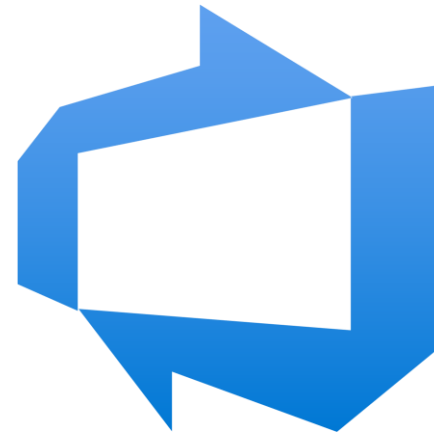# KevOps for Azure Databricks

Azure Databricks

Azure DevOps
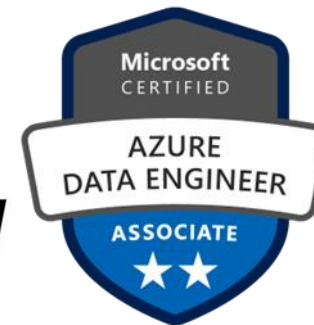
Kevin Chant

Data Saturday Rheinland

Agenda

- Advantages of Unity Catalog
- Introduction to Azure DevOps
- Databricks Repos
- CI/CD for notebooks
- Keep Databricks secrets secret
- Automating job information

Kevin Chant

- Data Engineering Manager in the Netherlands
- Worked in IT since Windows 95
- Experience in various sectors
- Various certifications, dual-category MVP

- Twitter: @kevchant
- LI: https://www.linkedin.com/in/kevin-chant/
- Blog: https://www.KevinRChant.com
- GitHub: https://github.com/kevchant

Advantages of Unity Catalog

Unifies workspaces

Easy admin

Credential Passthru options

Enable Identity Federation

Add workspace

Databricks Runtime 11.3 LTS & above

Azure DevOps overview

Collection of services

Azure Boards for work items

Azure Repos to store code

Azure Pipelines to deploy things

Azure Test Plans to manage tests

Azure Artifacts to store artifacts

CI/CD for notebooks
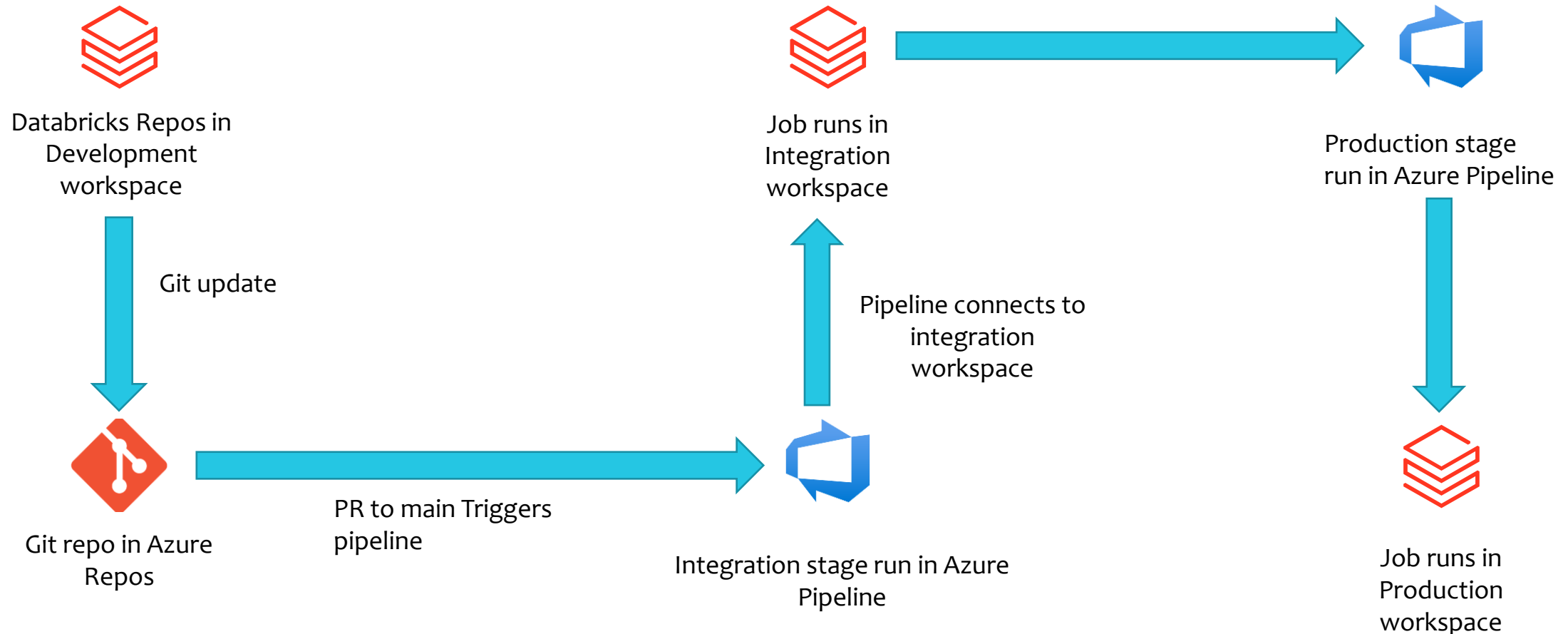
Databricks preferred method

Microsoft preferred method

Databricks method

Works directly with Databricks Repos

No actual build

Example based on YAML

Databricks method diagram

Databricks Repos in Development workspace

Git update

Git repo in Azure Repos

PR to main Triggers pipeline

Integration stage run in Azure Pipeline

Pipeline connects to integration workspace

Job runs in Integration workspace

Production stage run in Azure Pipeline

Job runs in Production workspace

Microsoft method

Whl (wheel) file method

More complex

Encourages IDE

Microsoft method diagram

Dev Databricks Repos using Dev cluster

Git request

Python Wheel file published in Azure Artifacts service.

Unit Tests passed

Integration stage run in Azure Pipelines

Connects via Databricks token

Job runs in Production workspace

Connects via Databricks token

Integration tests OK

Git repo in Azure DevOps

Triggers Build pipeline

Performs defined unit tests and if successful builds wheel

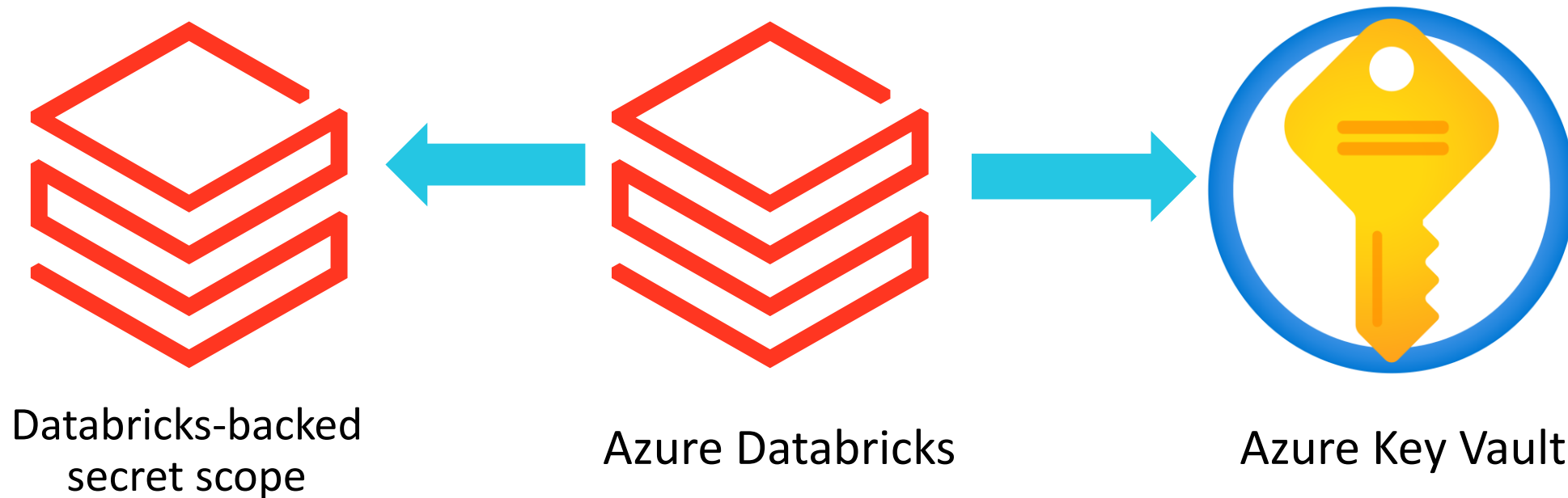Job runs in Integration workspace

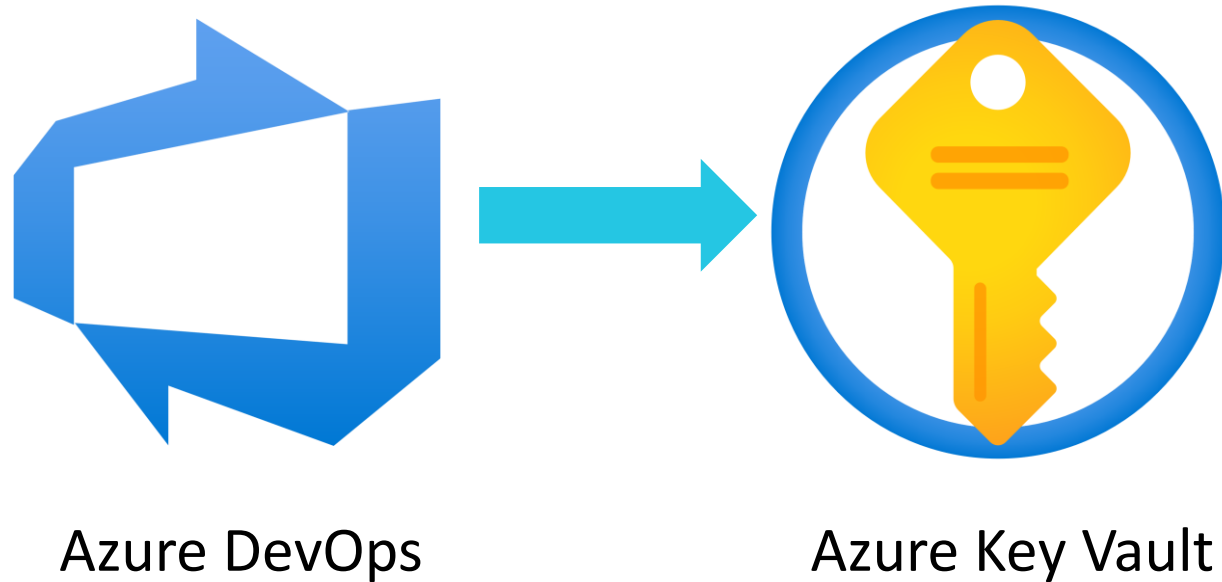Azure Pipelines get Production Credentials

Other options

Hybrid method

Multiple organizations

Keep working with notebook connections

Keeping Databricks secrets secret

Databricks-backed
secret scope

Azure Databricks

Azure Key Vault

Keeping Databricks secrets secret

Azure DevOps   Azure Key Vault

Automating job information

Extract JSON required

API method

Automate extraction via Azure Pipelines

Questions

Thank you

- Twitter: @kevchant

- LI: https://www.linkedin.com/in/kevin-chant/

- Blog: https://www.KevinRChant.com

- GitHub: https://github.com/kevchant