Azure Data Factory
Hybrid data integration, at global scale

Erika Harris – Senior Program Manager
AzureCAT
There are barriers to getting value from data

- Data silos
- Incongruent data types
- Complexity of solutions
- Multi cloud environment
- Rising costs
Derive real value from your data

- Data silos
  - One hub for all data
- Incongruent data types
  - Support for diverse types of data
- Performance constraints
  - Unlimited data scale
- Complexity of solutions
  - Familiar tools and ecosystem
- Rising costs
  - Lower TCO

On-premises, hybrid, Azure
AZURE DATA FACTORY
Hybrid data integration, at global scale

PRODUCTIVE
✓ Drag & Drop UI
✓ Codeless Data Movement

HYBRID
✓ Orchestrate where your data lives
✓ Lift SSIS packages to Azure

SCALABLE
✓ Serverless scalability with no infrastructure to manage

TRUSTED
✓ Certified compliant Data Movement
AZURE DATA FACTORY

Hybrid data integration, at global scale

INTTEGRATE VIA AZURE DATA FACTORY

INGEST
Data orchestration and monitoring

STORE
Data lake and storage

PREP & TRANSFORM
Hadoop/Spark and machine learning

MODEL & SERVE
Data warehouse & Data lake

Cloud

VNet

On-premise

LOB

CRM

Graph

Image

Social

IoT

Hybrid data integration, at global scale

Integrate via Azure Data Factory
SCENARIOS
Modernize your enterprise data warehouse at scale

**Azure Data Factory**

On-premises data
- Oracle, SQL, Teradata, fileshares, SAP

Cloud data
- Azure, AWS, GCP

SaaS data
- Salesforce, Dynamics

Microsoft Azure also supports other Big Data services like Azure HDInsight, Azure SQL Database and Azure Data Lake to allow customers to tailor the above architecture to meet their unique needs.
Microsoft Azure also supports other Big Data services like Azure HDInsight, Azure SQL Database, and Azure Data Lake to allow customers to tailor the above architecture to meet their unique needs.
Lift your SQL Server Integration Services (SSIS) packages to Azure

Azure Data Factory

SSIS Integration Runtime

Cloud data sources

SQL DB Managed Instance

SSIS Cloud ETL

Cloud

On-premises

VNET

On-Premise data sources

Microsoft SQL Server Integration Services

SQL Server
Hybrid and Multi-Cloud Data Integration

Author, orchestrate and monitor with Azure Data Factory

On-Prem
- SAP
- cloudera
- Microsoft SQL Server
- Hortonworks

SaaS Apps
- Oracle
- Teradata
- Salesforce
- Adobe
- Microsoft Dynamics

Public Cloud
- Azure
- Amazon Web Services
- Google Cloud Platform
Integration Runtime
ADF Integration Runtime (IR)

ADF compute environment with multiple capabilities:
- Activity dispatch & monitoring
- Data movement
- SSIS package execution

To integrate data flow and control flow across the enterprises’ hybrid cloud, customer can instantiate multiple IR instances for different network environments:
- On premises (similar to DMG in ADF V1)
- In public cloud
- Inside VNet

Bring a consistent provision and monitoring experience across the network environments

Portal Application & SDK

Self-Hosted IR

Data Movement & Activity Dispatch on-prem, Cloud, VNET

Azure IR

Data Movement & Activity Dispatch In Azure Public Network, SSIS VNET coming soon
Integration Runtime

- Activity Dispatch/Monitor (spark, copy, ml, etc)
- Data Movement
- SSIS Package Execution
Azure Data Factory “Integration Runtime” deployed on premises for transformation and then moved to cloud
Azure Data Factory “Integration Runtime” deployed inside VNet
Depends On

For-each Activity

VNET Integration

Flexible Scheduling

SSIS Runtime

Conditional Branching

Parameters

KeyVault Integration

Trigger Execution

Custom State Passing

Expressions

Self Hosted IR

Over 70 Data Source Connectors

Developer Experience

Self Hosted IR High Availability & Scalability

PolyBase Data Movement
Get started with Azure Data Factory


View pricing


Hands-on-lab

https://aka.ms/adflab2