

DATA() { EXPOSED;

DATA EXPOSED SPECIAL

Around the Clock with Azure SQL and Azure Data Factory

Americas

February 3, 2021

09:00 - 17:00 PT

Asia

February 4, 2021

09:00 - 17:00 SGT

16 Sessions | 2 Ask the Expert Panels | 1 Hackathon



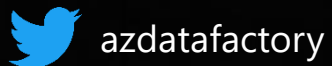
HOSTED BY

Wee Hyong Tok & Anna Hoffman

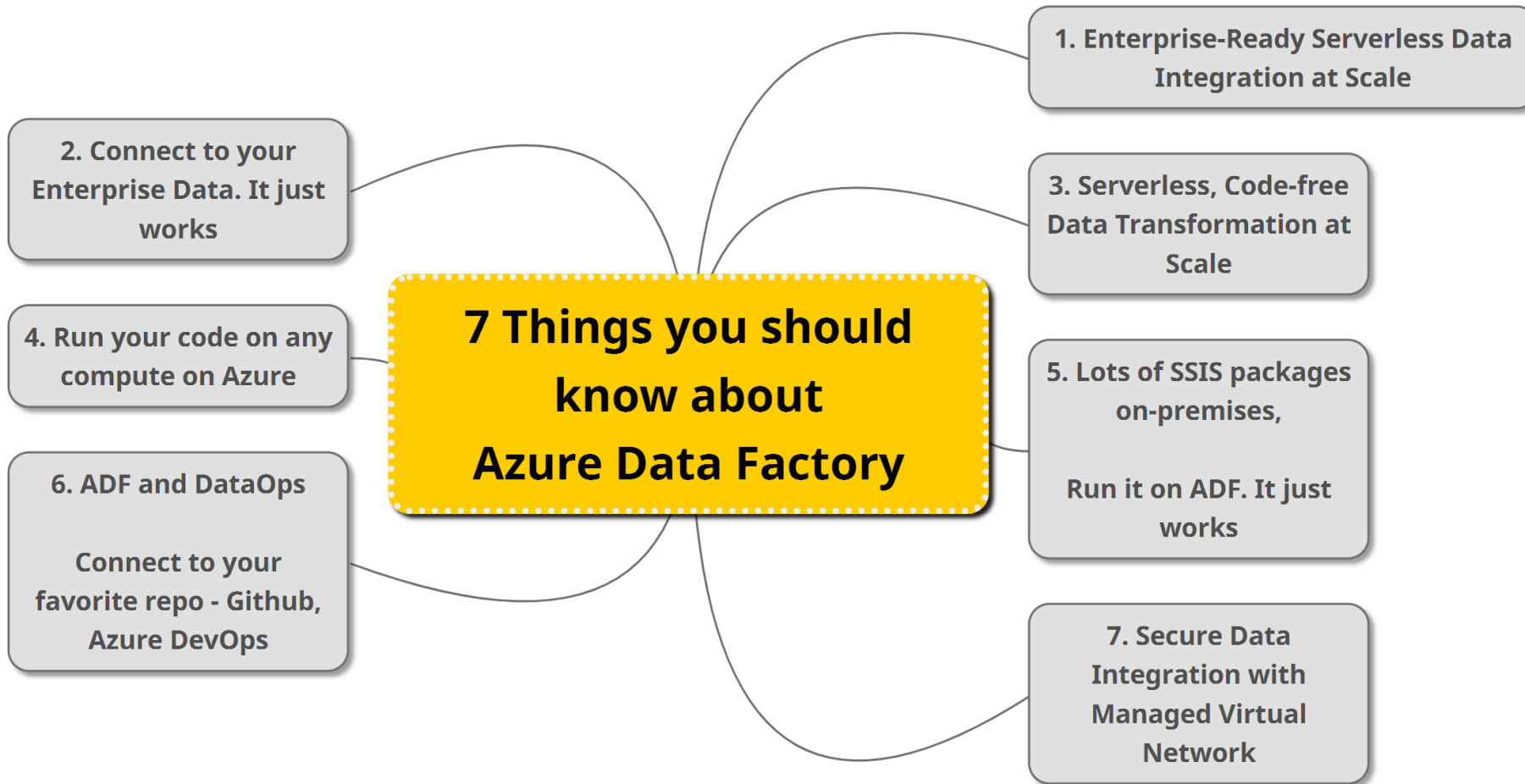
DATA() {
EXPOSED;

7 Things you should know about Azure Data Factory

Shirley Wang & Wee Hyong Tok
Group Product Manager
Azure Data Governance, Microsoft



Agenda

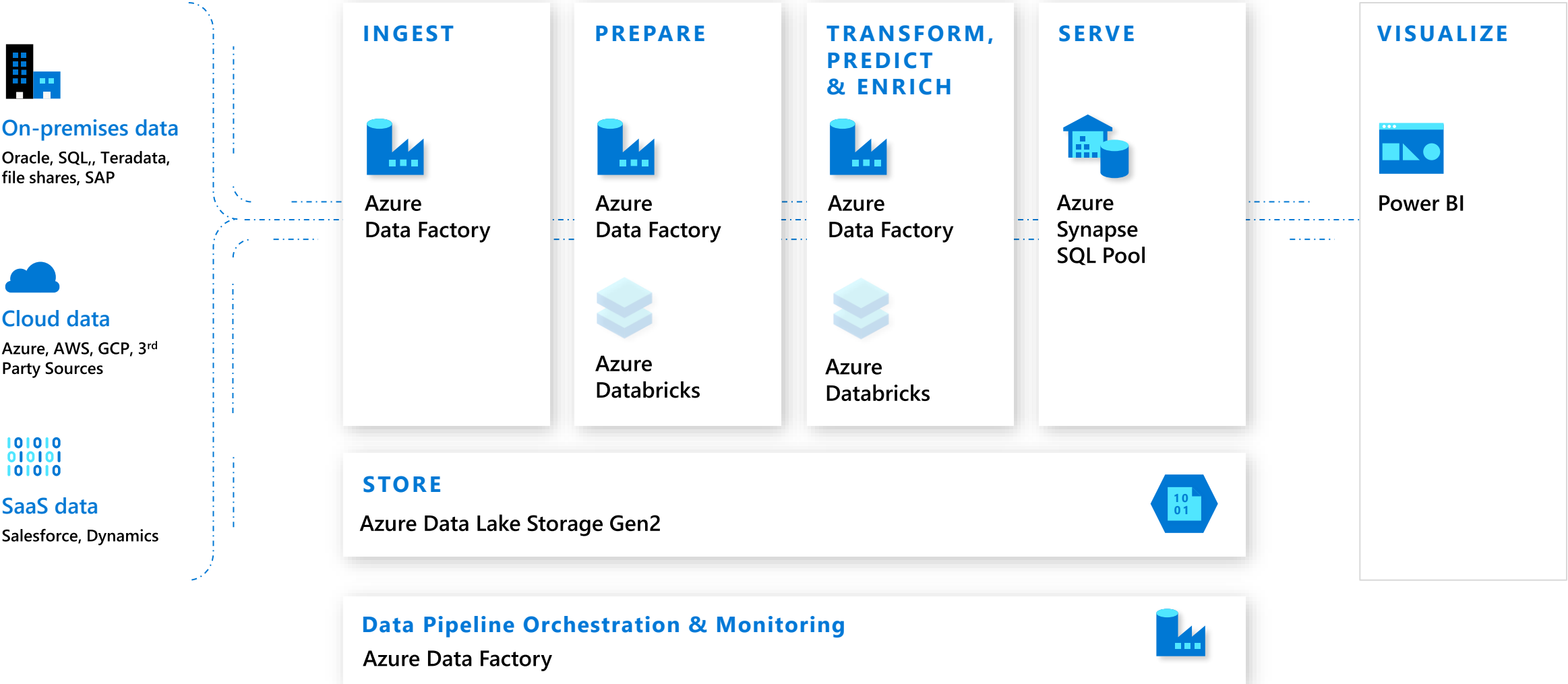




1

Enterprise-ready,
Serverless Data
Integration at Scale

Modern Data Warehouse (MDW)



Azure Data Factory

Enterprise Data integration, simplified



Easy-to-use

- Code-free ETL/ELT
- SSIS – it just works in ADF!
- Built-in Git & CI/CD



Cost-effective

- Pay-as-you-go
- Fully managed, Serverless
- Scales on demand



Powerful

- Connect to all the enterprise data sources
- Orchestrate and monitor at scale



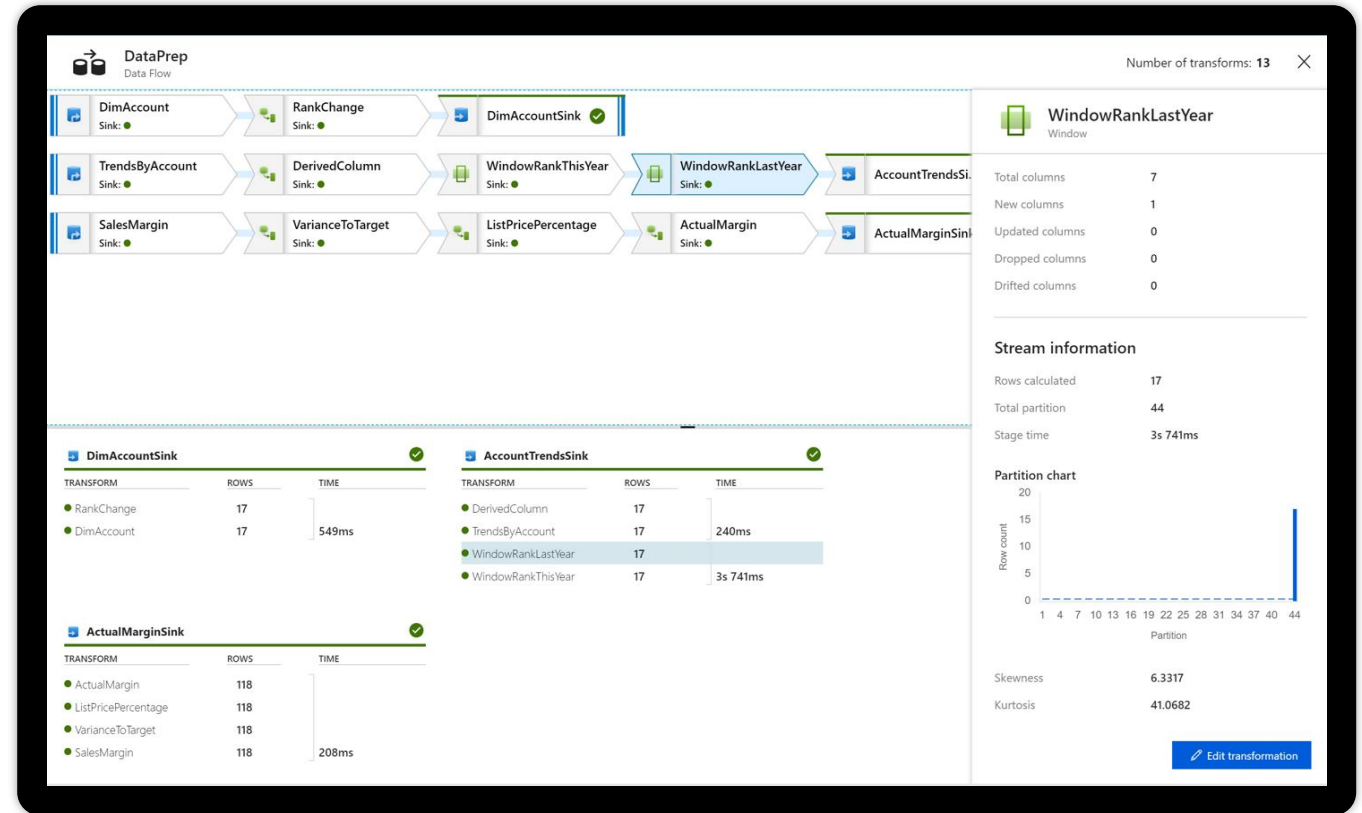
Intelligent

- Autonomous ETL
- AI based Copy
- Build Predictive pipelines

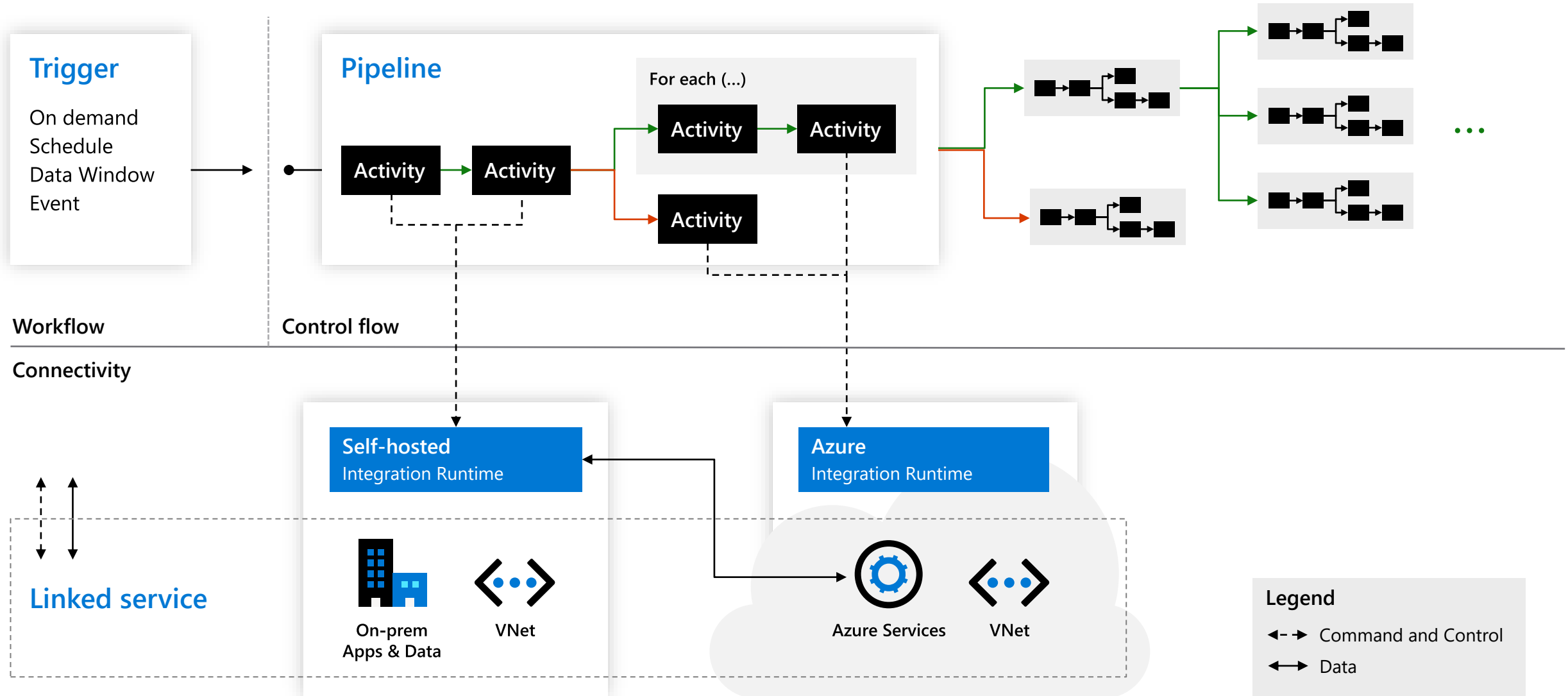
Monitor: Best-in-class management

Lower your cost of deployment, testing, and maintenance by monitoring pipeline and activity runs with pre-built templates and UI-based views

- Query Runs with rich language
- Reference operational lineage between parent-child pipelines
- Integrate Azure Monitor for diagnostics loggings, metrics and alerts, and events.
- Restate pipeline and activities



Orchestrate at scale



How does ADF help you Move Data at Global Scale?

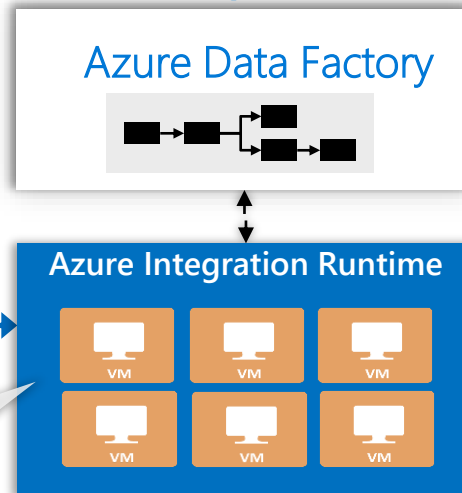
1 Copy Config Settings

Data integration unit ⓘ
☐ Edit
[Add dynamic content \[Alt+P\]](#)

Degree of copy parallelism ⓘ
☐ Edit

Configure Data Integration Units and Parallelism in code-free authoring UX

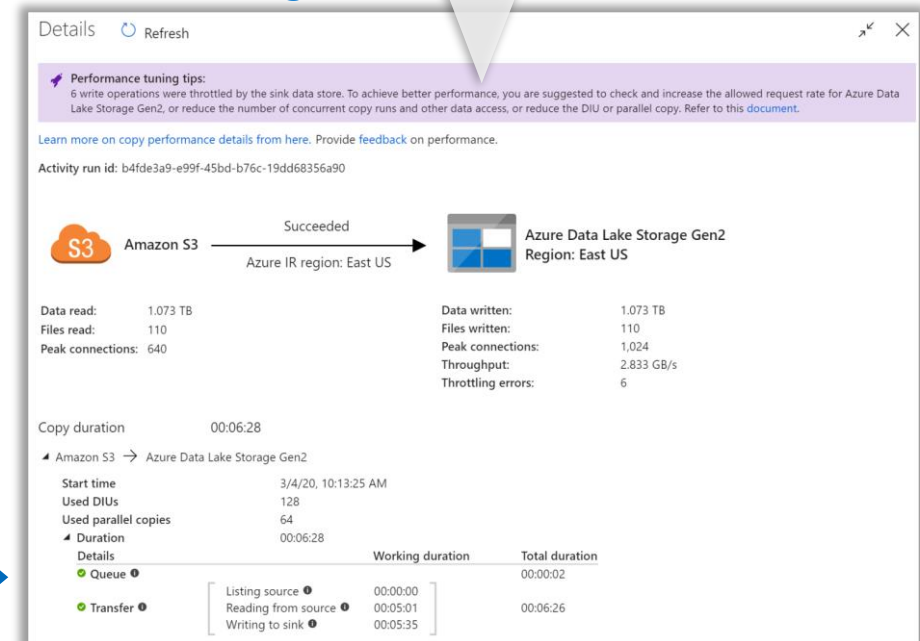
2 JIT compute scalability



Corresponding amount of VM horsepower made available on demand to execute copy activity at scale.

Results of copy run stats, including # of DIUs used, timing breakdown, perf suggestions, will appear in monitoring view.

3 Monitoring View



Data Factory Availability

Access Data Factory is available in 25+ global regions and growing



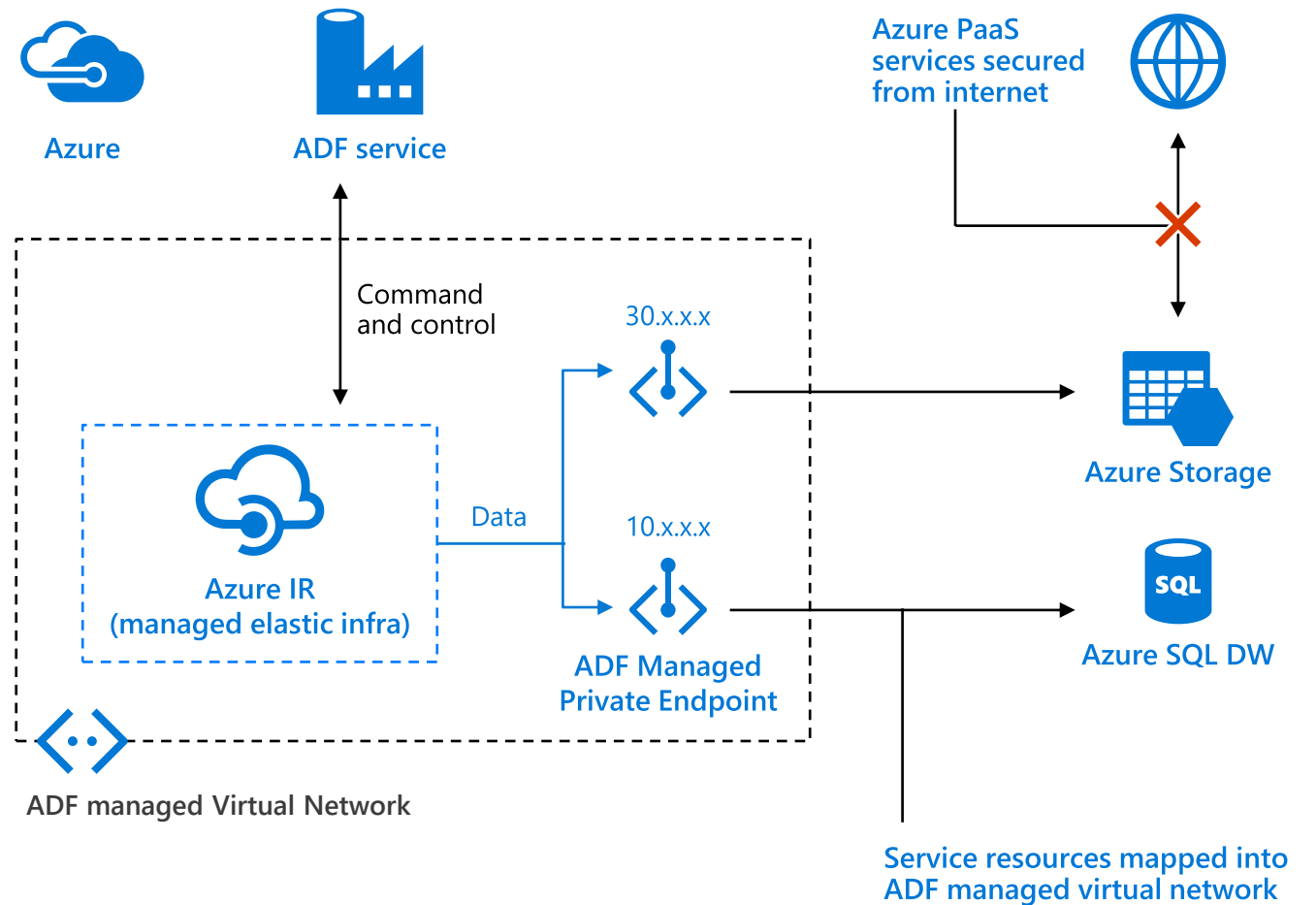
- | | | | | |
|------------------------|--------------------|--------------------|----------------------|---------------|
| 1. South Africa North | 7. US Gov Texas | 13. West Europe | 19. Central US | 25. West US |
| 2. East Asia | 8. US Gov Virginia | 14. France Central | 20. East US | 26. West US 2 |
| 3. Southeast Asia | 9. Brazil South | 15. Central India | 21. East US 2 | |
| 4. Australia East | 10. Canada Central | 16. Japan East | 22. North Central US | |
| 5. Australia Southeast | 11. China East 2 | 17. Korea Central | 23. South Central US | |
| 6. US Gov Arizona | 12. North Europe | 18. UK South | 24. West Central US | |

Data Factory Security

Data Factory is certified by HIPAA and HITECH, ISO/IEC 27001, ISO/IEC 27018, and CSA STAR.

Connect [securely](#) to Azure data services with Managed Identity and Service Principal. Store your credentials with Azure Key Vault.

Managed Virtual Network (in public preview) gives you an isolated and highly secure environment to run your data integration pipelines.





2

Connect to your
enterprise data

It just works!

Connect to all your Enterprise Data



Azure
Synapse
Analytics



Azure
Storage



Azure SQL
DB



Cosmos DB



File
System



Dynamics



SAP HANA
and Table



Oracle
Exadata



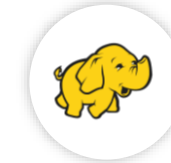
Teradata



Netezza



Mongo DB



HDFS



Salesforce



Marketo



ServiceNow



Amazon S3
and Redshift



Google Cloud
Storage and
BigQuery

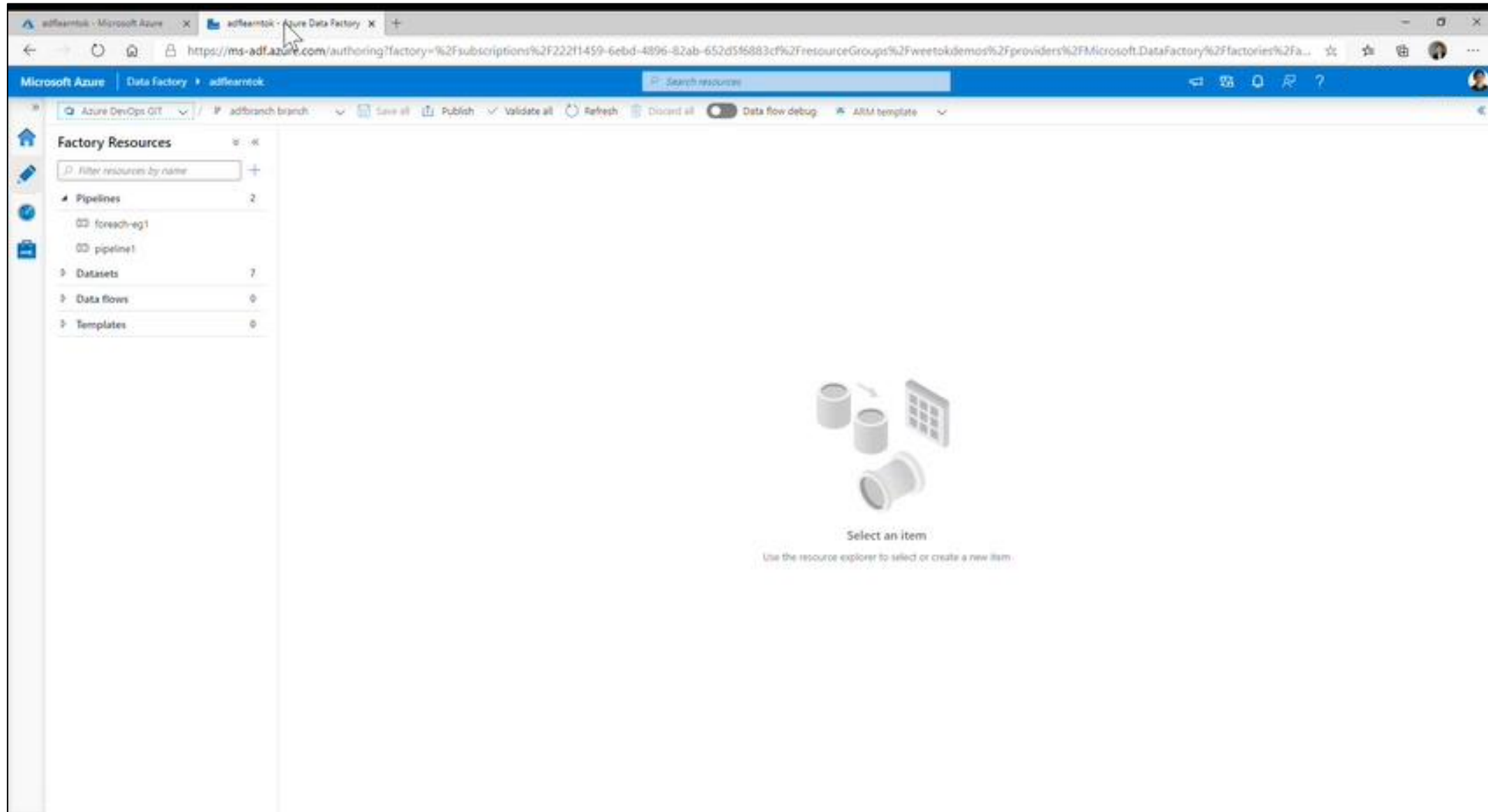


FTP/SFTP

90+

Built-in Connectors
and more...

Connect to all your Enterprise Data





3

Serverless, Code-free
Data Transformation
at scale

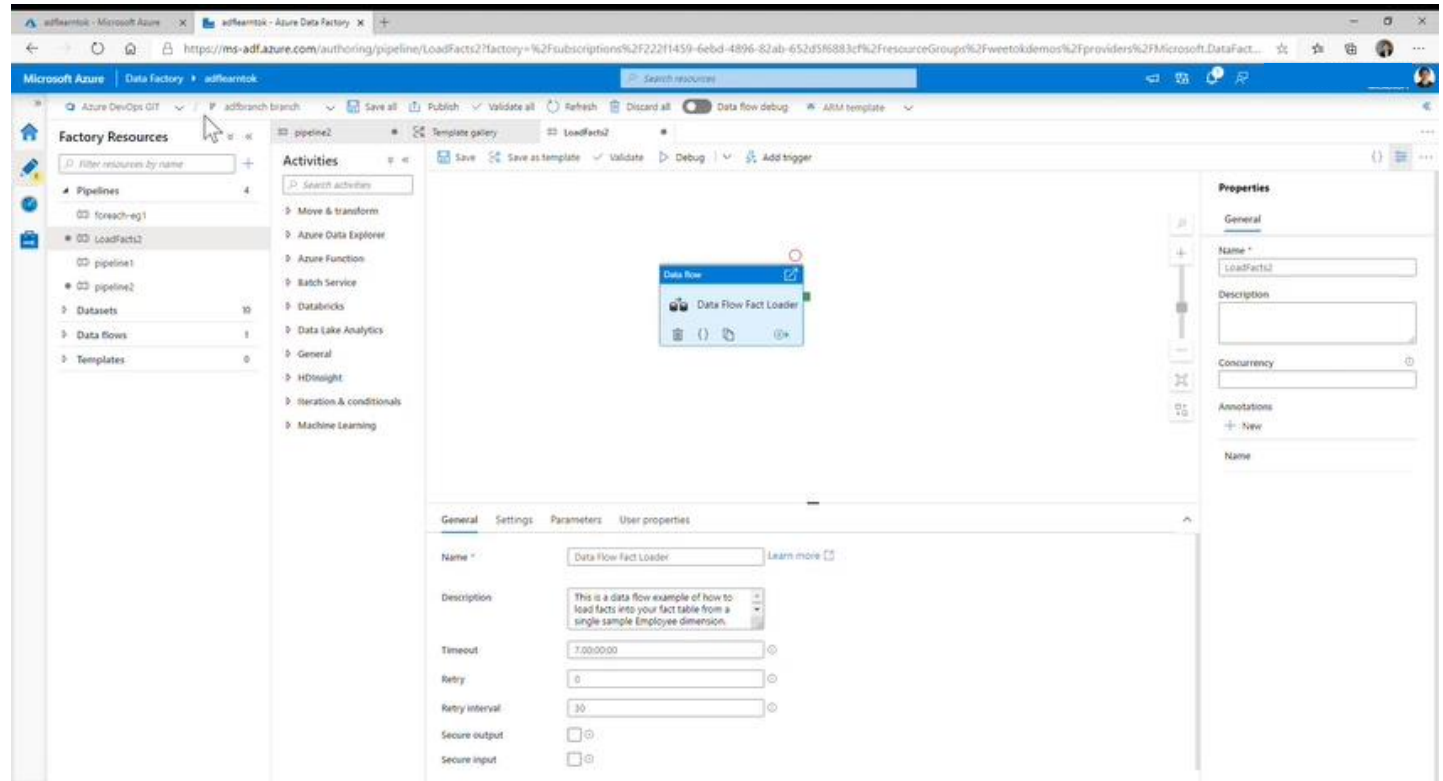
Transform code-free and/or code-centric, you choose

Code-free transformation at scale with Mapping Data Flow

Construct and perform ETL and ELT processes in an intuitive environment

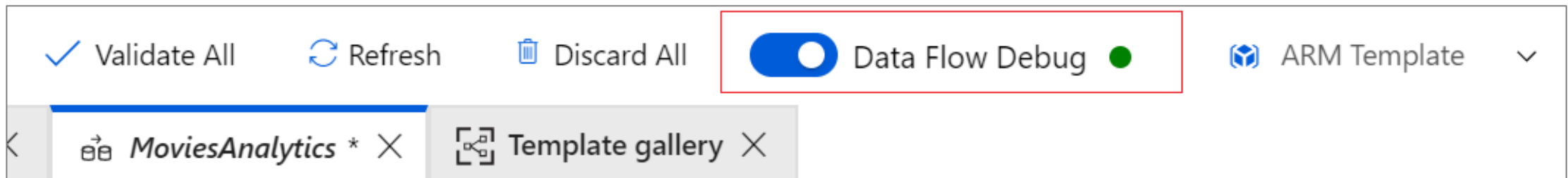
Spend less time on design, testing, and maintenance, while focusing on business logic

- Data cleansing, transformation, aggregation, and conversion
- Cloud scale via Spark execution
- Resilient, easily built data flows



Debug mode

- Quickly verify logic during development on small interactive cluster
 - 4 core, 60-minute time to live
- Enables the following:
 - Get data preview snapshot at each transformation
 - Preview output of expression in expression builder
 - Run debug pipeline with no spin up
 - Import Spark projection of source schema
- **Rule of thumb:** If developing Data Flows, turn on right away
 - Initial 3-5-minute start up time



Debug mode: Data Preview

Connections × TripData × TripFare × MoviesAnalytics * × Template gallery ×

Save Validate Debug Settings Hide Graph Code

MoviesDB

DropColumn

FilterYearsAfter1950

GetGenre

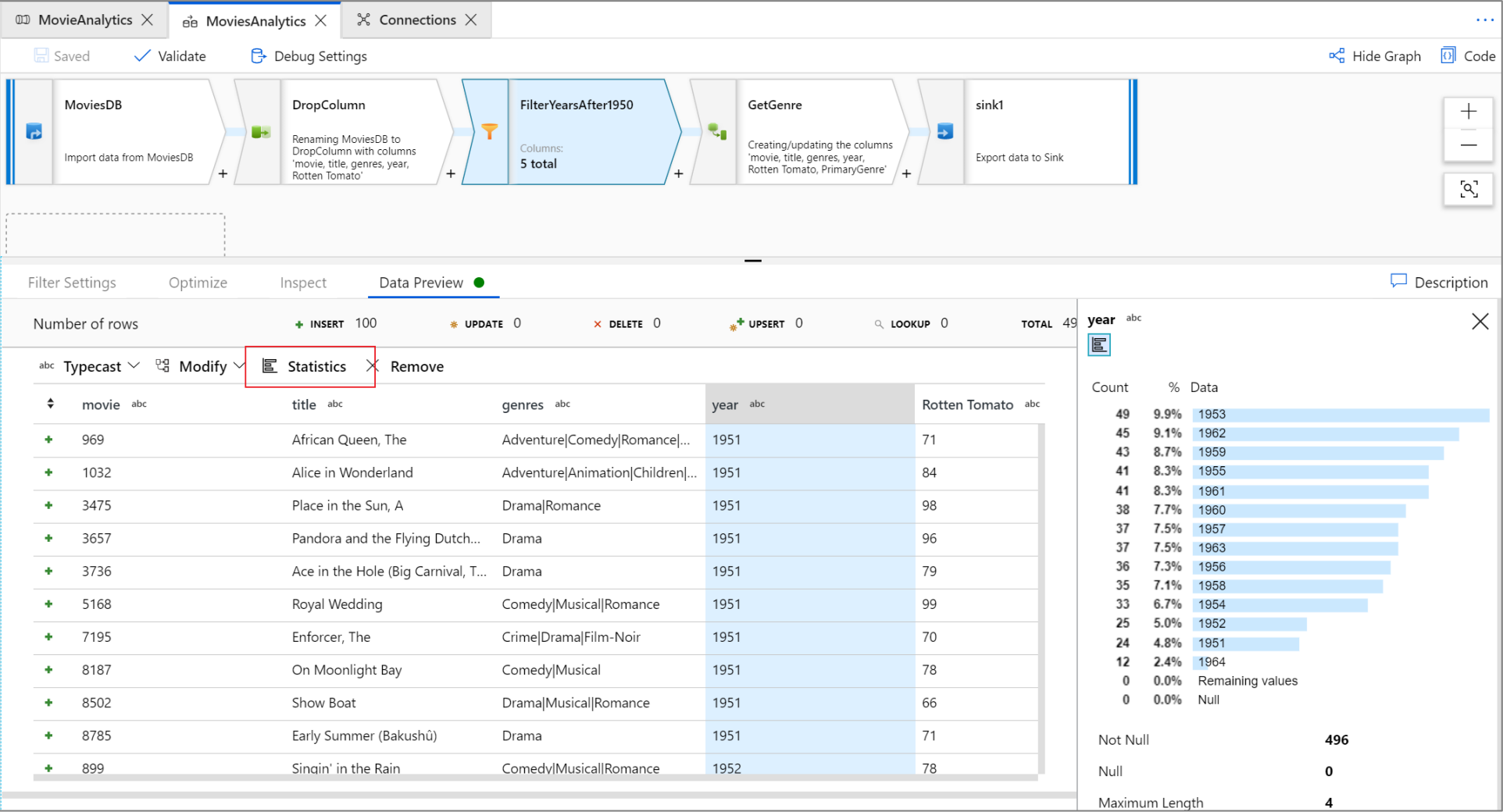
Filter Settings Optimize Inspect Data Preview ● Description

Number of rows INSERT 100 UPDATE 0 DELETE 0 UPSERT 0 LOOKUP 0 TOTAL 8609

Typecast Modify Statistics Remove

	movie	title	genres	year	Rotten Tomato
+	969	African Queen, The	Adventure Comedy Romance War	1951	71
+	1032	Alice in Wonderland	Adventure Animation Children Fa...	1951	84
+	3475	Place in the Sun, A	Drama Romance	1951	98
+	3657	Pandora and the Flying Dutchman	Drama	1951	96
+	3736	Ace in the Hole (Big Carnival, The)	Drama	1951	79
+	5168	Royal Wedding	Comedy Musical Romance	1951	99
+	7195	Enforcer, The	Crime Drama Film-Noir	1951	70
+	8187	On Moonlight Bay	Comedy Musical	1951	78
+	8502	Show Boat	Drama Musical Romance	1951	66
+	8785	Early Summer (Bakushū)	Drama	1951	71
+	899	Singin' in the Rain	Comedy Musical Romance	1952	78

Debug mode: Data Profiling



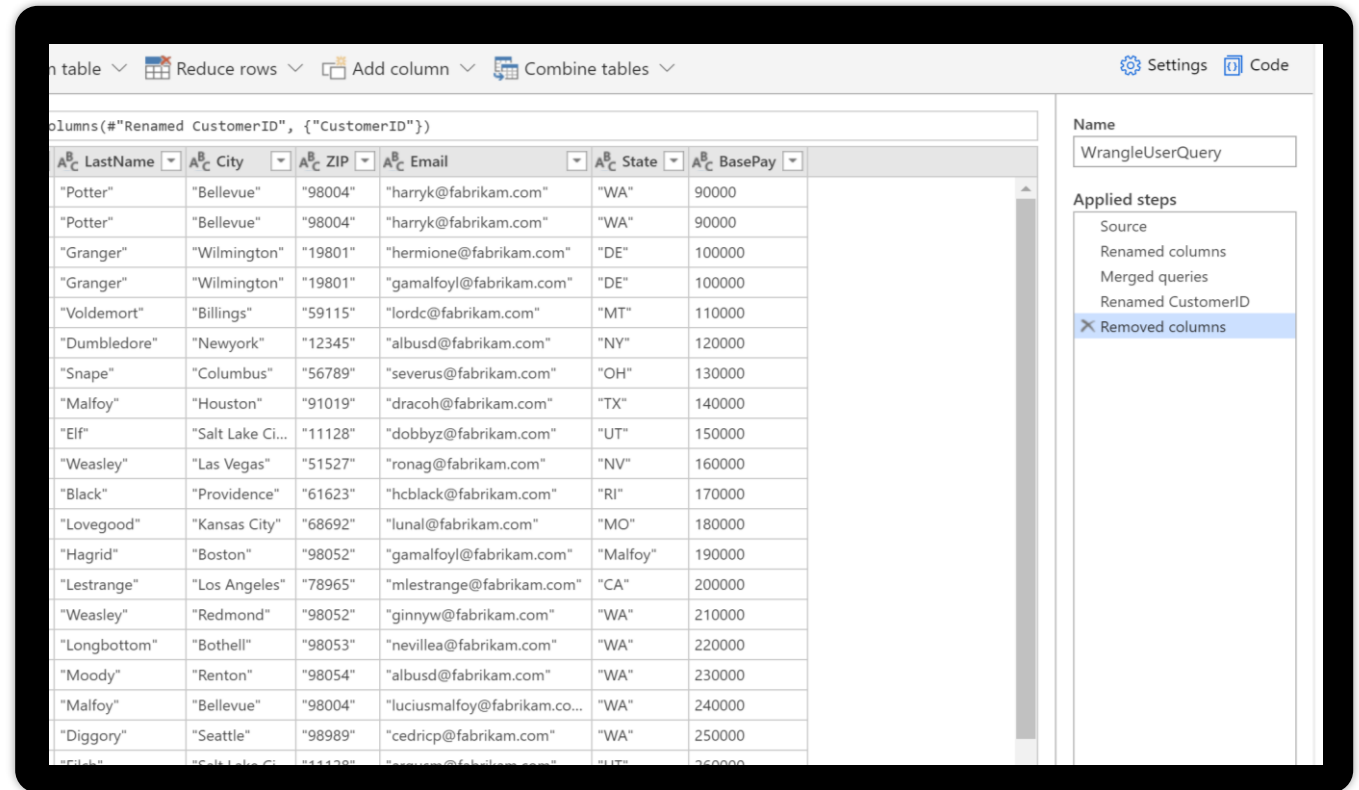
How do I jumpstart using ADF Templates?

I love Power Query

Can I do Power Query in Azure Data Factory?

Visual Data Prep with Power Query in ADF

- Perform code-free, agile data preparation at scale via Spark execution
- Get started easily with familiar Microsoft Excel like UI
- Choose from a variety of wrangling functions like combining tables, adding columns, and reducing rows
- Create, debug, schedule, and monitor the wrangling data flow in a one-stop-shop





4

Run your code on any
compute on Azure

Transform code-free and/or code-centric, you choose

Fully contained operationalization for code-centric computing

ADF comes with completely integrated external compute engines



Azure Databricks

Notebook, Jar, Python



Azure HDInsight

Hive, Pig, Spark, MapReduce,
Streaming



**Synapse SQL Pool,
Azure SQL DB & SQL Server**
Stored Procedure



Machine Learning

Batch Execution, Update Resource



Azure Functions

Function calls



Azure Batch

Custom Executable

A large blue circle with a slight drop shadow is positioned on the right side of a white background. Inside the circle, the number '5' is at the top, followed by two lines of text.

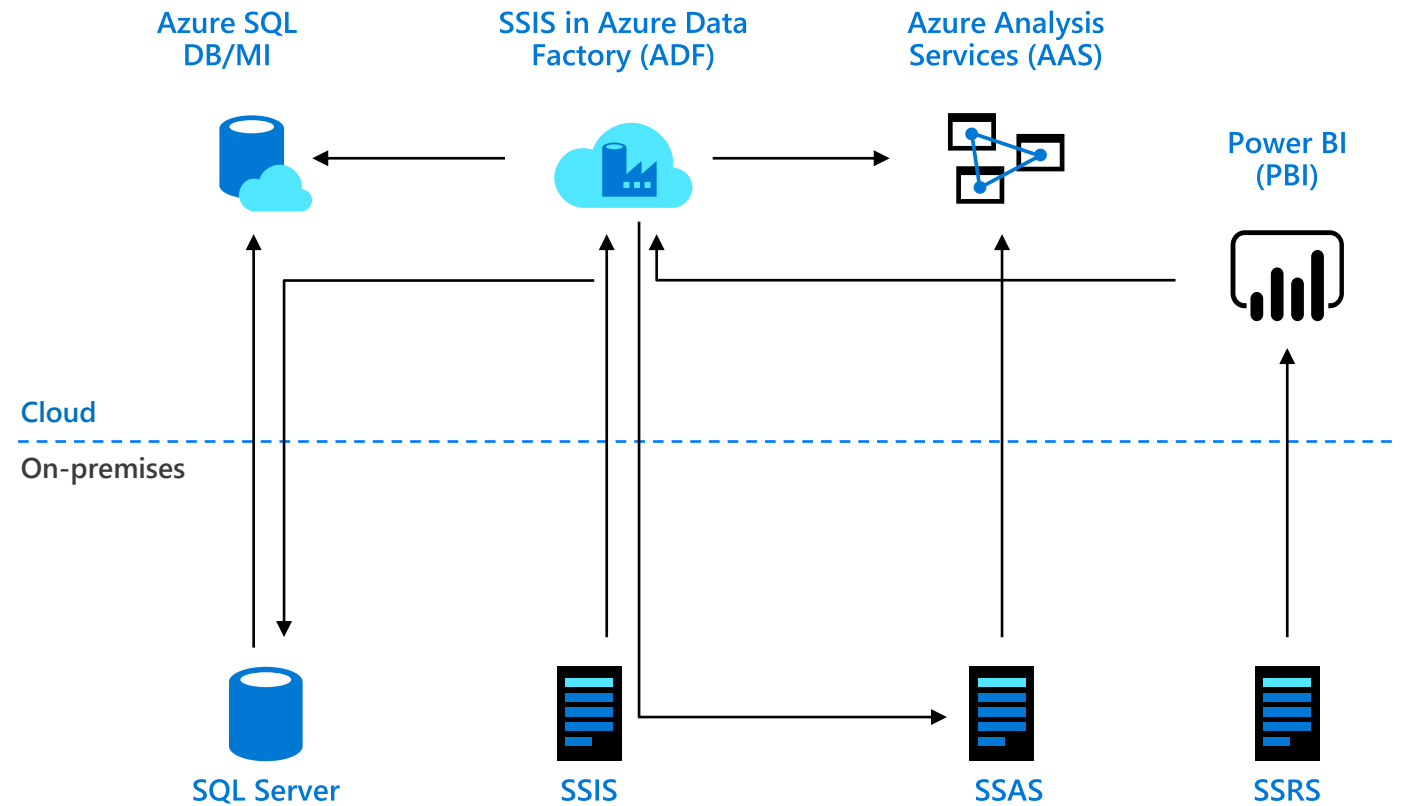
5

Lots of SSIS packages
on-premises?

Run it on ADF
It just works!

SSIS in ADF - It just works!

1. **Reduce operational costs** necessary to maintain on-premises SSIS infrastructure
2. **Streamline rehosting process** aided by the Data Migration Assistant and Data Factory Deployment Wizard
3. **Operationalize ETL** in Data Factory with Azure Monitor and Azure DevOps



Benefits of Modernizing SSIS

Greater potential – enable saving up to 88%*

Modernize SQL server and SSIS in tandem to lower TCO significantly.

The only fully compatible solution

Migrate all your workloads, custom and third-party components to the cloud.

Easy-to-use

Use the Data Factory deployment wizard tool and refer to how-to documentation.

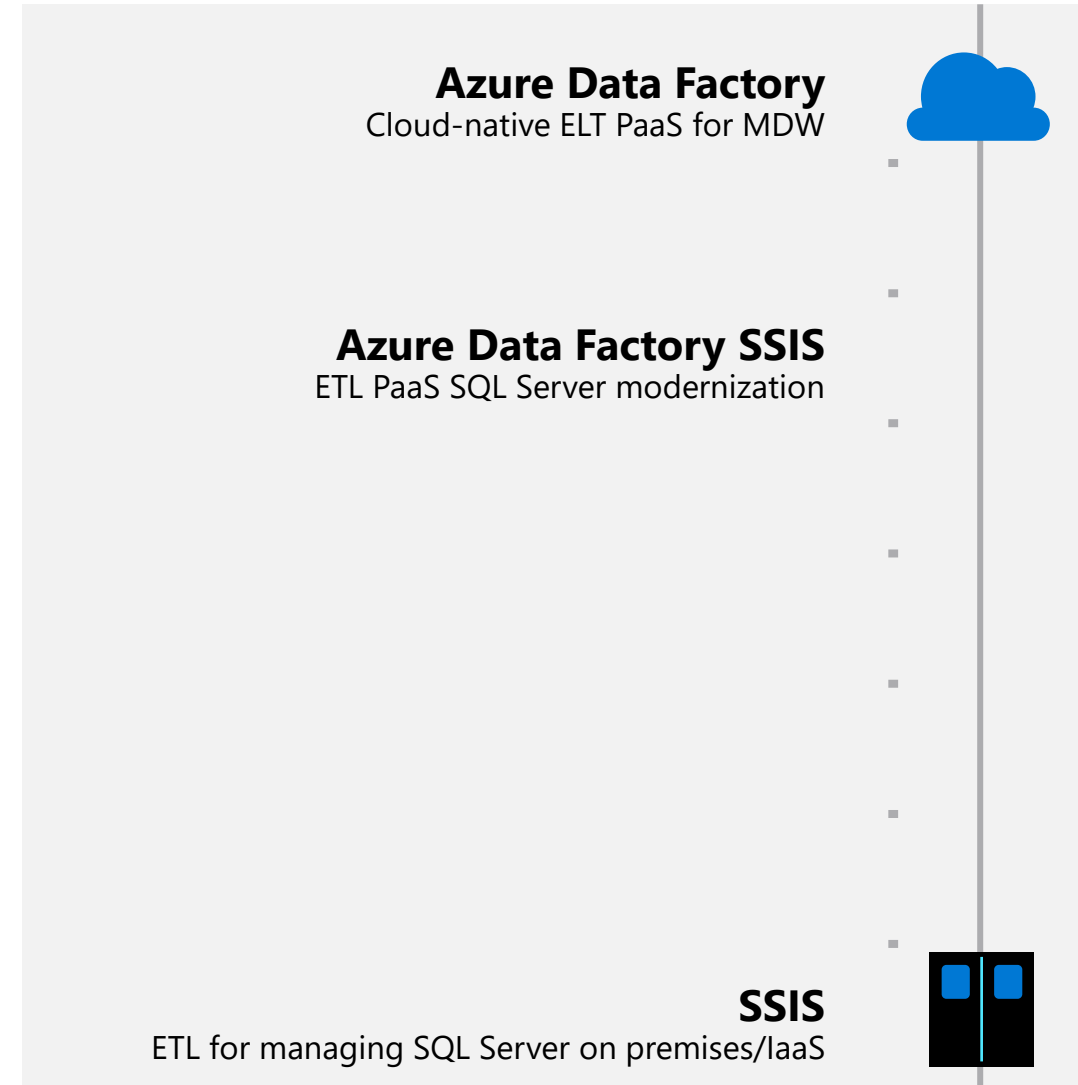
Experience the benefits of the cloud

Focus on your data while dedicated data integration does the rest.

Unlock limitless insights with Azure Synapse

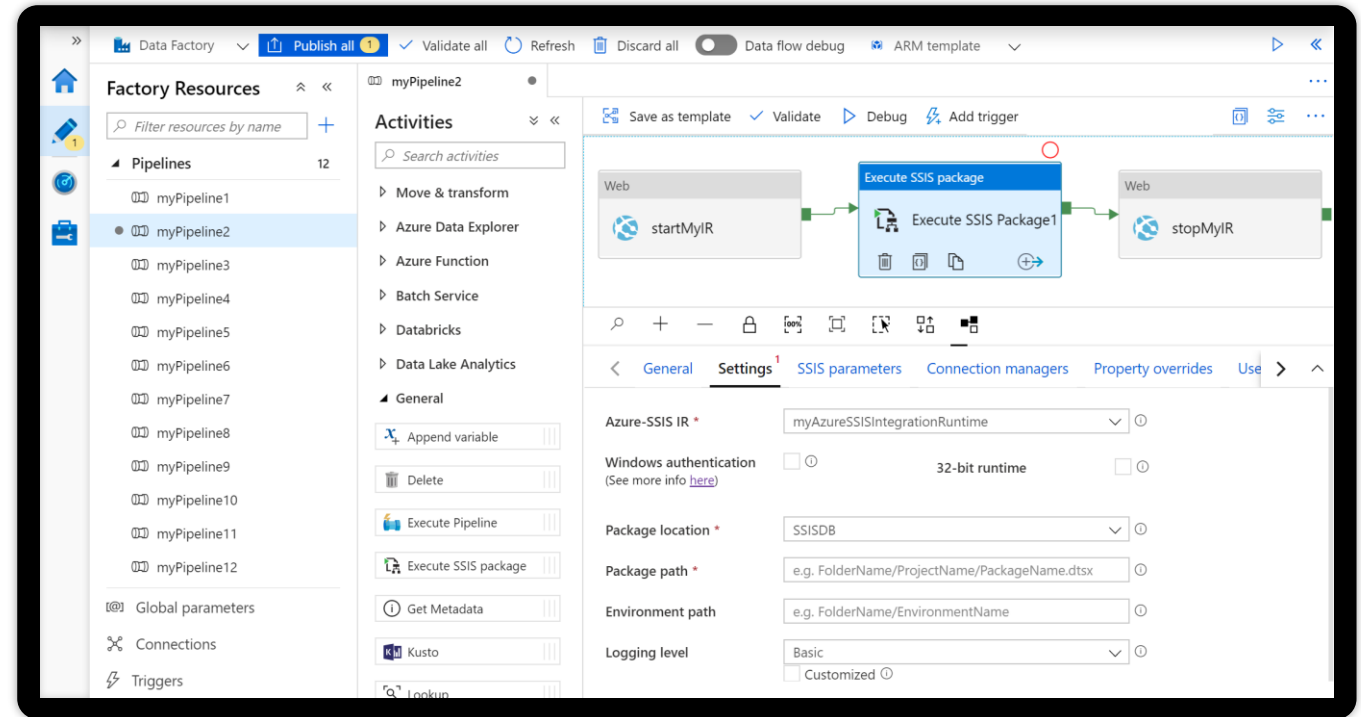
Query data on your terms using dedicated on-demand or provisioned resources – at scale

[*Azure Data Factory pricing details](#)



Modernize your ETL packages with Data Factory

1. **Refine enormous stores of raw** data into actionable business insights
2. **Develop complex ETL processes** to help you integrate data silos
3. **Complement ETL** with native Data Factory capabilities and perform first-class SSIS activities

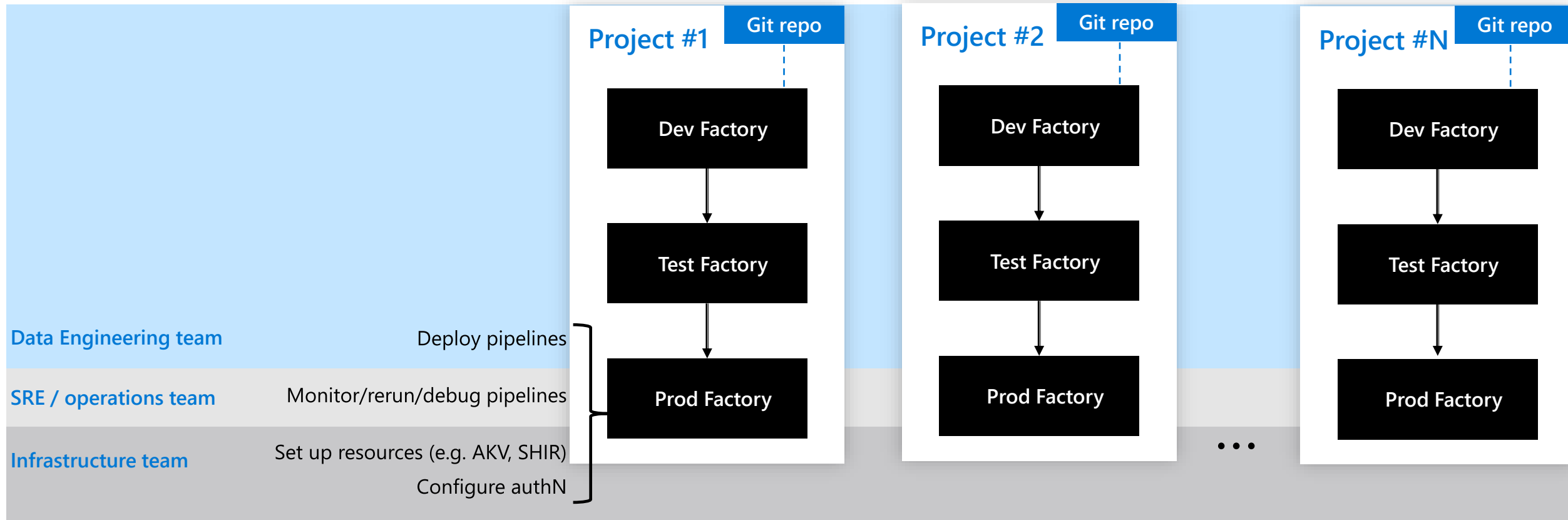


6

DataOps
and ADF

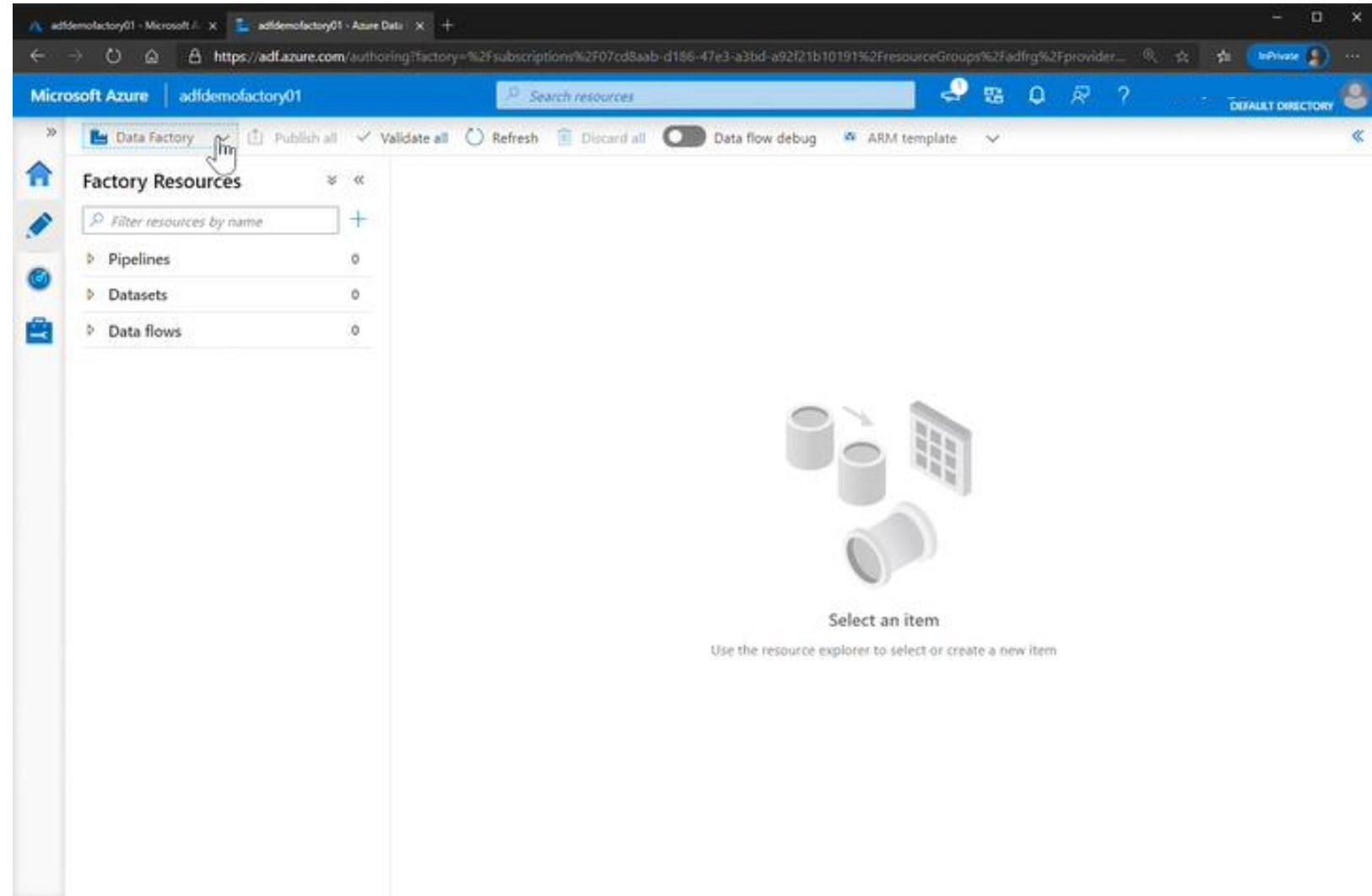
Source Control and CI/CD

- Separate factory for each project & each environment
- Dev factory is associated with Git repo
- Changes are deployed from Dev to Test & Prod using Azure DevOps pipeline or ARM template
- Use custom roles for granular RBAC to resources within factory



Source Control and CI/CD with Data Factory

1. **Source Control** with Azure DevOps or GitHub
2. **Automate Deployment of Data Factory** into multiple environments (e.g. Dev, Test, Production) using Azure pipelines
3. **Simplify Complex Data Factories**
ARM Templates ADF automatically generates linked resource manager templates for large factories



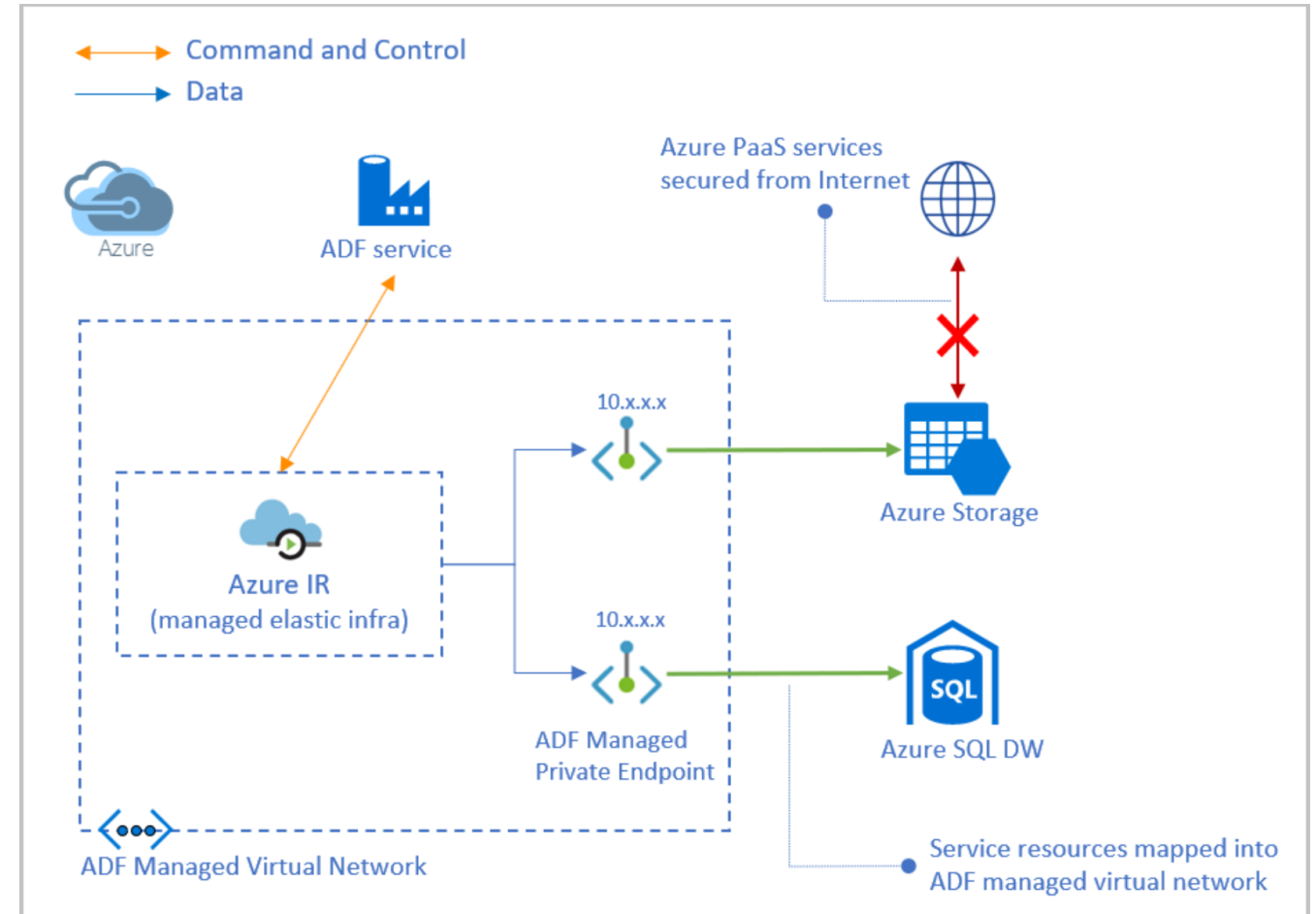
A large blue circle with a slight drop shadow, containing the number 7 and the title text.

7

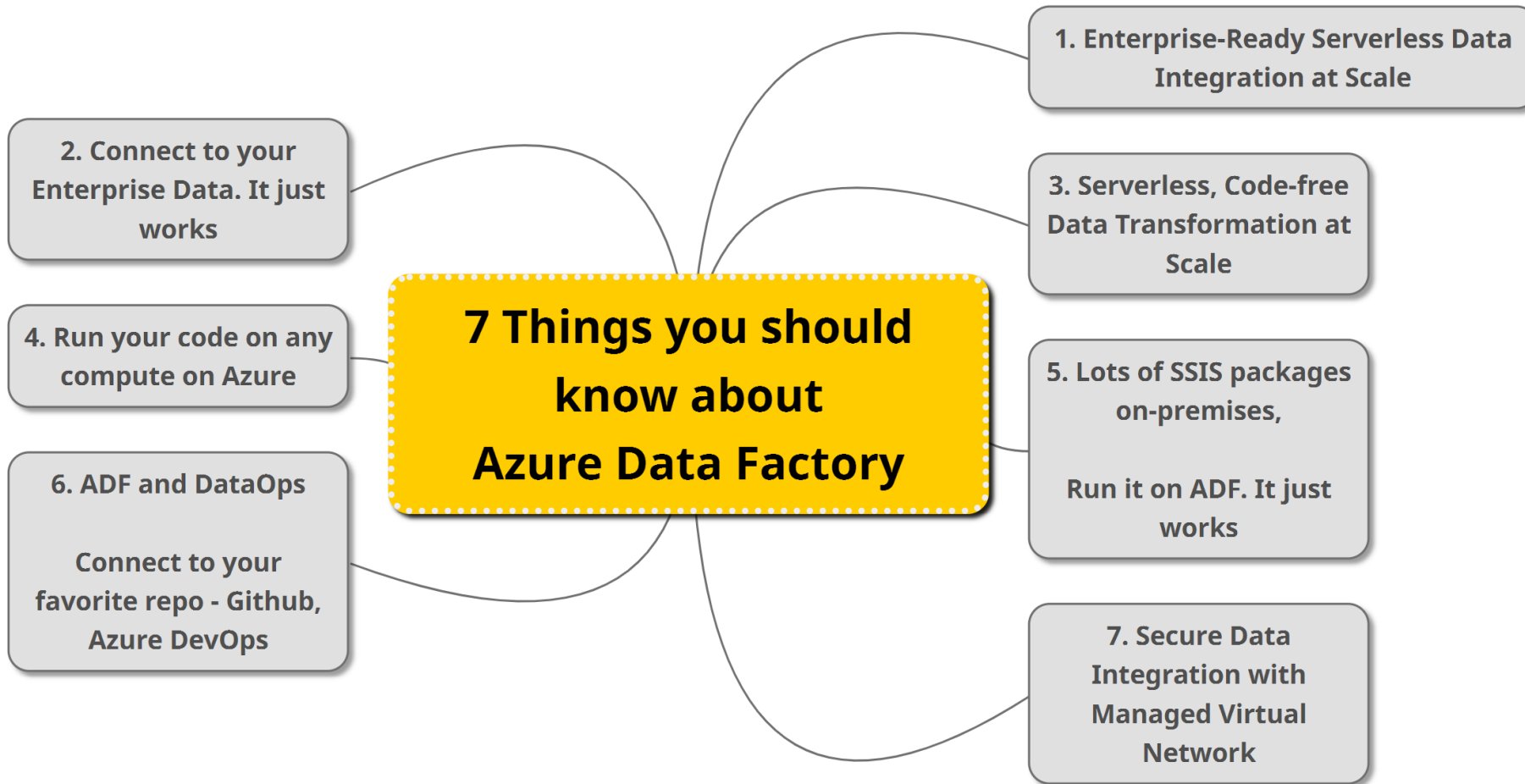
Secure Data
Integration with
Managed Virtual
Network

Managed Virtual Network (preview)

1. **Managed Virtual Network + Managed Private Endpoints** - Secure data integration managed by Azure Data Factory
2. **Simplify Complex networking** - Setup end to end data integration using secure virtual network without deep networking knowledge
3. **Protect against data exfiltration** - Traffic between data stores and managed virtual network entirely on Microsoft backbone



Summary



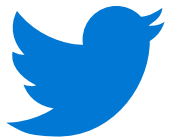
How do I get started with Azure Data Factory?



<https://azure.microsoft.com/services/data-factory/>



<https://docs.microsoft.com/learn/paths/data-integration-scale-azure-data-factory/>



Stay in touch

<https://twitter.com/azdatafactory>

Youtube Channel

aka.ms/adfvideos

DATA() {
EXPOSED;

Learn with us!

View our on-demand playlist:
aka.ms/azuresqlandadf

@AzureSQL
@AzDataFactory



