microsoft Fabric | Lab 03

Kamil Data Geek - Azure explained • 192 views • 1 month ago

Use Apache Spark in Microsoft Fabric | Lab 02

Kamil Data Geek - Azure explained • 294 views • 1 month ago

Get started with Lakehouses in Microsoft Fabric | Lab 01

Kamil Data Geek - Azure explained • 313 views • 2 months ago

Ask SQL Family

▶ Play all

This set of videos promote episodes of "Ask SQL Family" podcast.

Grant Fritchey: Why "Scary DBA" and what's his favourit...

Kamil Data Geek - Azure explained
Premieres 16/05/2024, 17:50

Notify me

Dejan Sarka - is Python in SQL Server good idea? | Ask...

Kamil Data Geek - Azure explained
32 views • 12 days ago

Pedro Lopes - should DBAs worry about their job? | Ask...

Kamil Data Geek - Azure explained
38 views • 3 weeks ago

Michael Rys - father of U-SQL | Ask SQL Family 008

Kamil Data Geek - Azure explained
21 views • 1 month ago

Cathrine Wilhelmsen - Women in IT | Ask SQL Fami...

Kamil Data Geek - Azure explained
23 views • 2 months ago

Bob Ward - SQL Server during the last 30 years | Ask SQL...

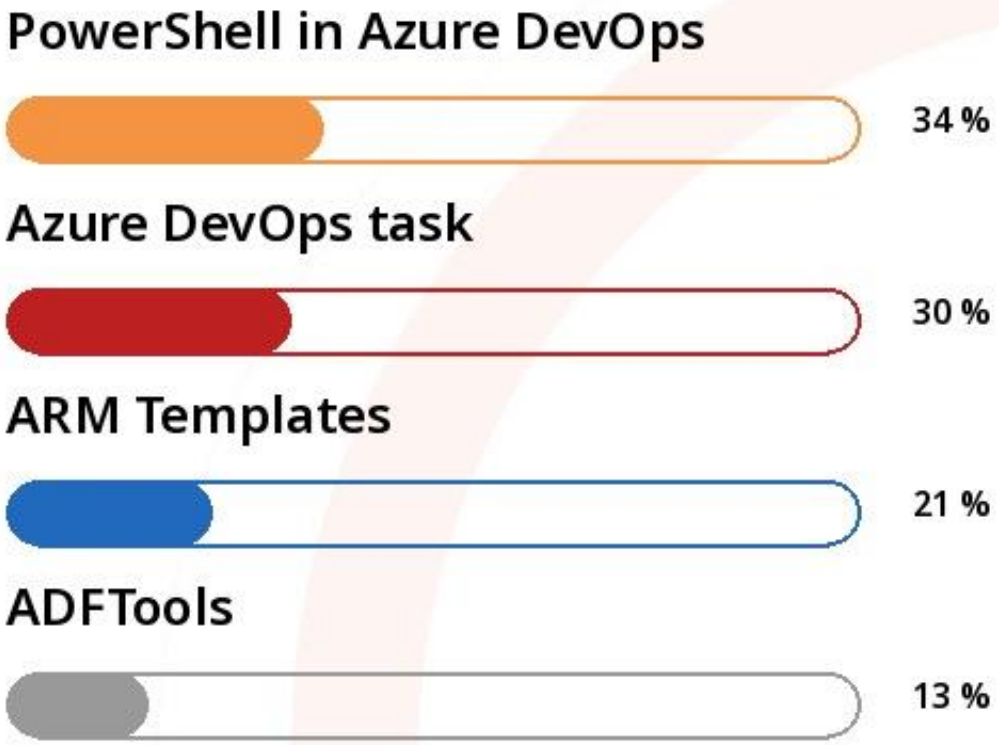
Kamil Data Geek - Azure explained
67 views • 2 months ago

www.AzurePlayer.net/YouTube

SQLDay 2024

Poll question

What do you prefer to use when deploying Azure Data Factory? (Choose 2)



My response

ADFTools

AGENDA

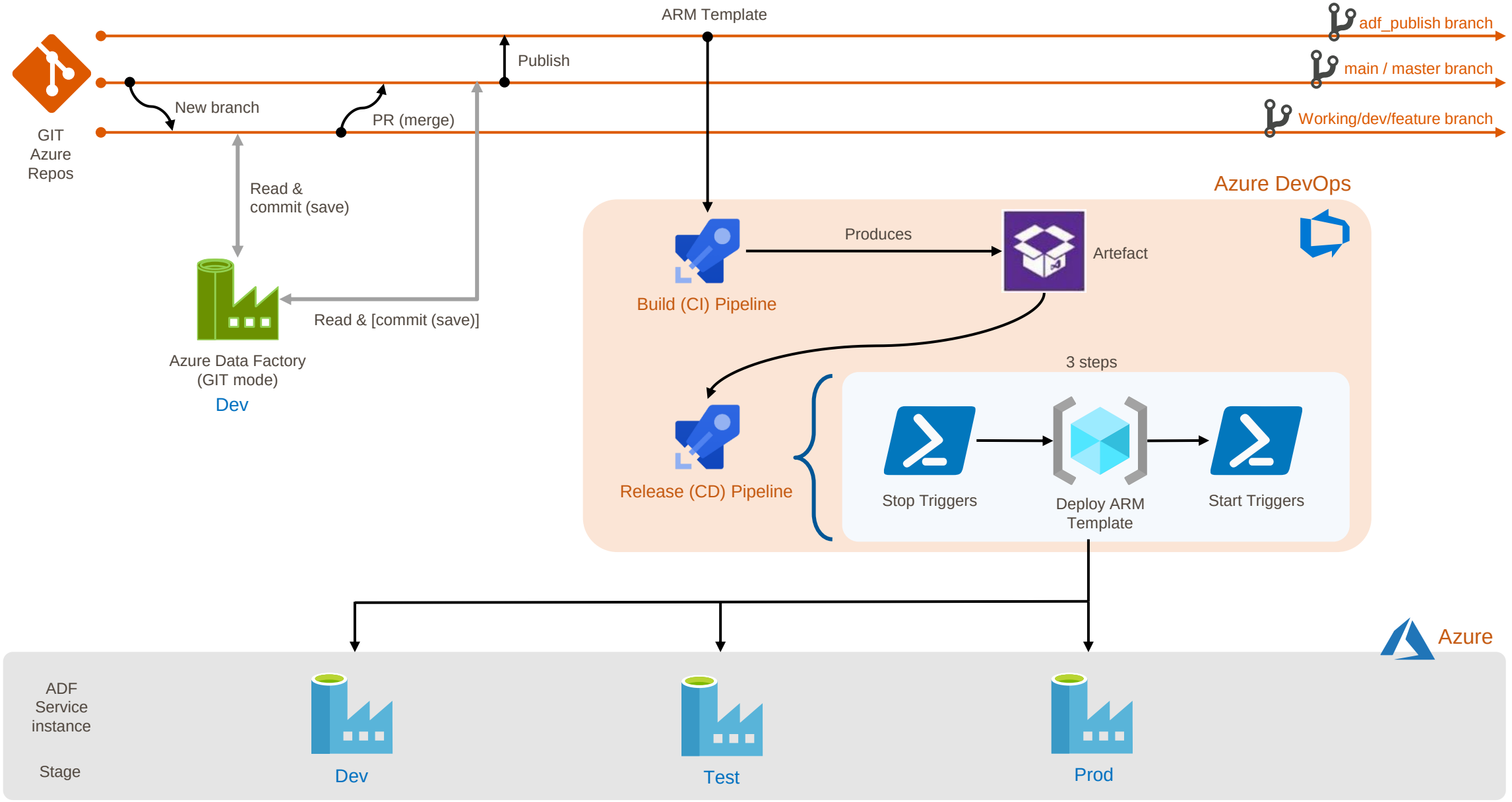
Azure Player
Play with data & have fun!

- 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!

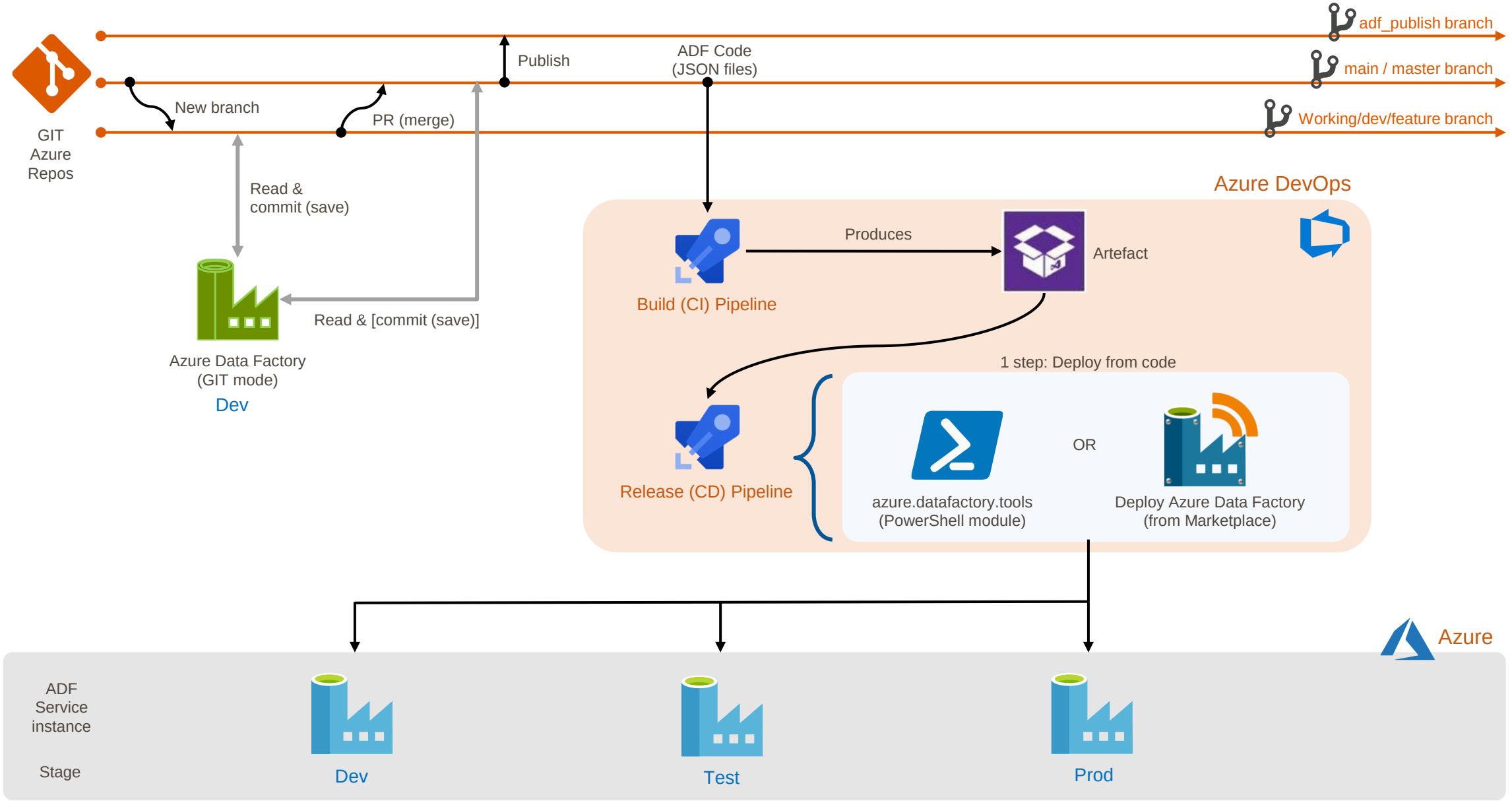
Slides available:

https://azureplayer.net/slides

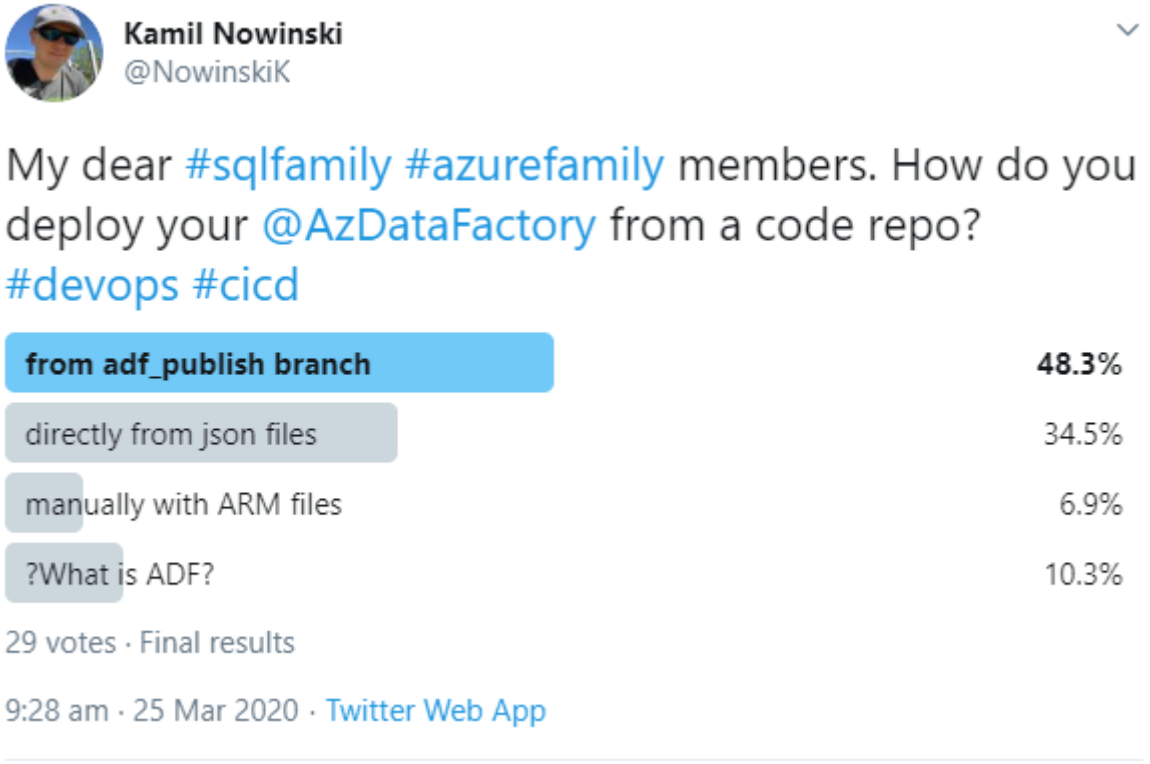
Deployment (1): Microsoft's method (ARM Template)



Deployment (2): Directly from code method



Who use what?



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

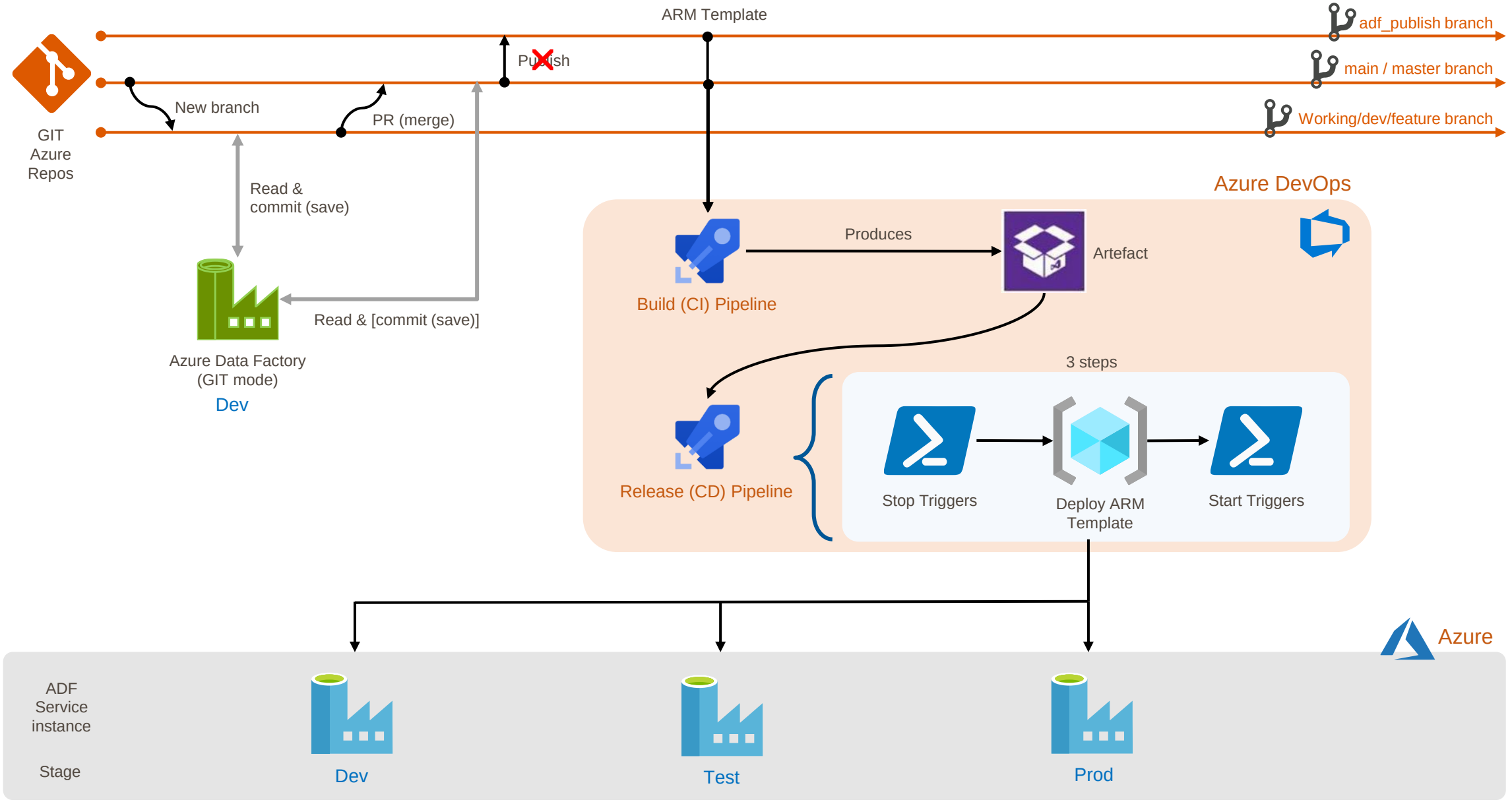
Rest-Api/PowerShell script from code (JSON objects)

- ~~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

Azure Player
Play with data & have fun!



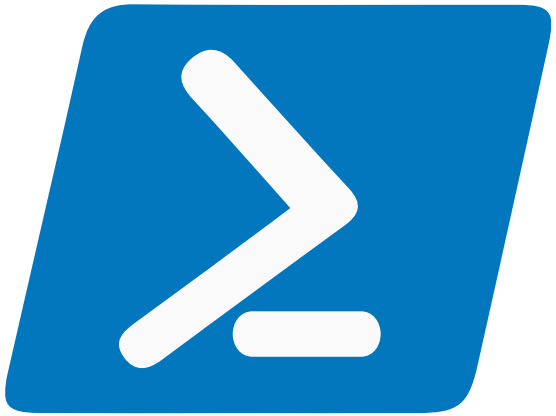
SQL Day

DEMO

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

Azure Player
Play with data & have fun!

What do you prefer?



SQL Day

DEMO #1

ARM Template deployment

Azure Player
Play with data & have fun!

Azure Player

#adftools

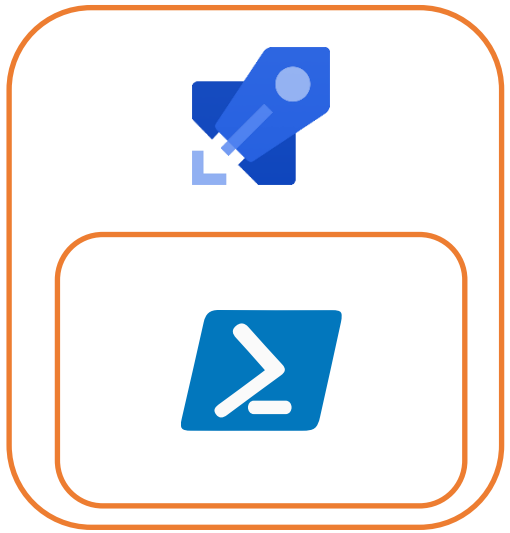
#adftools

Azure Player
Play with data & have fun!

Two tools:

PowerShell module (azure.datafactory.tools)

Azure DevOps extension (3 tasks)



https://azureplayer.net/adftools/

SQL Day

PowerShell Module

azure.datafactory.tools

Azure Player
Play with data & have fun!

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, Managed VNET
  - 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
  - Incremental Deployment (new!)
  - Generates mermaid dependencies diagram

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

SQL Day

DEMO #2

Deployment with PowerShell

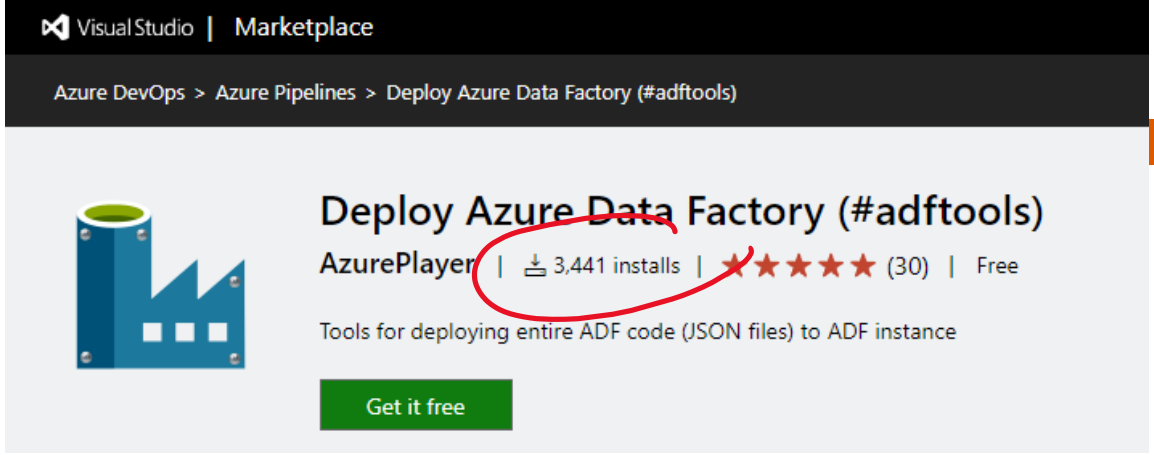
Azure Player
Play with data & have fun!

SQL Day

Azure DevOps Extension

Azure Player
Play with data & have fun!

Azure DevOps: Custom Tasks



Deploy Azure Data Factory

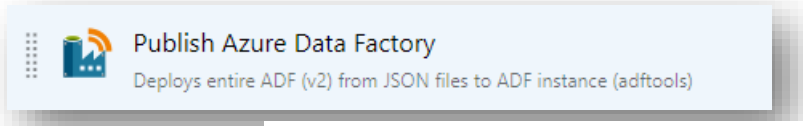
Key concept:

- 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

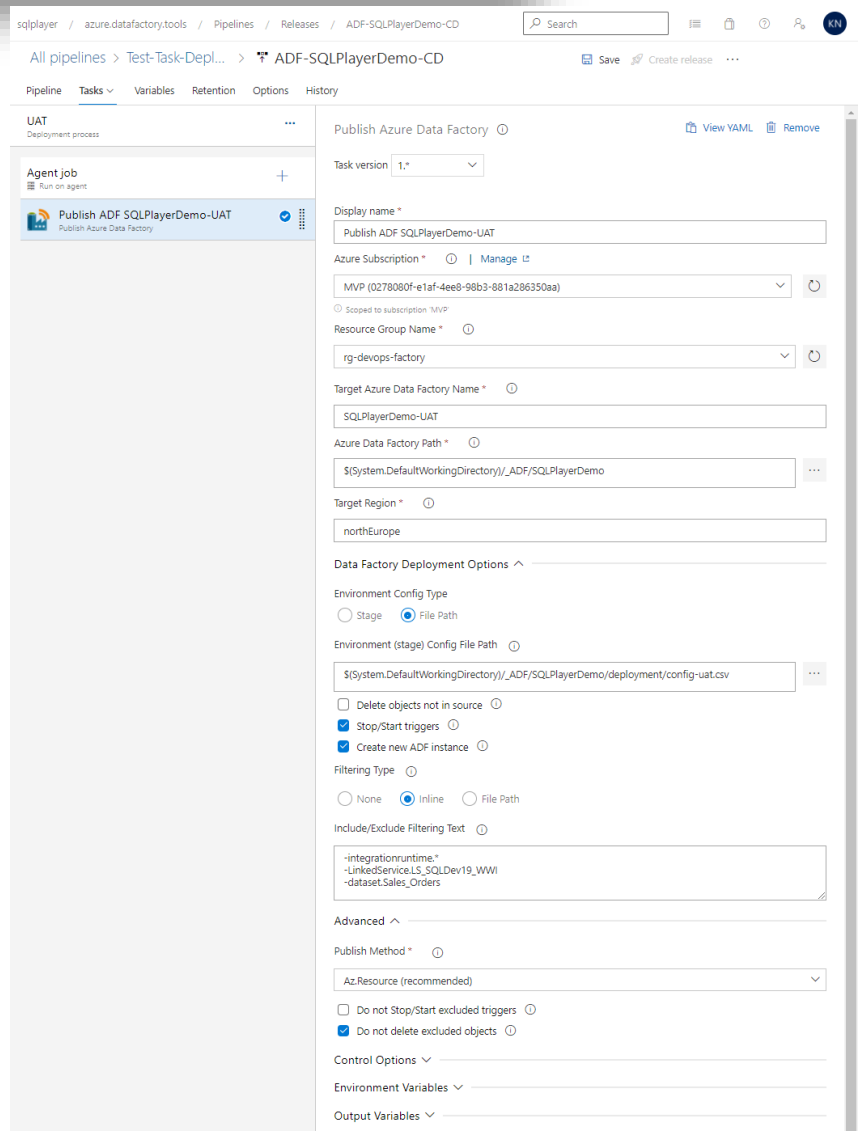
Task in Azure DevOps: Publish ADF (CD)

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
- Incremental Deployment



Selective deployment with PowerShell module

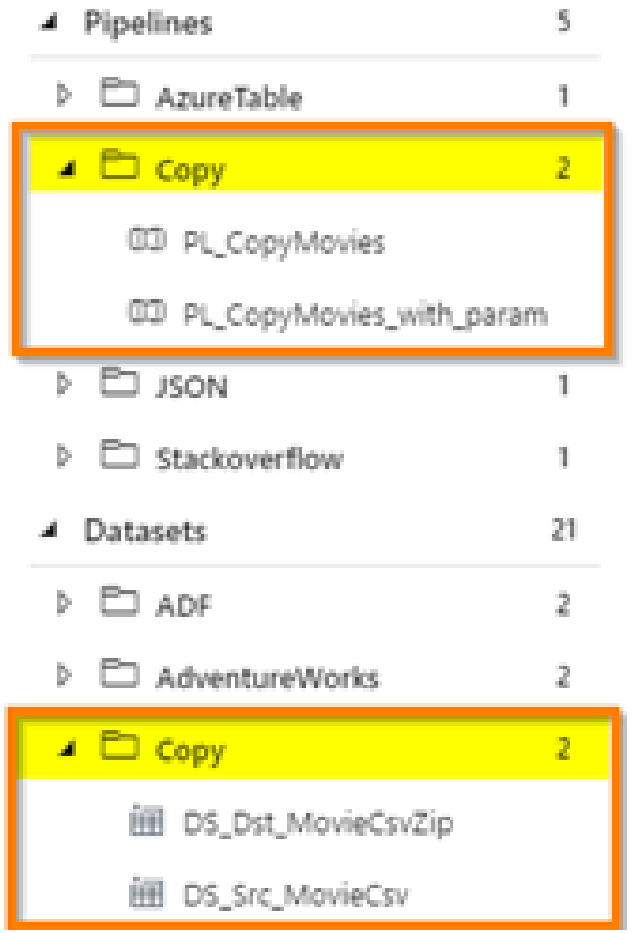
Azure Player
Play with data & have fun!

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]
*.*@Team1
managedVirtualNetwork*.*
*managedPrivateEndpoint.*
```



SQL Day

DEMO #3

Deployment with Azure DevOps

Azure Player
Play with data & have fun!

SQL Day

Bonus capabilities

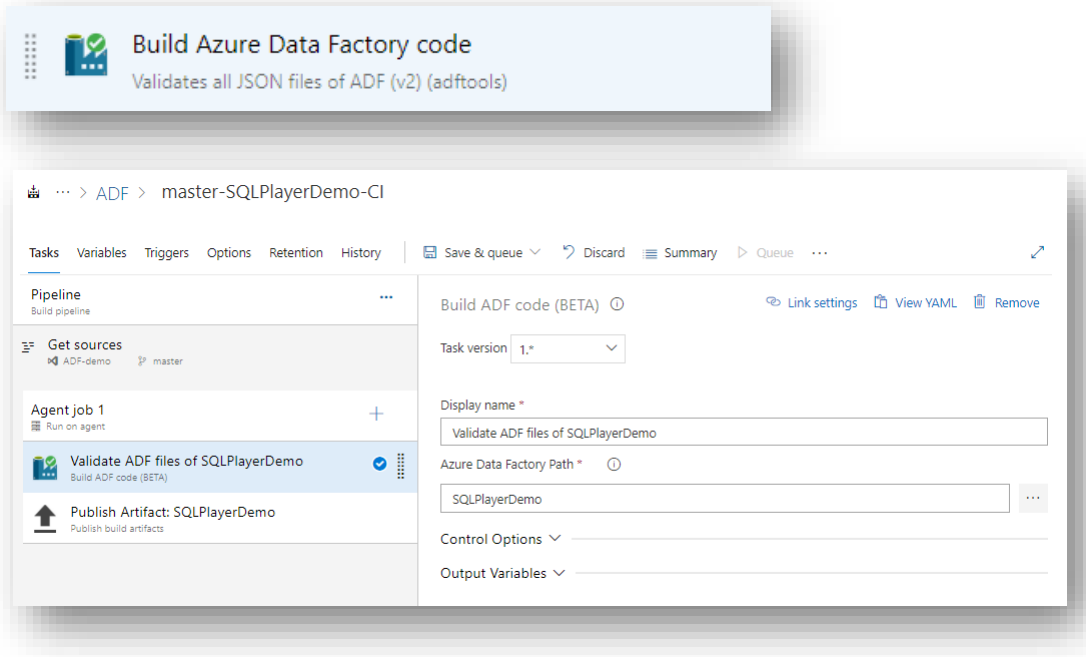
New task in Azure DevOps: Build ADF (CI)

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

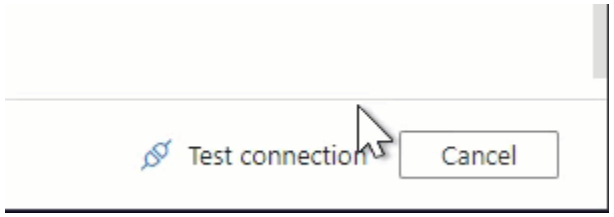
- **Validate & Export ARM Template**
  - uses ADFUtilities NPM package provided by Microsoft
  - Counterpart of Validate all and Export ARM Template in ADF UI

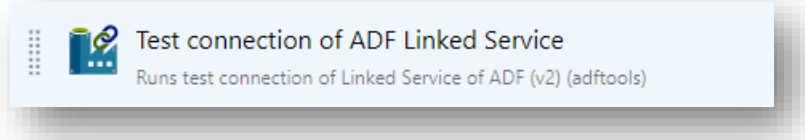


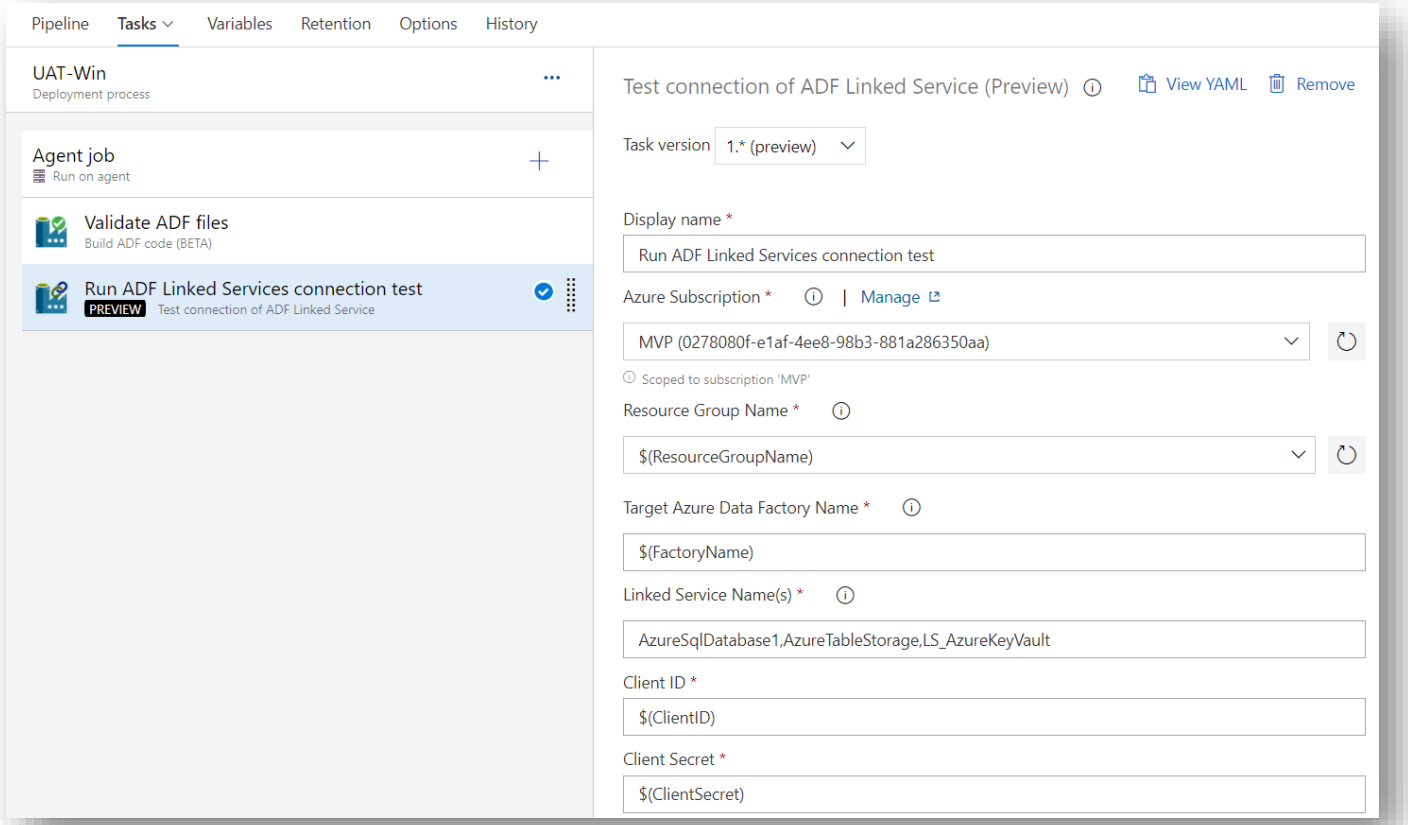
New task in Azure DevOps: Test Connection

Azure Player
Play with data & have fun!

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







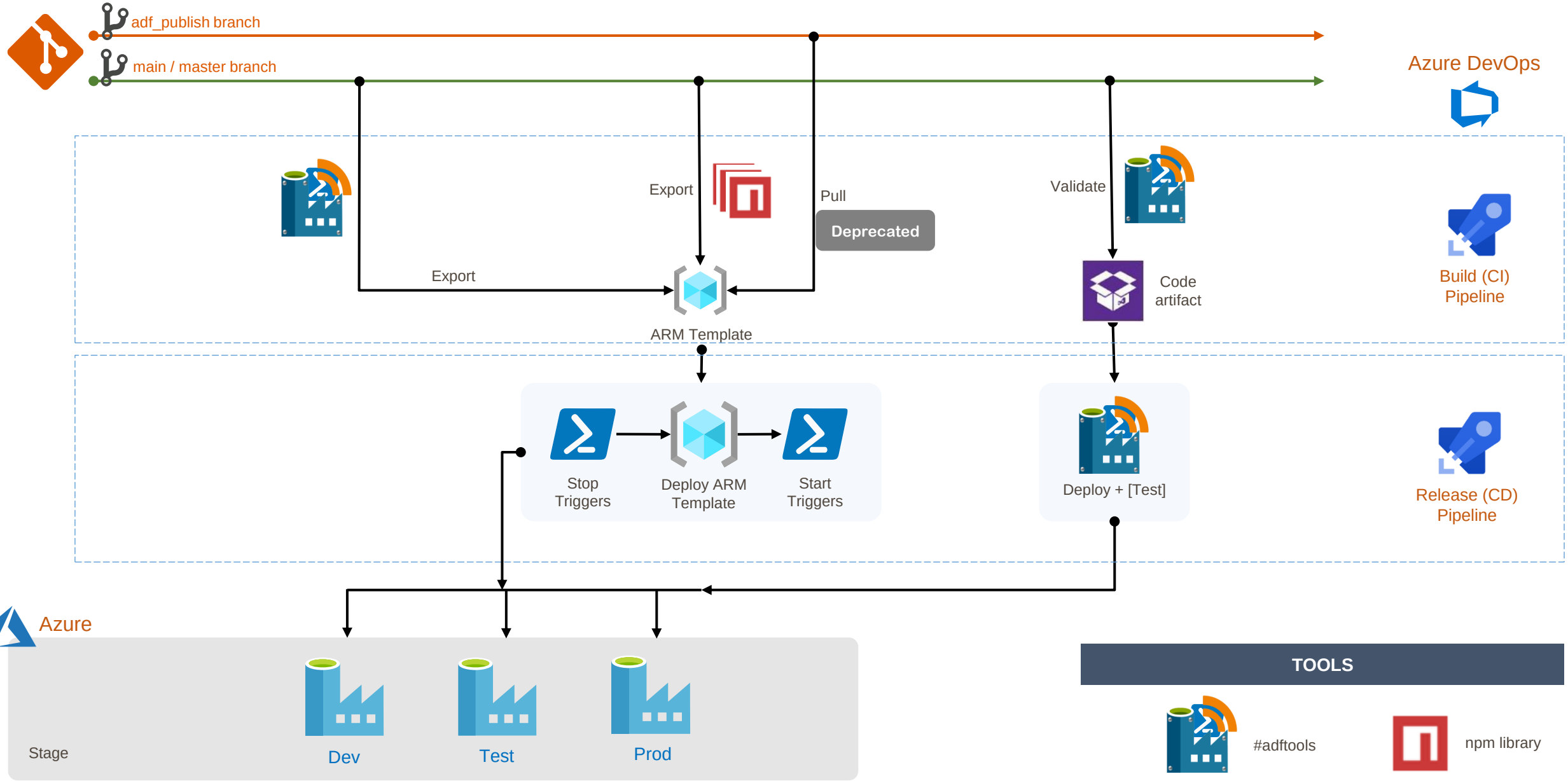
SQL Day

Summary

Azure Player
Play with data & have fun!

Takeaway: ADF deployment - Possible paths

Azure Player
Play with data & have fun!



#Adftools - Resources

- 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

SQL Day

16 edycja konferencji SQLDay

Data Community

13-15 maja 2024, WROCŁAW + ONLINE

partner platynowy

lingaro

partner złoty

DXC TECHNOLOGY

VOLVO

software one

partner srebrny

accenture

Nordcloud an IBM Company

Capgemini

TECHNOLOGY INNOVATION DATA KNOWLEDGE hdk

elitmind think bright

C&F