



Module 9: Data Factory in Production

- Testing
- Bootstrapping
- Best Practices

Testing

A green arrow originates from the word 'Testing' and points horizontally towards the word 'Quality'. The arrow is a solid green color and tapers slightly as it moves to the right.

Quality

Types of Testing

📖 **Integration Tests** - A test of a pipeline as-is, without eliminating any effects of external dependencies.

📖 **Functional Tests** - An isolated test of whether the pipeline is doing things right – is the pipeline producing the desired result?

📖 **Unit Tests** - An isolated test of whether the pipeline is doing the right things – do the pipeline's activities get executed in the way you expect?

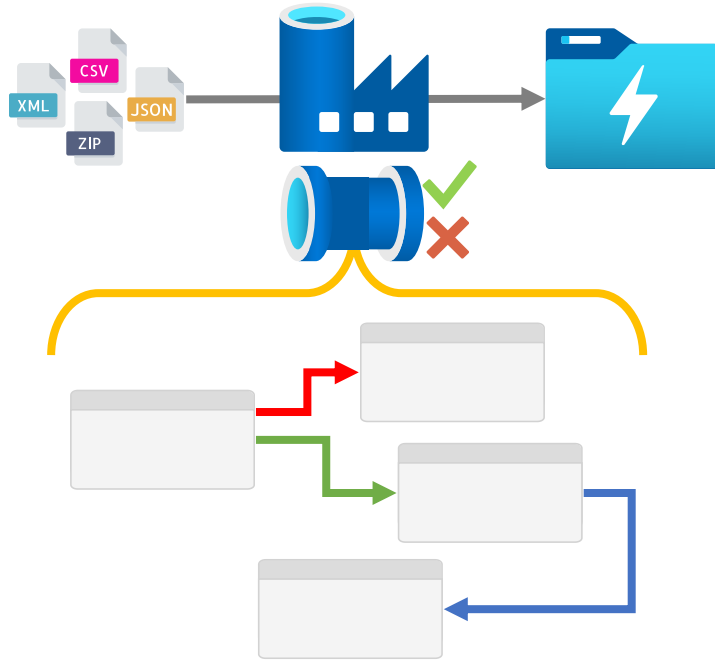
Source: Richard Swinbank richardswinbank.net



Types of Testing – What do they mean in our pipelines?

Integration Tests

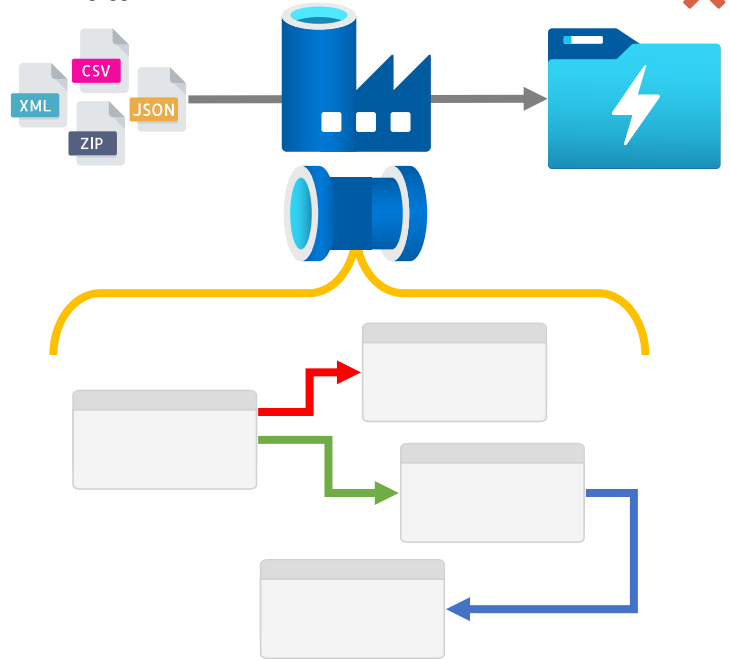
Simple Run Outcome



Functional Tests

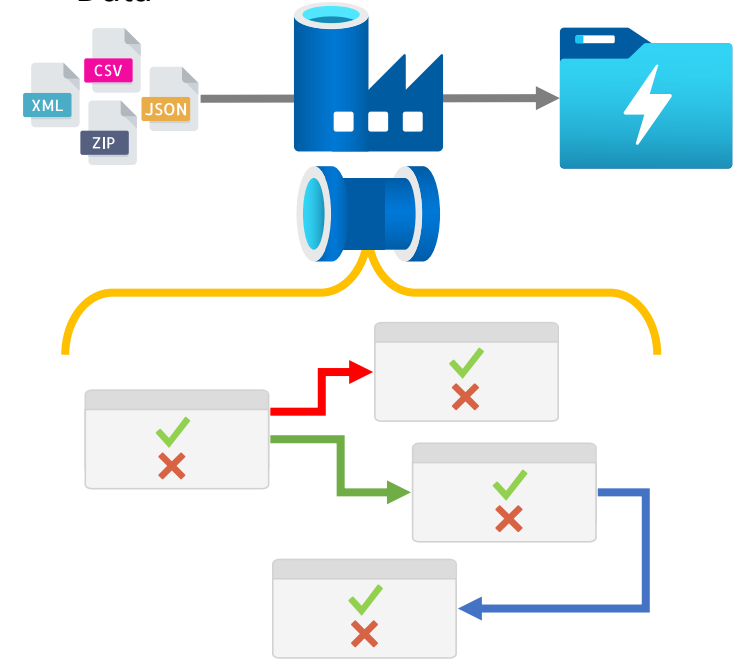
Sample Data

Specific Results



Unit Tests

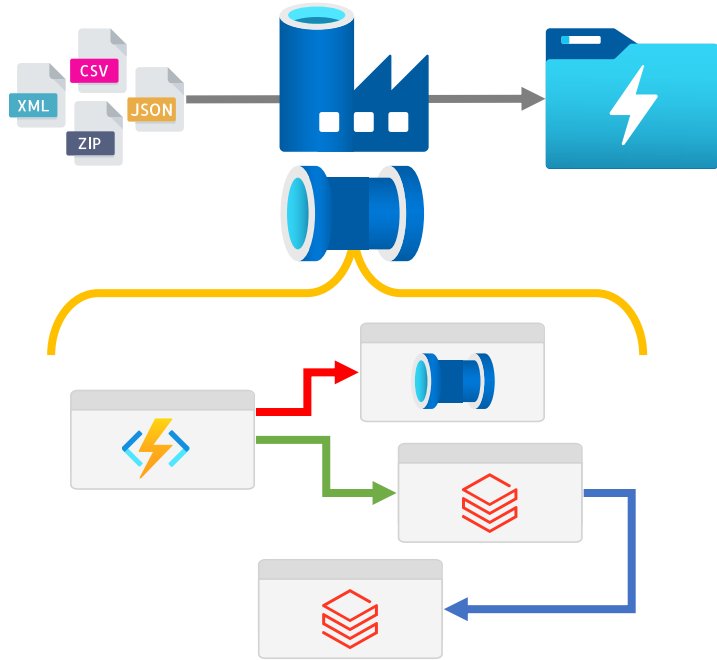
Sample Data



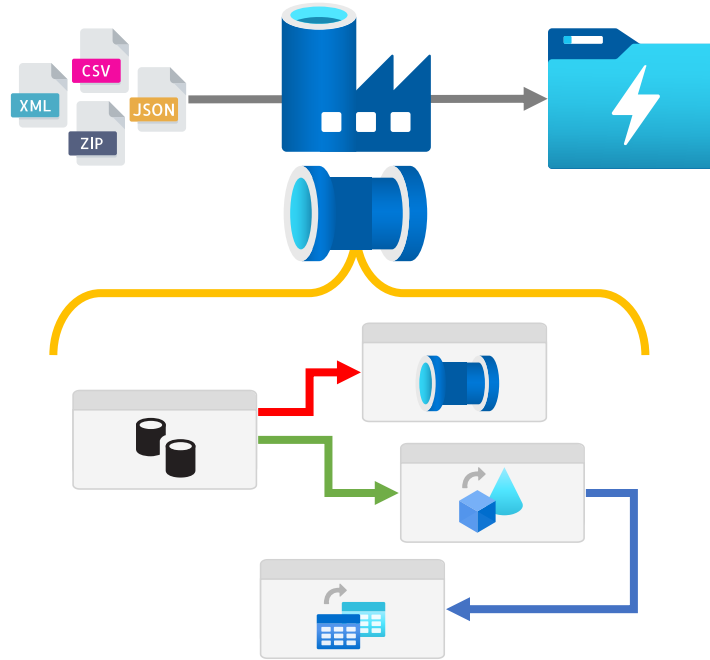
Specific Activity Inputs & Outputs

Types of Testing – Does it depend on external resources?

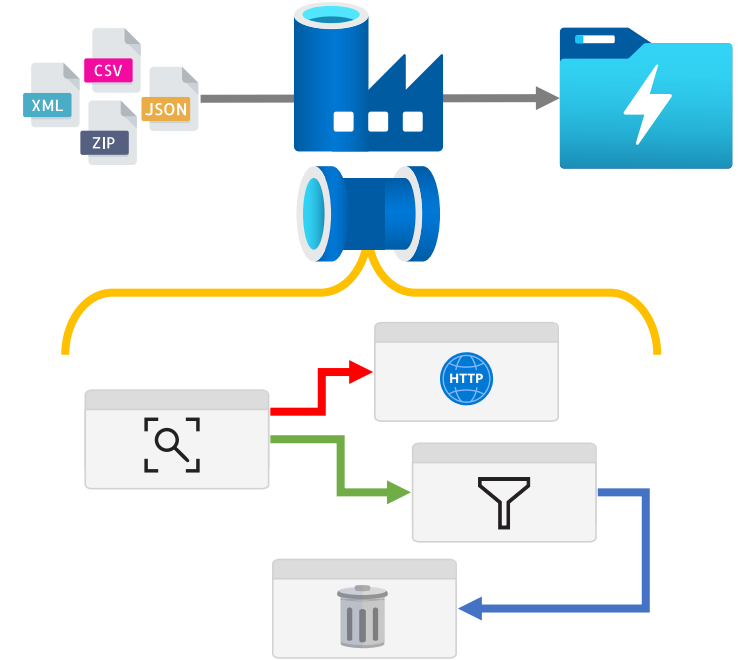
Integration Tests



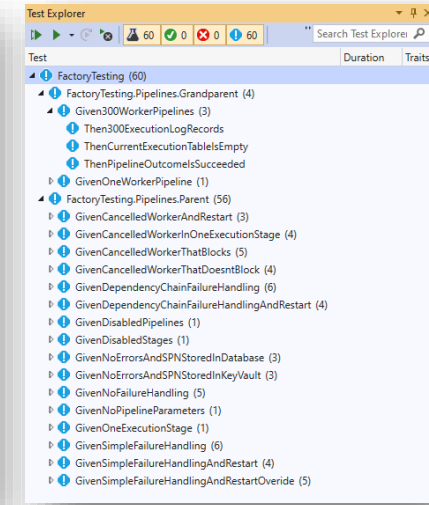
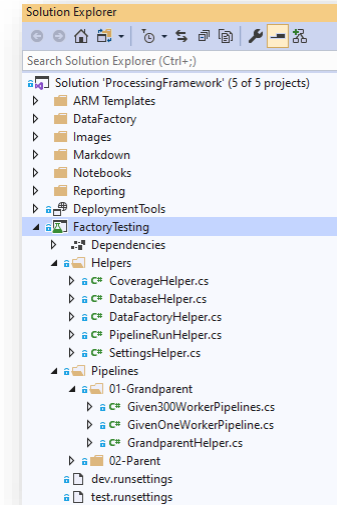
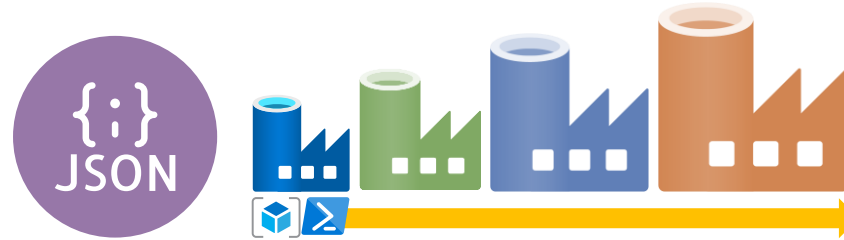
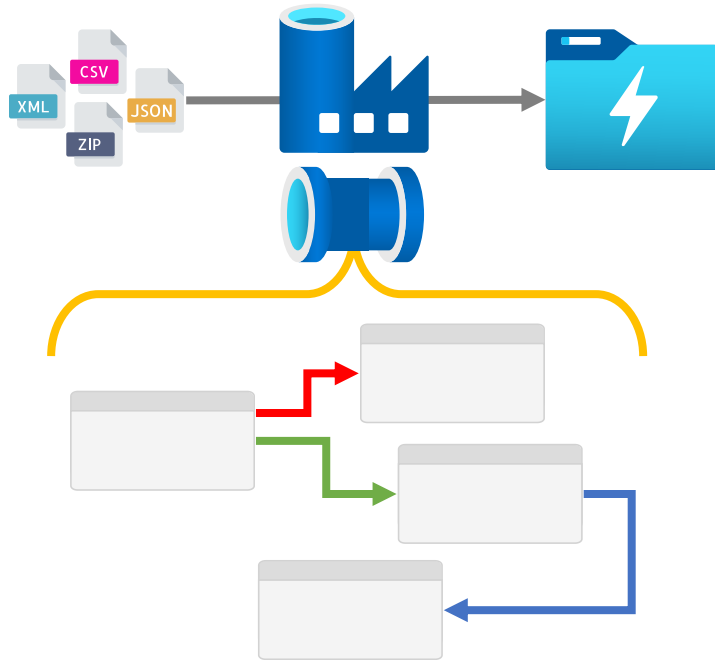
Functional Tests



Unit Tests

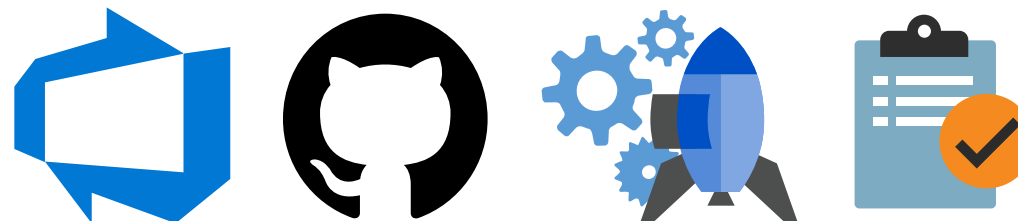


How To Run & Automate Tests



github.com/mrpaulandrew/ADF.procfwk/tree/master/FactoryTesting

github.com/richardswinbank/community/tree/main/adf-testing-series

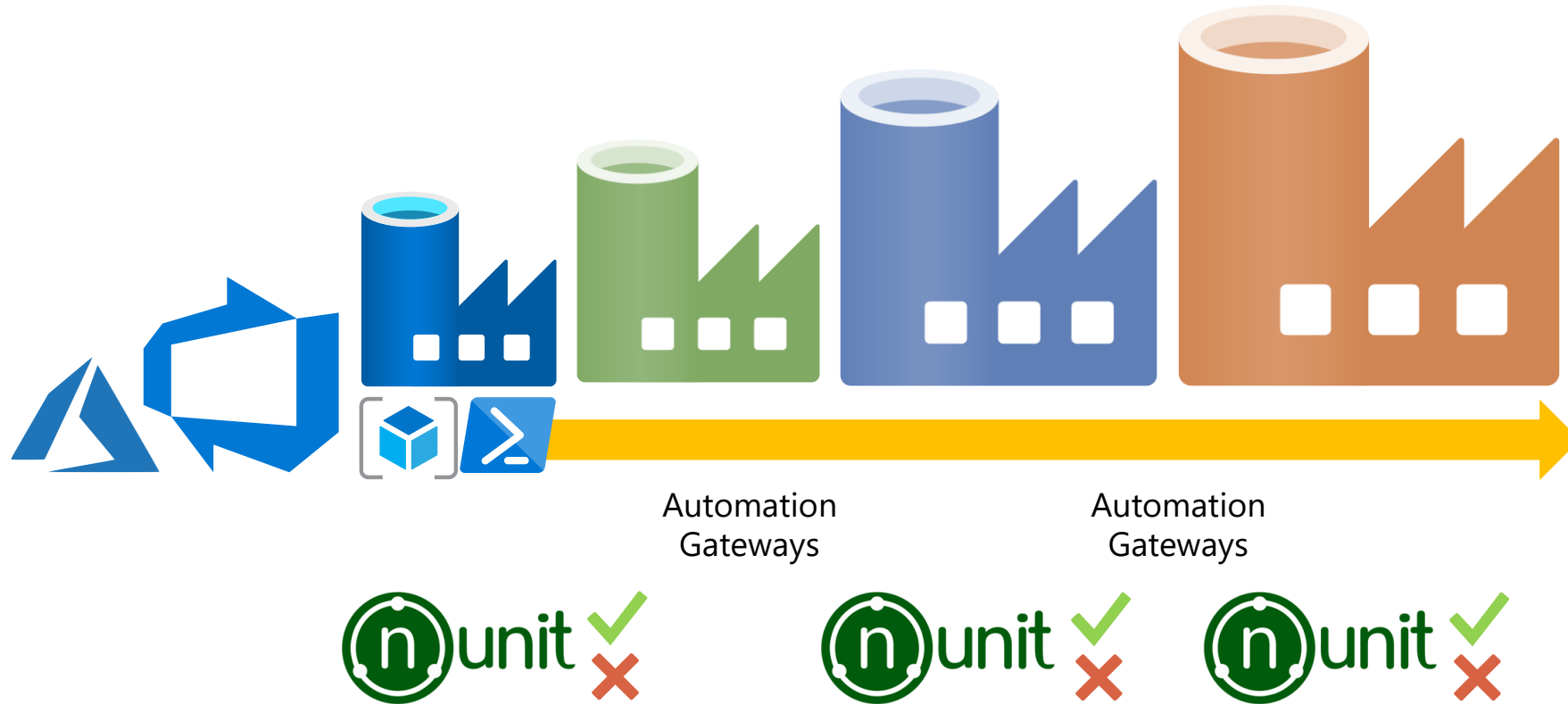


DEMO



Testing Code vs Actual Code

Data Factory – Chicken vs Egg

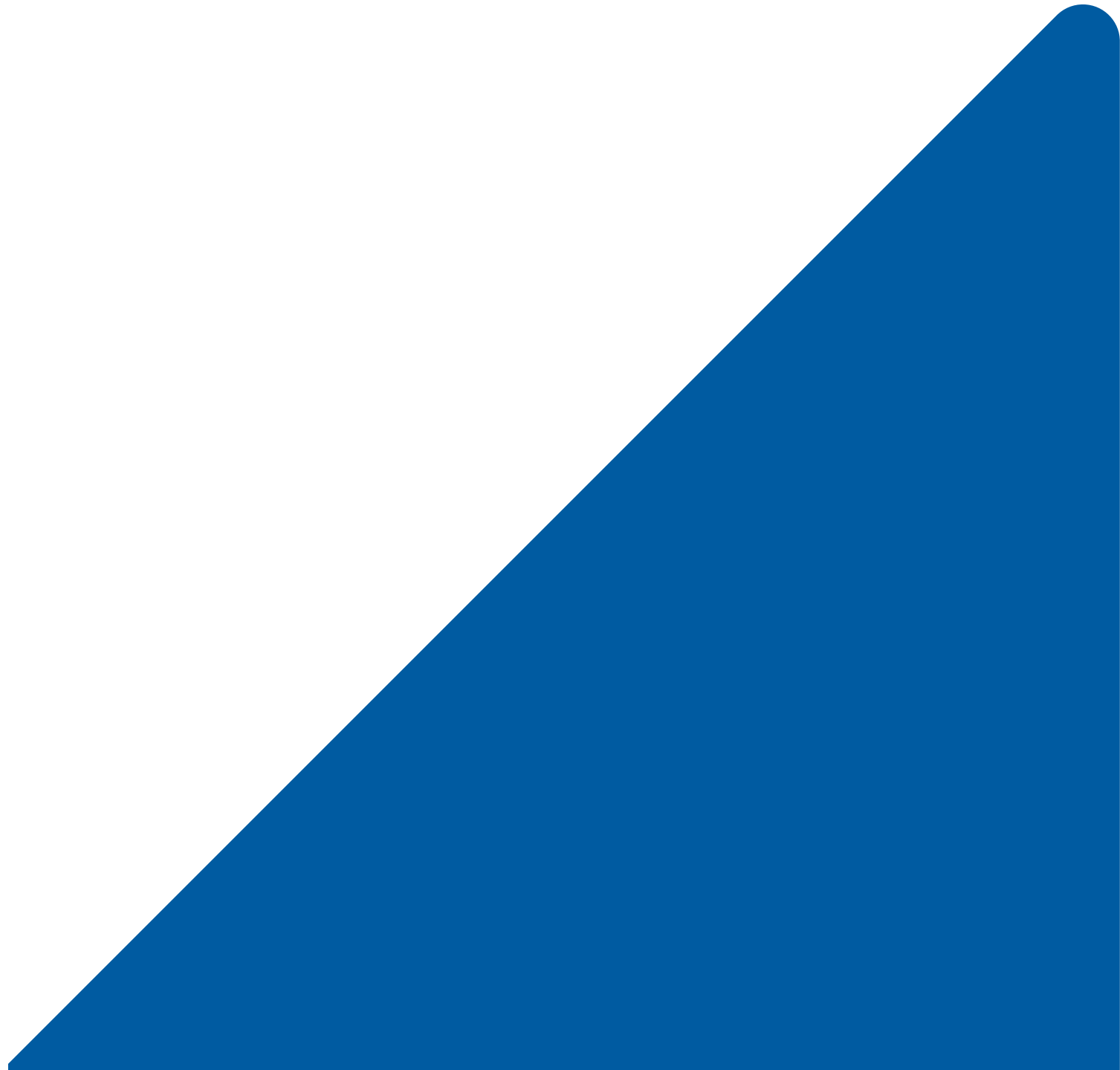


Should automate this?

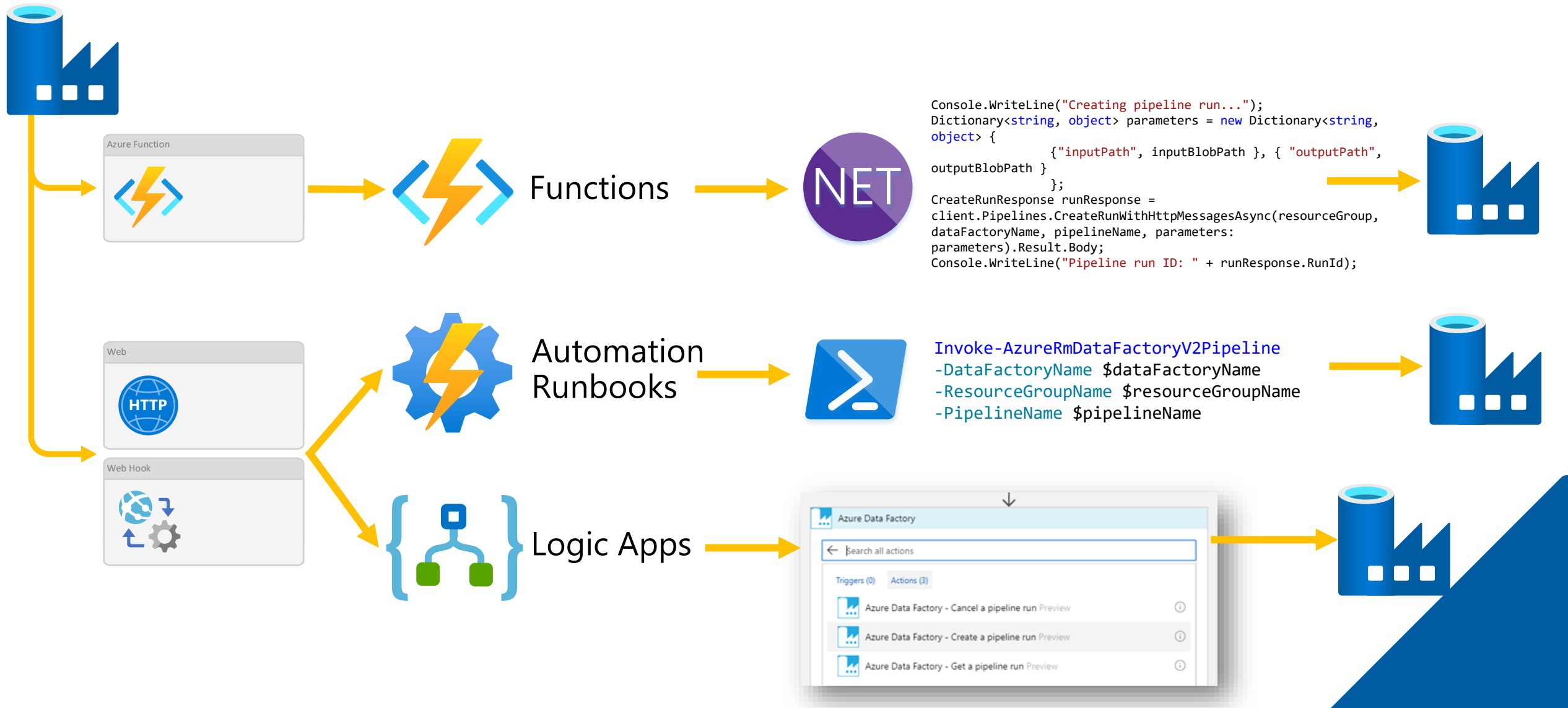
When should testing scripts run?

What if a test case fails?

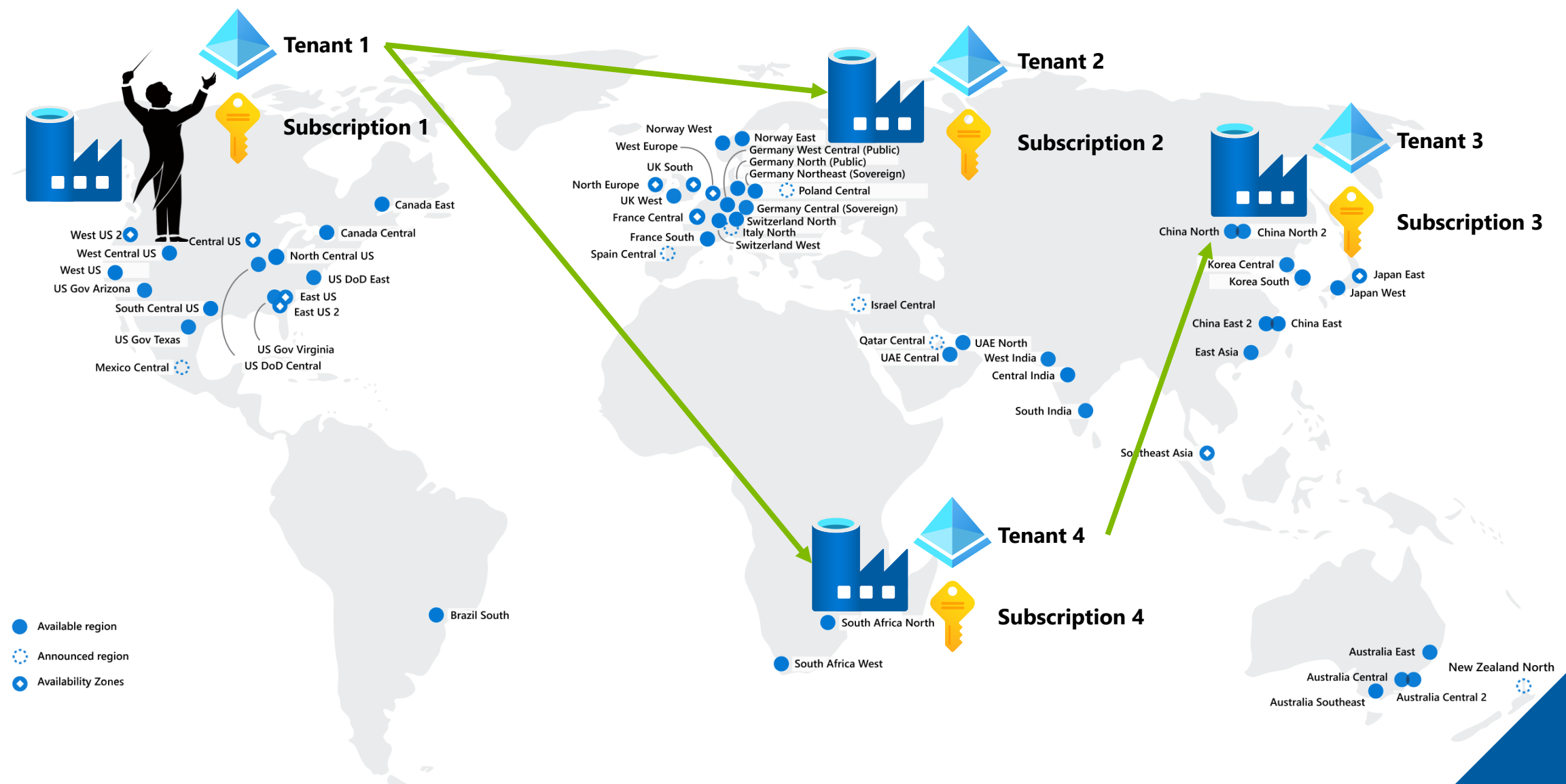
Bootstrapping



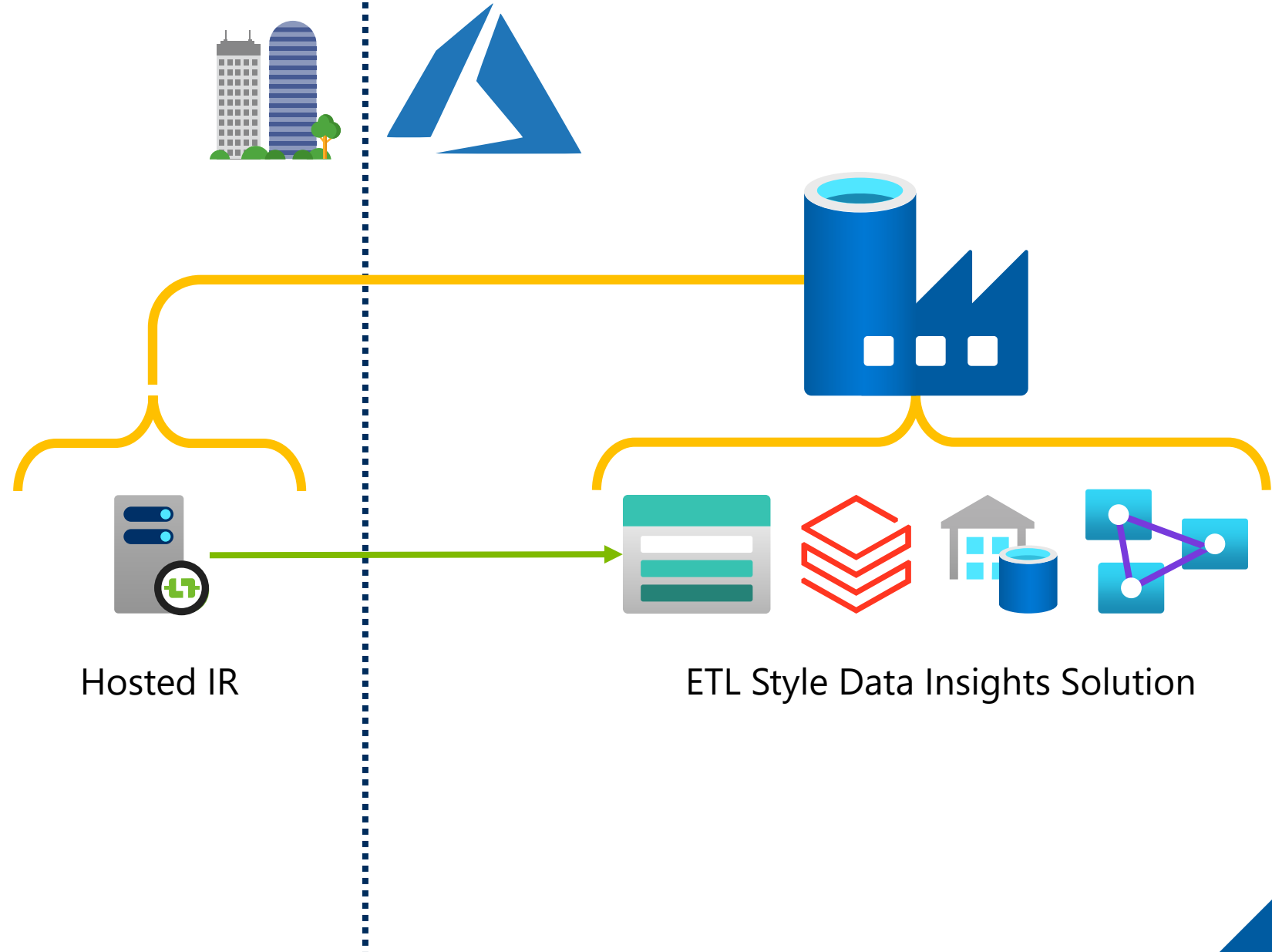
Bootstrapping – How?



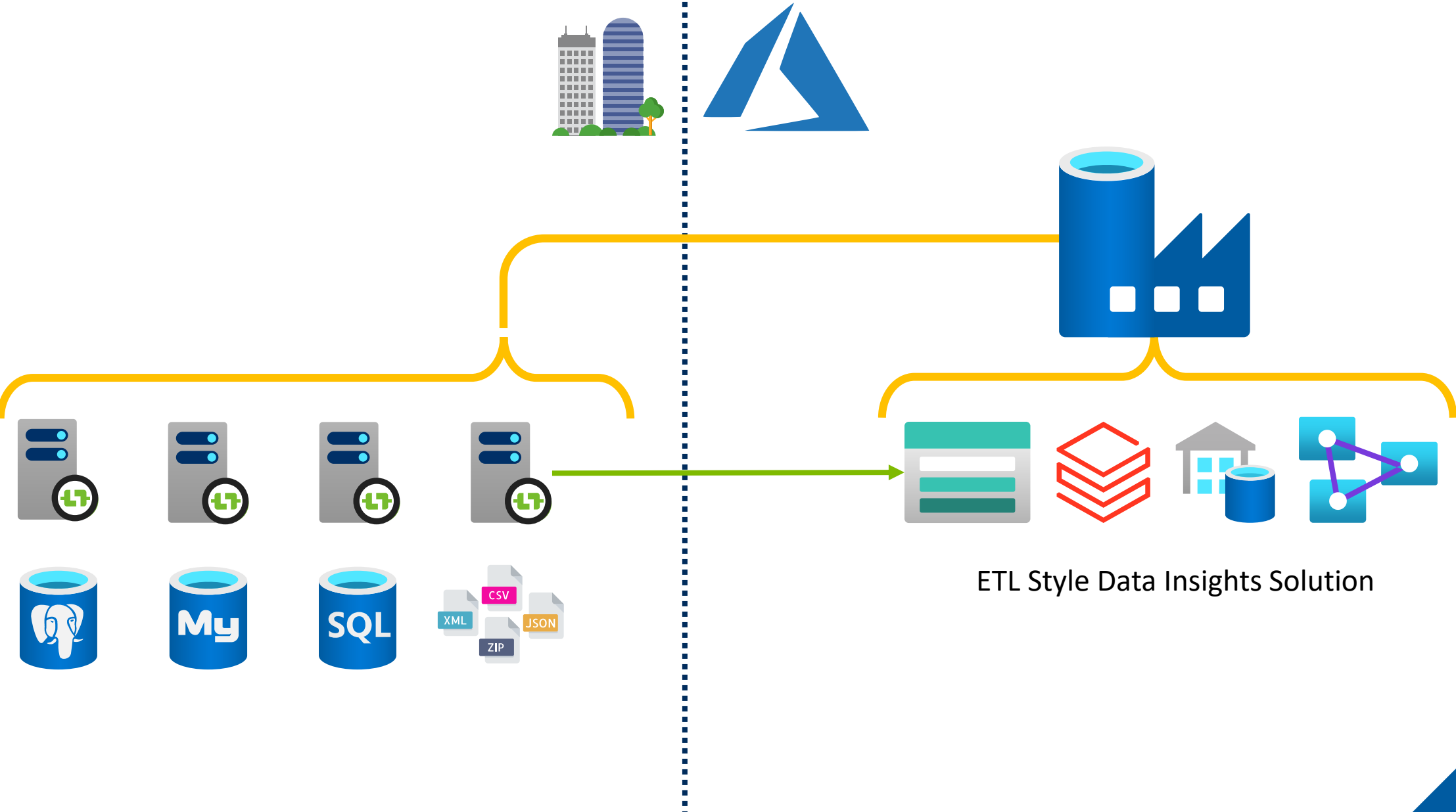
Bootstrapping – Why?



Bootstrapping – Why?



Bootstrapping – Why?



Best Practices

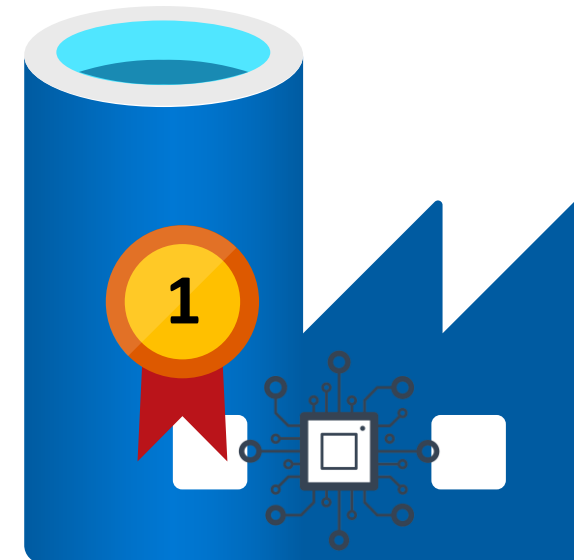


Key Points

- Environment Setup & Developer Debugging
- Deployments
- Automated Testing
- Naming Conventions
- Pipeline Hierarchies
- Pipeline & Activity Descriptions
- Factory Component Folders
- Linked Service Security via Azure Key Vault
- Dynamic Linked Services
- Generic Datasets
- Metadata Driven Processing
- Parallel Execution
- Hosted Integration Runtimes
- Azure Integration Runtimes
- Wider Platform Orchestration
- Custom Error Handler Paths
- Monitoring via Log Analytics
- Service Limitations
- Using Templates
- Documentation



mrpaulandrew.com/2019/12/18/best-practices-for-implementing-azure-data-factory/



Module 9:

Data Factory in Production

🔊 Testing ✓

🔊 Bootstrapping ✓

🔊 Best Practices ✓