Unleash the power of Azure Data factory

Ashish Agrawal
Enterprise Architect – Cloud & Data

AZURE DATA FACTORY

Fully-managed data integration service in the cloud

Flexible
Data integration

Hybrid
Data orchestration

Data movement
As-a-service

Security and compliance

Flexible
Data integration

Hybrid
Data orchestration

Data movement
As-a-service

Modernize your data warehouse with Azure big data and advanced analytics services such as HDInsight and Data lake Analytics

Build custom data-driven SaaS applications unique to your customer data using your language of choice

Bring together all your sources of data to understand your customers and drive impactful business decisions

Flexible
Data integration

Hybrid
Data orchestration

Data movement
As-a-service

Orchestrate your data pipeline wherever your data lives – in cloud or in self-hosted environment

Meet your security and compliance needs while taking advantage of truly hybrid integration capabilities

Execute your SQL Server Integration Services (SSIS) packages in the cloud

Flexible
Data integration

Hybrid
Data orchestration

Data movement
As-a-service

Accelerate integration with managed data movement as-a-service

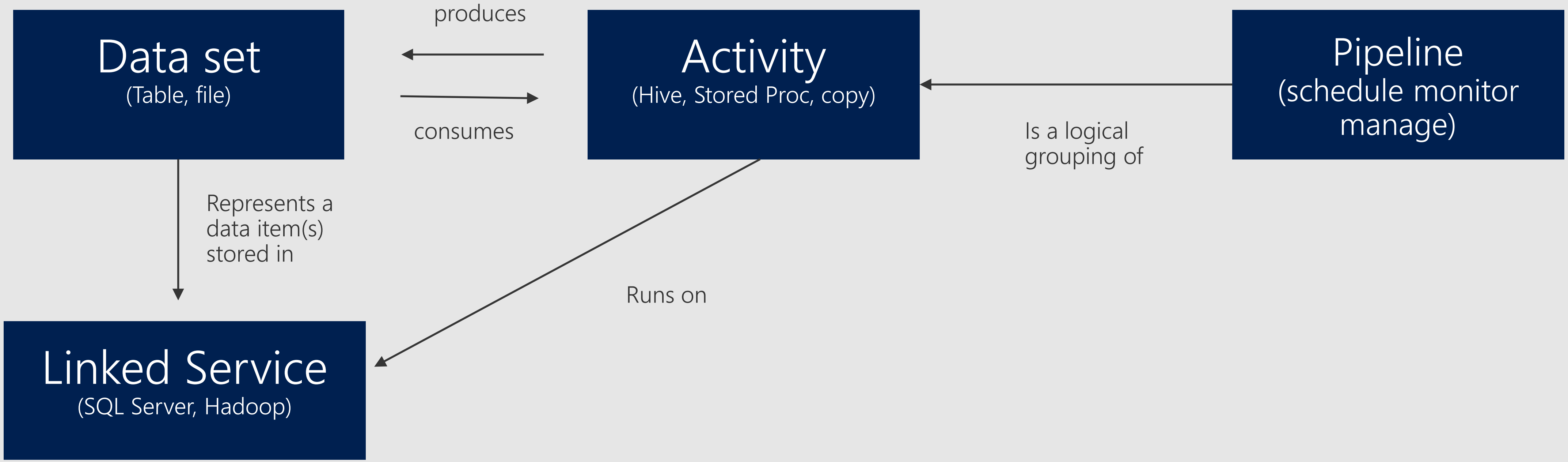
Visually integrate data sources using more than 90+ natively built and maintenance-free connectors at no added cost

Elastic data movement at scale

Serverless data movement with no infrastructure to manage

ADF: Key Components

- Pipelines
- Activities
- Datasets
- Linked services
- Data Flows
- Integration Runtimes



ADF: Cloud-First Data Integration Scenarios

Cloud Transformation

- Migrate on-prem DW to Azure
- Lift and shift existing on-prem SSIS packages to cloud
- No changes needed to migrate SSIS packages to Cloud service

DW Modernization

- Modernizing DW arch to reduce cost & scale to needs of big data (volume, variety, etc)
- Flexible wall-clock and triggered event scheduling
- Incremental Data Load

Build Data-Driven, Intelligent SaaS Application

- C#, Python, PowerShell, ARM support

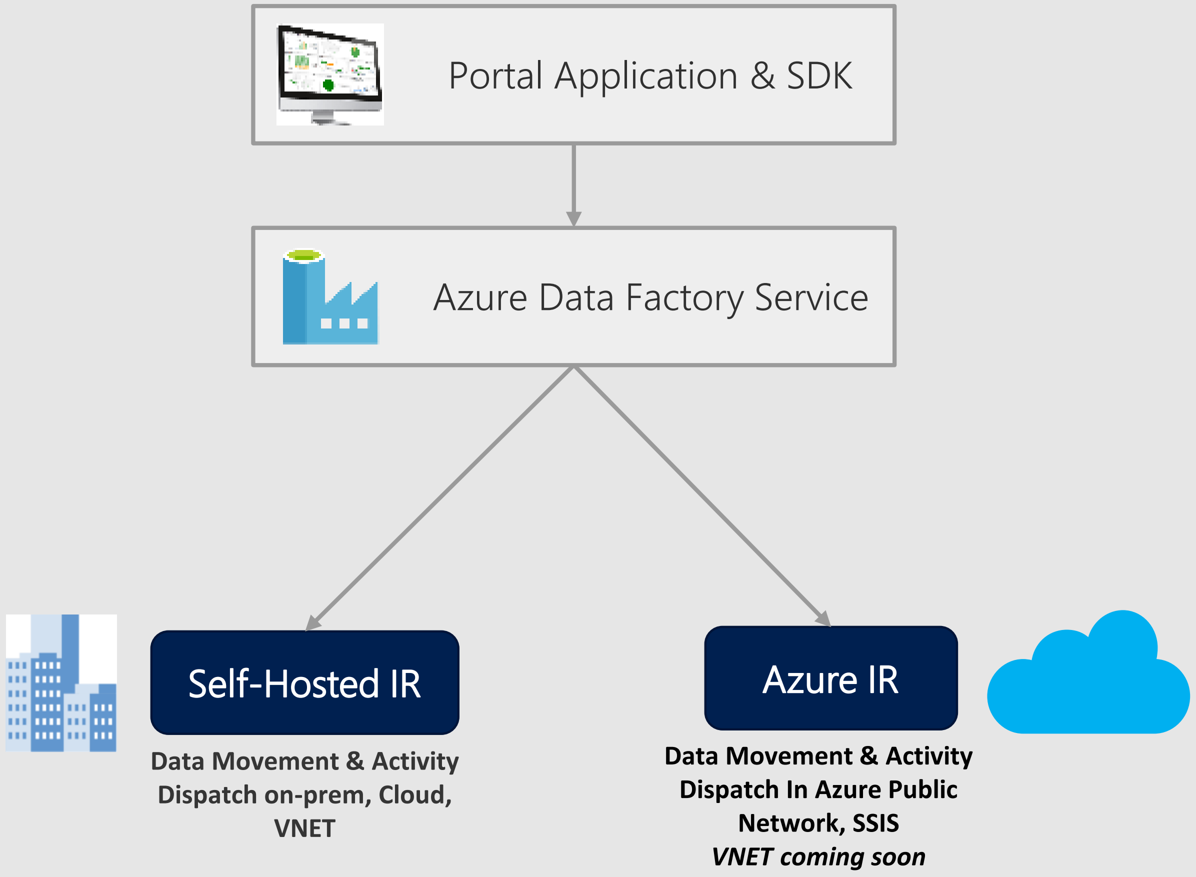
Big Data Analytics

- Customer profiling, Product recommendations, Sentiment Analysis, Churn Analysis, Customized offers, customer usage tracking, customized marketing
- On-demand Spark cluster support

Load your Data Lake

- Separate control-flow to orchestrate complex patterns with branching, looping, conditional processing

ADF Integration Runtime (IR)



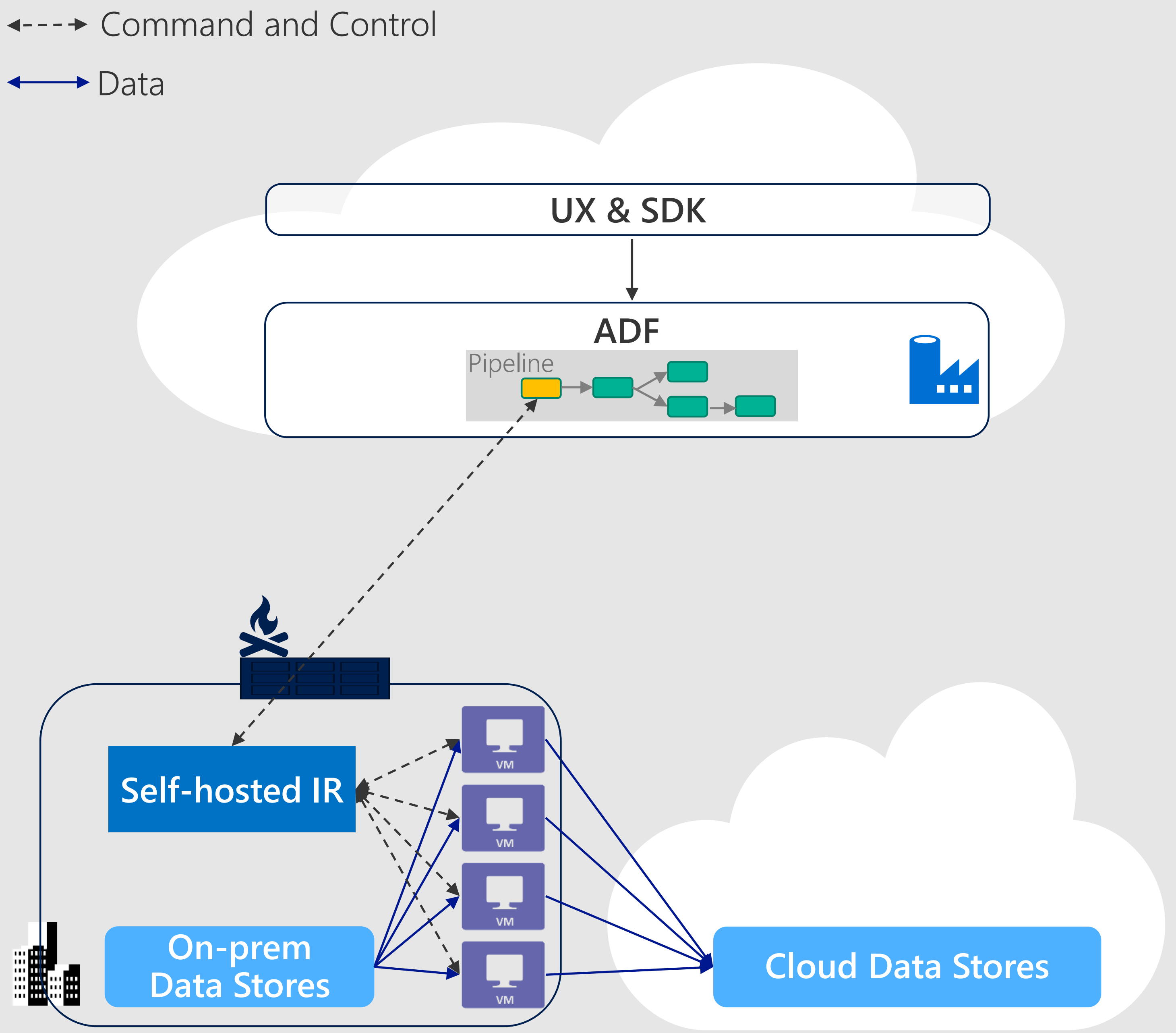
Data Movement & Activity Dispatch on-prem, Cloud, VNET

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

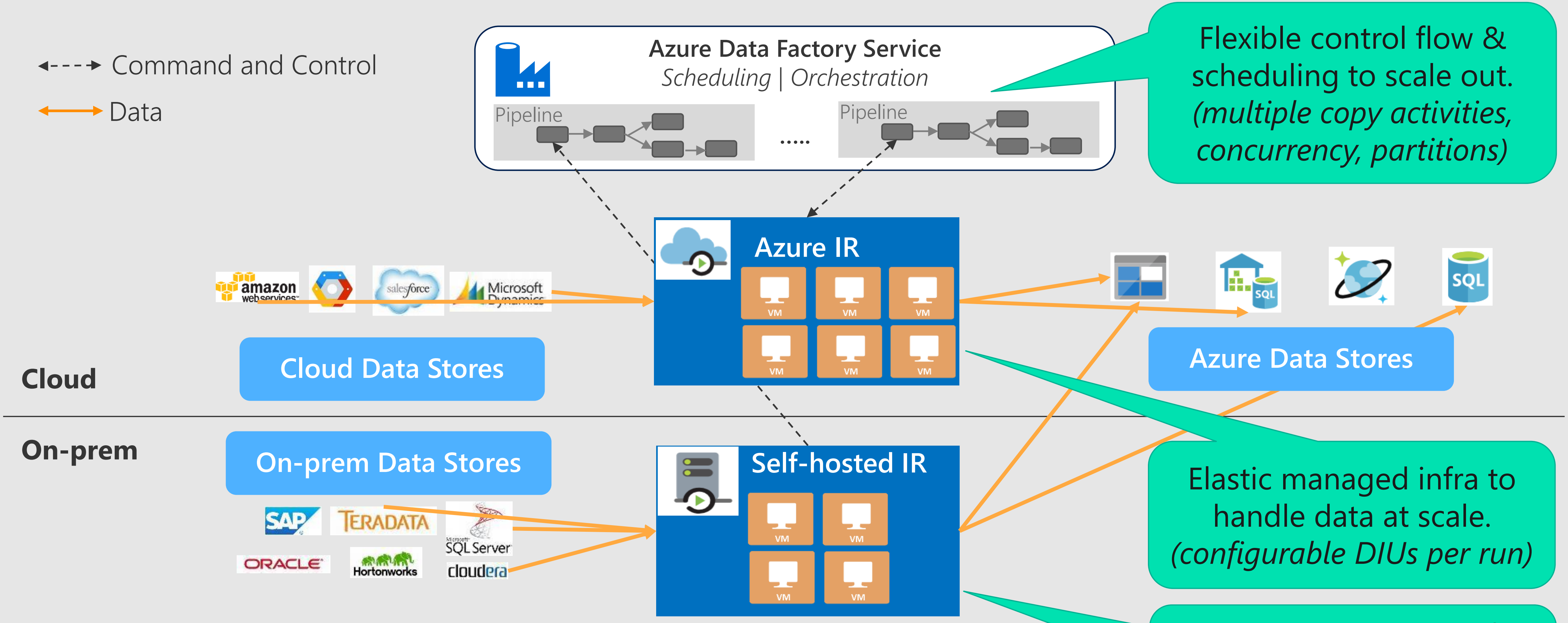
- ADF compute environment with multiple capabilities:
  - Activity dispatch & monitoring
  - Data movement
  - SSIS package execution
- To integrate data flow and control flow across the enterprises' hybrid cloud, customer can instantiate multiple IR instances for different network environments:
  - On premises (similar to DMG in ADF V1)
  - In public cloud
  - Inside VNet
- Bring a consistent provision and monitoring experience across the network environments

Hybrid Copy

- **Self-hosted Integration Runtime:** component installed on machine on-prem or VM in cloud
- **Touchless:** latest version automatically pushed down to machine during downtime
- **HA and scale-out:** register up to 4 nodes for each self-hosted IR.
- **Active-active mode:** requests are dispatched to nodes using round-robin.
- **Single-node concurrency:** configure # of concurrent activity runs, default behavior is determined based on IR CPU/memory



Understand How ADF Copy Scales



Flexible control flow & scheduling to scale out. *(multiple copy activities, concurrency, partitions)*

Elastic managed infra to handle data at scale. *(configurable DIUs per run)*

Customer managed infra with scaling options. *(powerfulness, concurrency)*

Addition:
Parallelism within each copy activity run (degree of parallelism & partition options)
Do not specify “max concurrent connection” unless you want to limit the # of conn to store