Module 5

Pipeline Extensibility

```sql
BEGIN --get ready

    SELECT
        [Contents]
    FROM
        [Training]
    WHERE
        [Module] = '5';
```

- Azure Batch Service
  - Tasks
  - Compute Pools
  - Scaling
- Pipeline Custom Activities
- Azure Management API
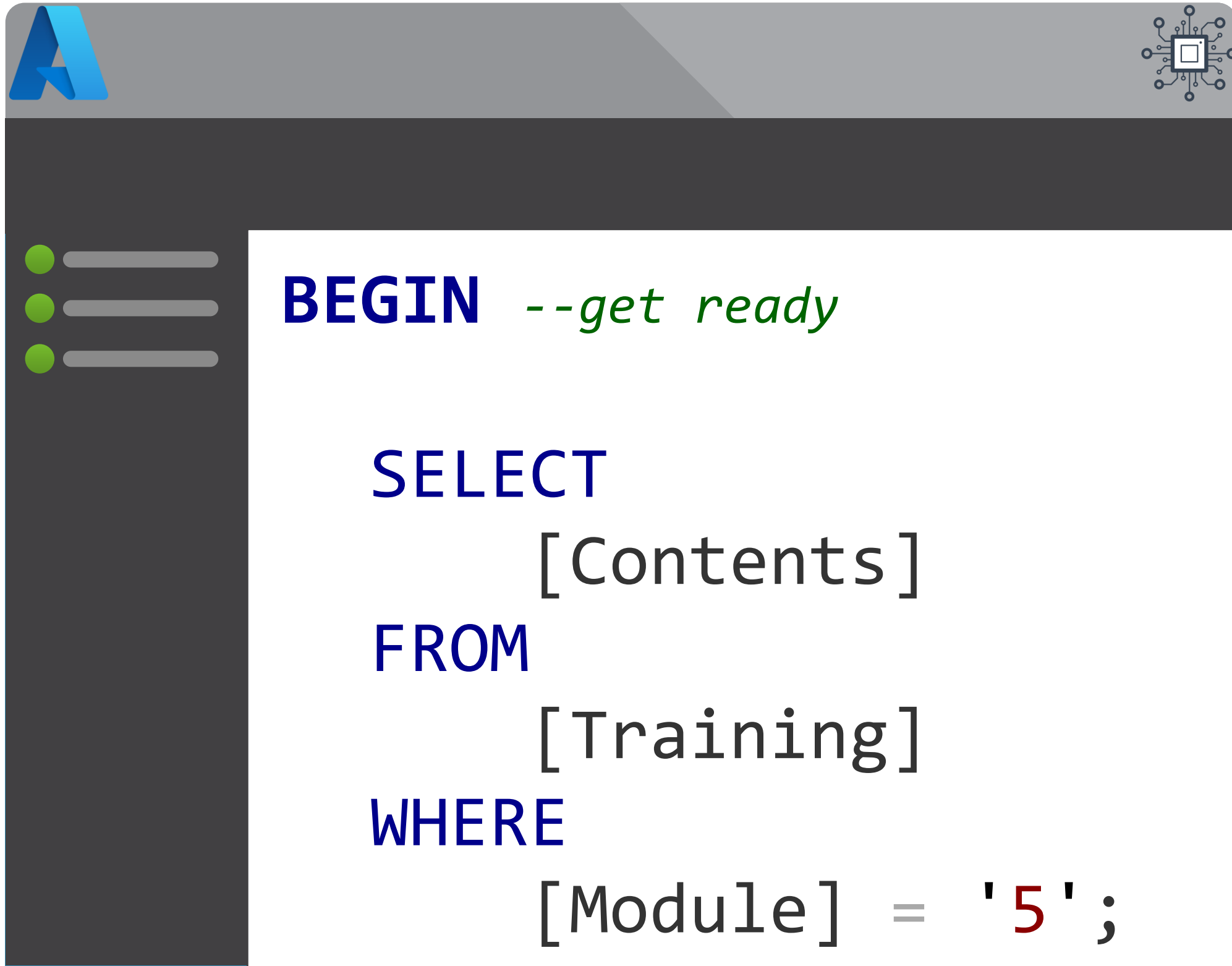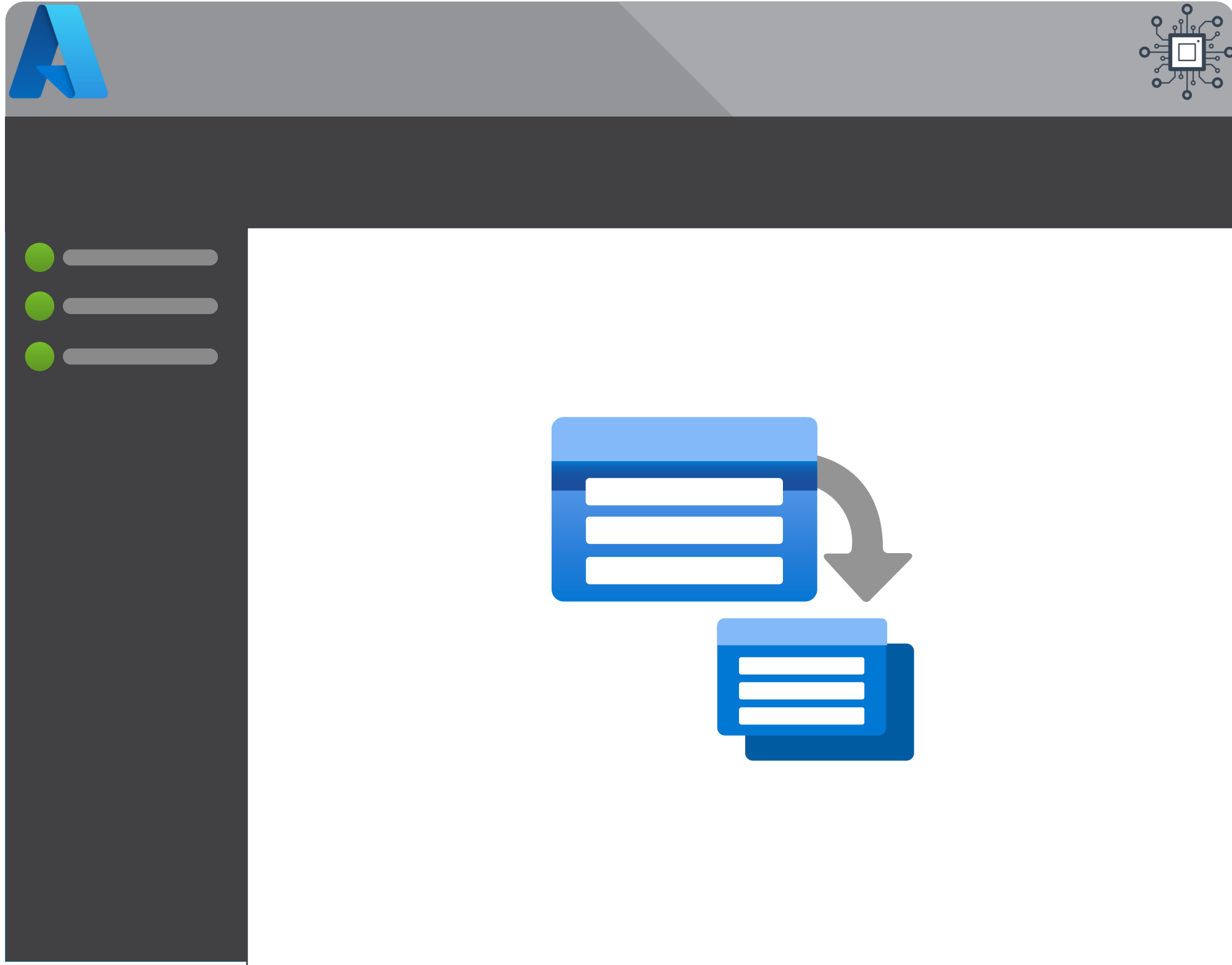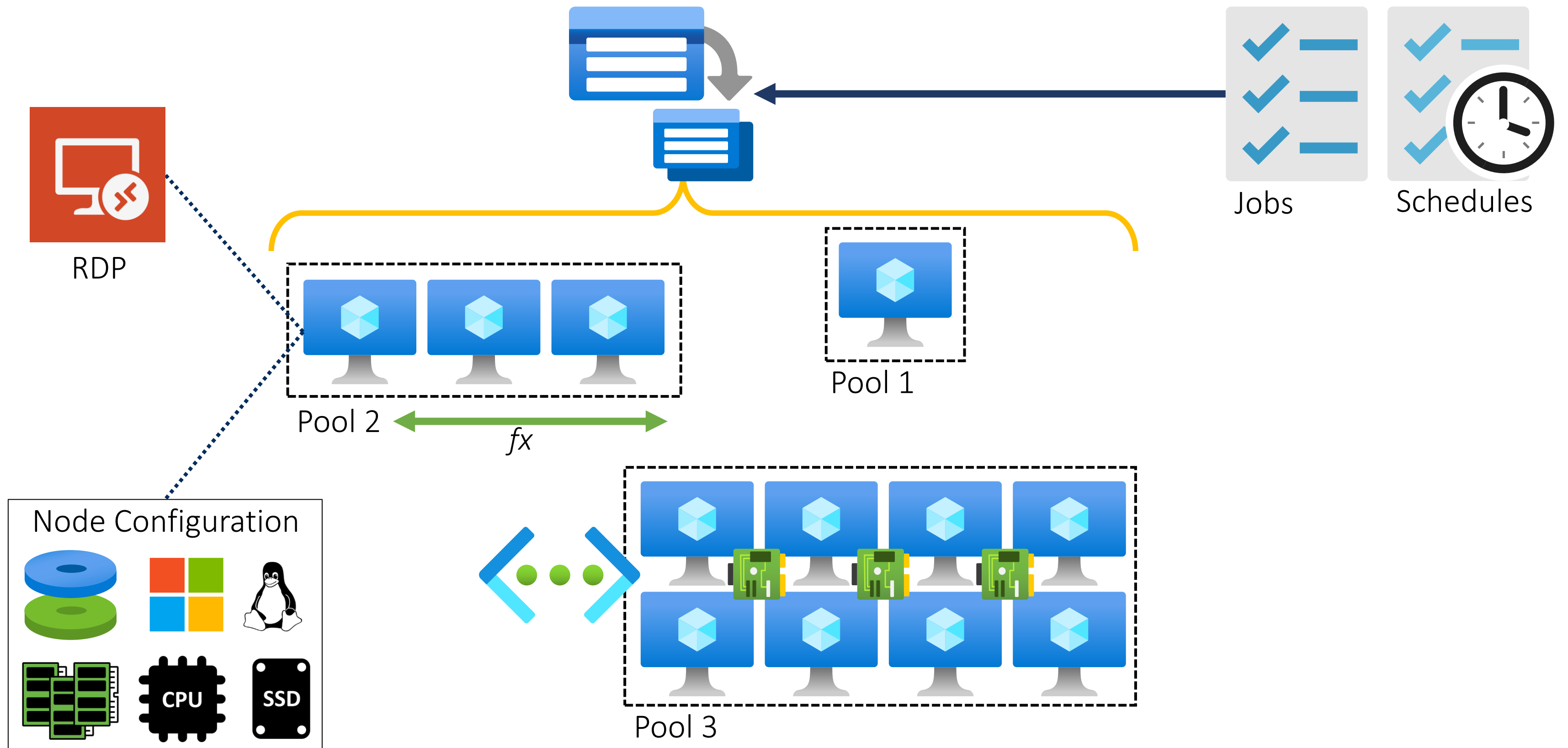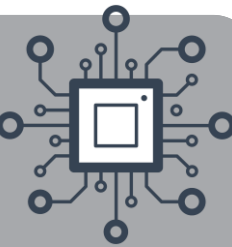- Azure Functions

Module 6

Pipeline Extensibility

- **Azure Batch Service**
  - Tasks
  - Compute Pools
  - Scaling
- Pipeline Custom Activities
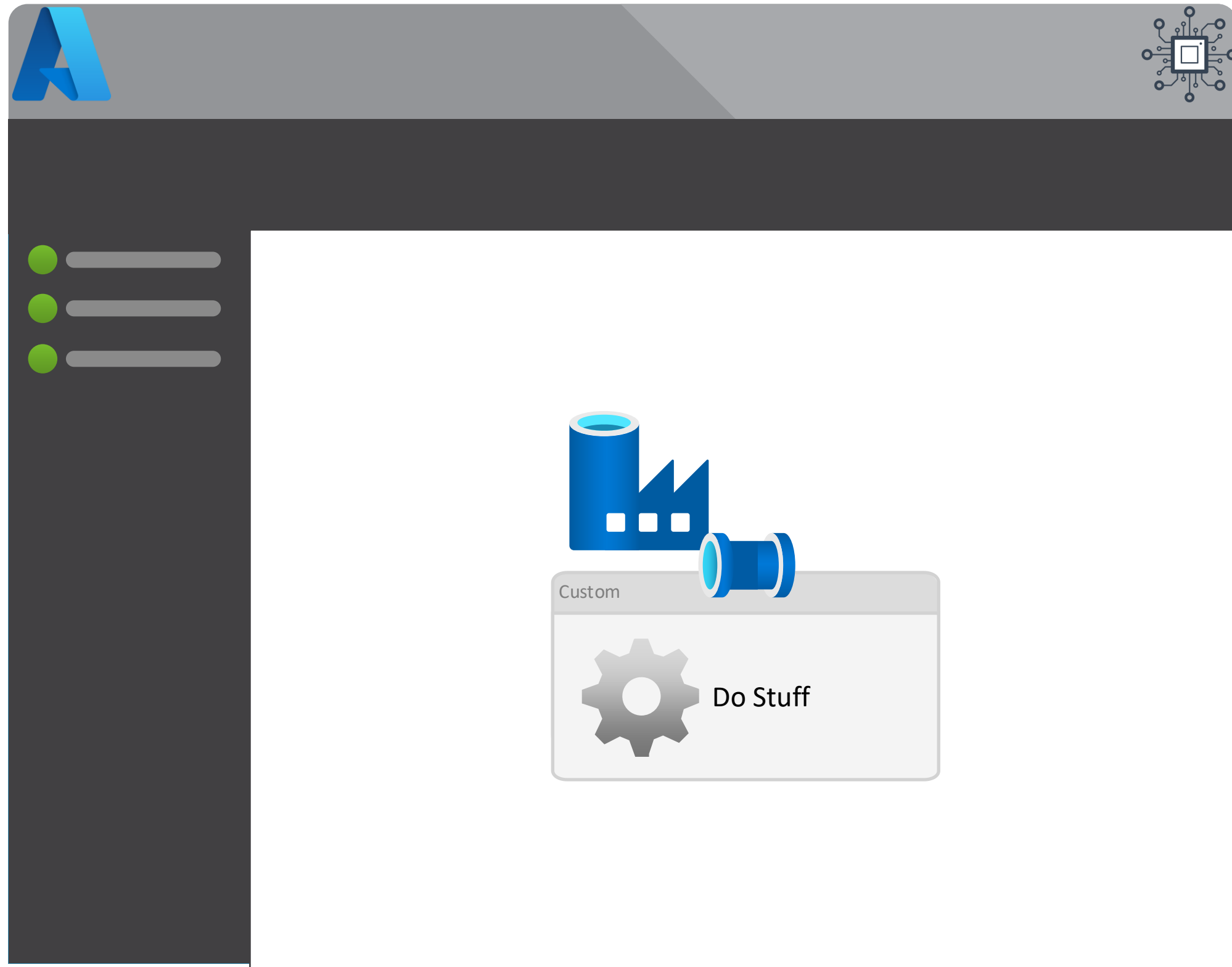- Azure Management API
- Azure Functions

Azure Batch Service

Jobs

Schedules

RDP

Pool 2 ←→ *fx*

Pool 1

Node Configuration

CPU

SSD

Pool 3

Module 6

Pipeline Extensibility

- Azure Batch Service
  - Tasks
  - Compute Pools
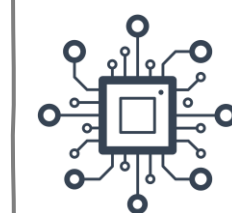 - Scaling
- Pipeline Custom Activities

Custom

Do Stuff

Custom

Extend Data Factory with Dedicated Scalable Compute and Custom Code

References Objects

Datasets: []

Linked Services: []

JSON JSON JSON JSON

EXE

Custom

LOG

JSON

EXE

Linked Services

Azure Batch

Azure Blob Storage

Do Stuff

Custom

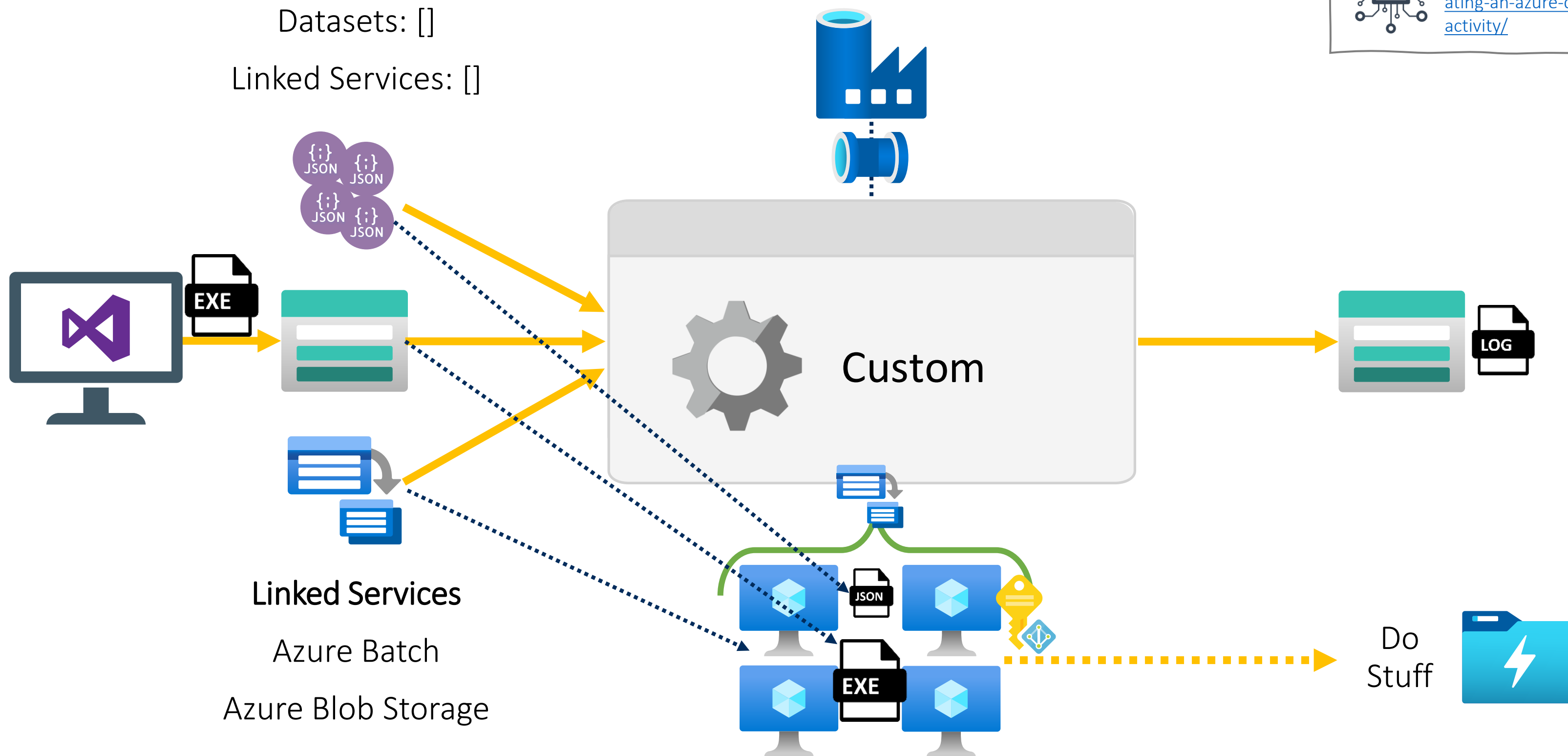## Extend Data Factory with Dedicated Scalable Compute and Custom Code
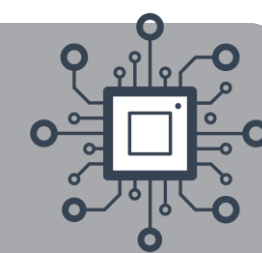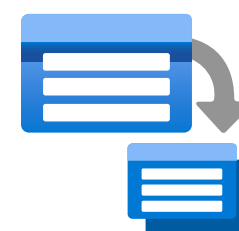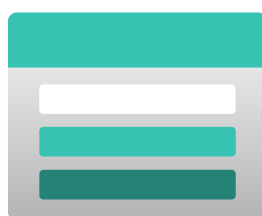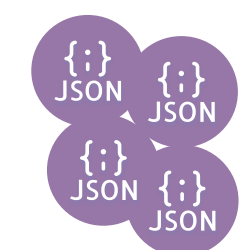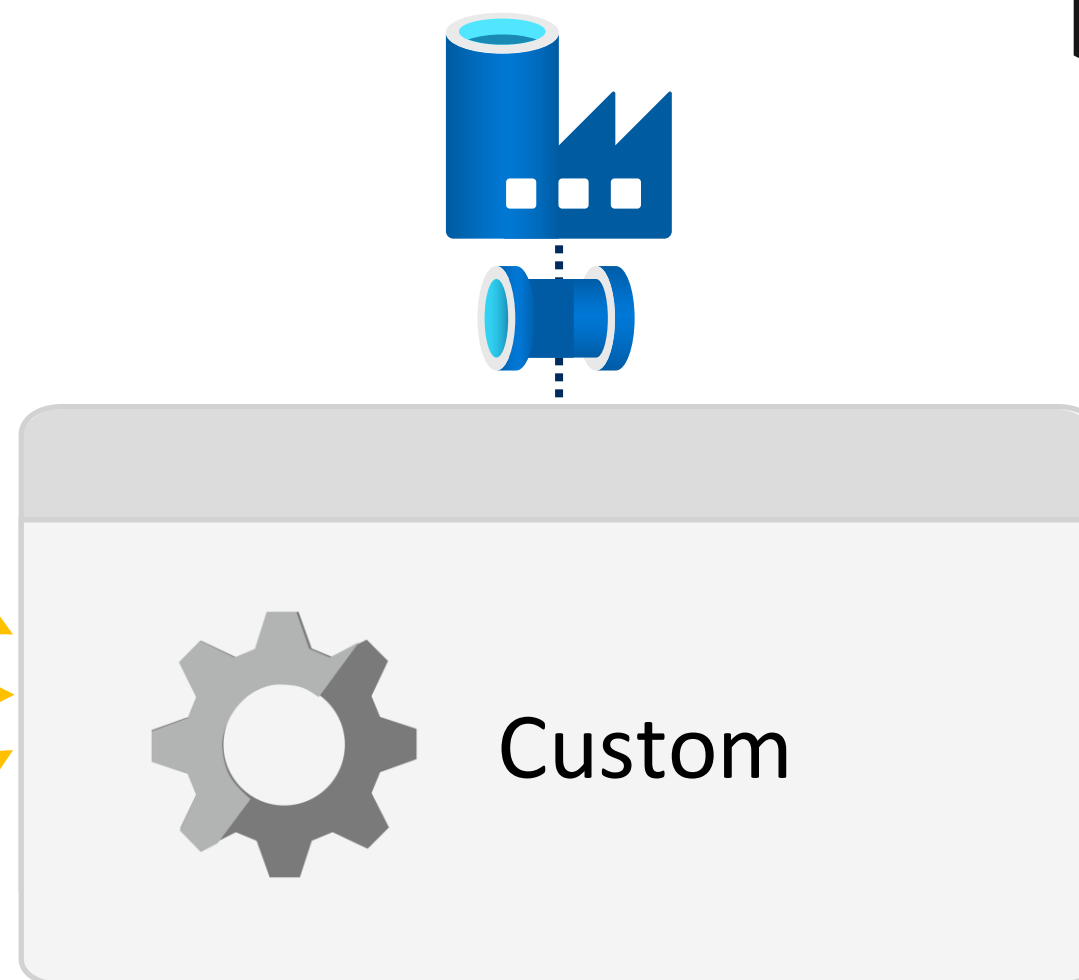
References Objects

Datasets: []

Linked Services: []

Custom

Linked Services

Azure Batch

Azure Blob Storage

New linked service

Azure Batch Learn more

Name *

BatchForTraining01

Description

Connect via integration runtime *

AutoResolveIntegrationRuntime

Authentication method *

Managed Identity

Managed identity name: **TrainingFactoryDev**
Managed identity object ID: **f818bfb8-5602-4e49-9bd5-81f3855cc81f**
Grant Data Factory service managed identity access to your Azure Batch.

Account name *

batchfortraining01

Batch URL *

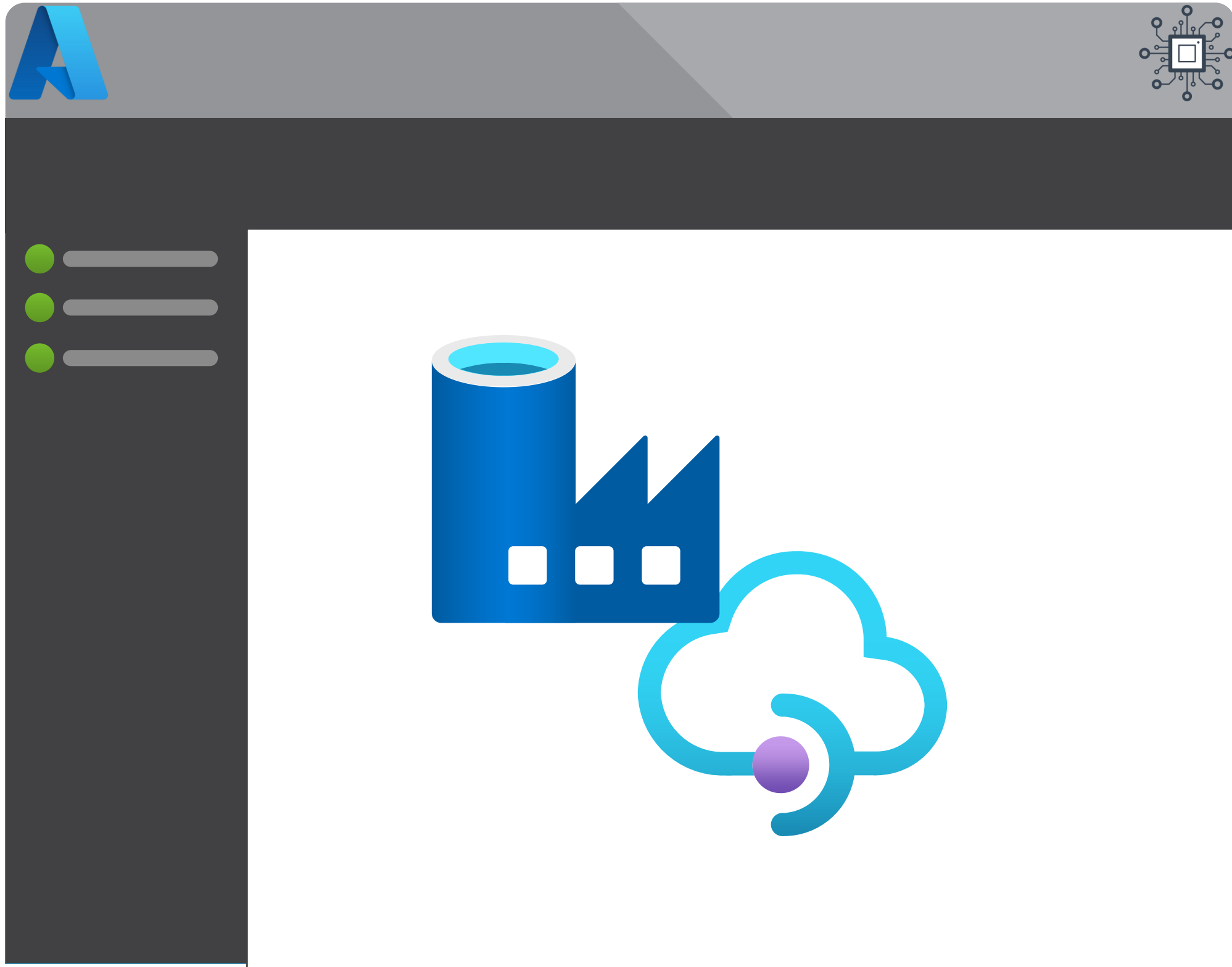https://batchfortraining01.uksouth.batch.azure.com

Pool name *

Pool ID

Storage linked service name *

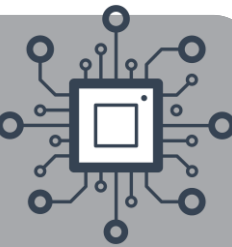TrainingStore01
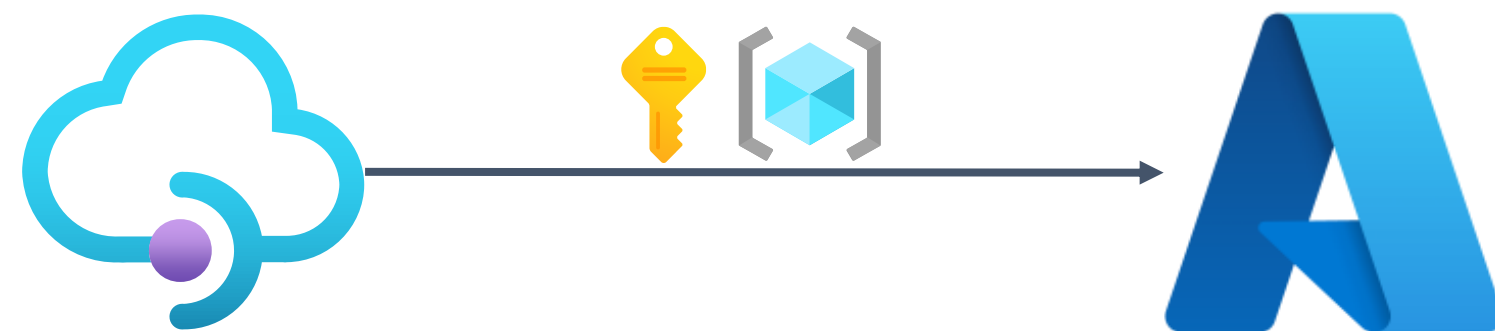
Module 5

Pipeline Extensibility

- Azure Batch Service
  - Tasks
  - Compute Pools
  - Scaling
- Pipeline Custom Activities
- **Azure Management API**
- Azure Functions

Management API's

Azure Management API

Make REST calls to interface with your Azure Resources

REST

Azure API Manager
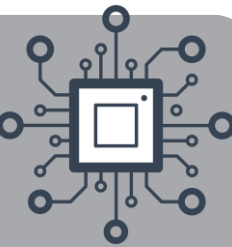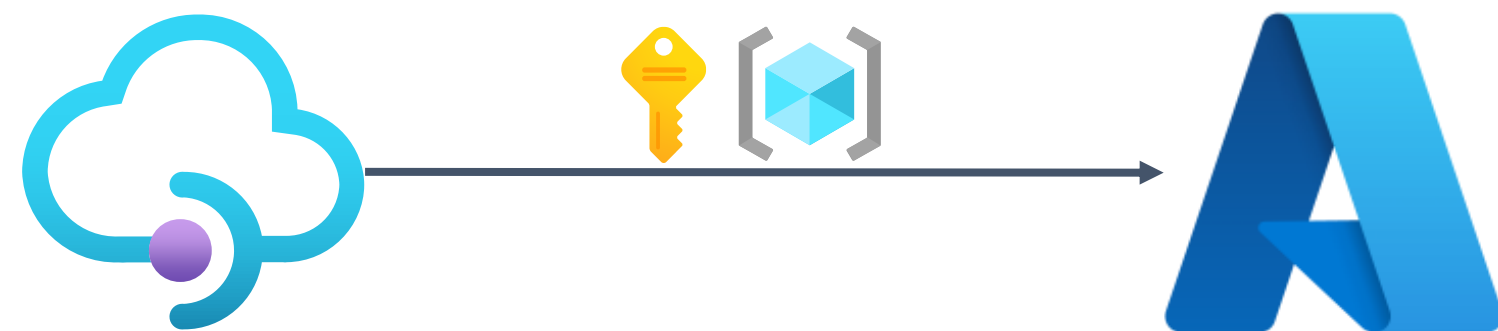
Build your own REST interface as an Azure Resource

REST

Azure Management API

Make REST calls to interface <u>with your</u> Azure Resources
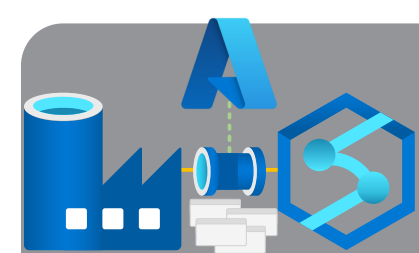
REST

Azure API Manager
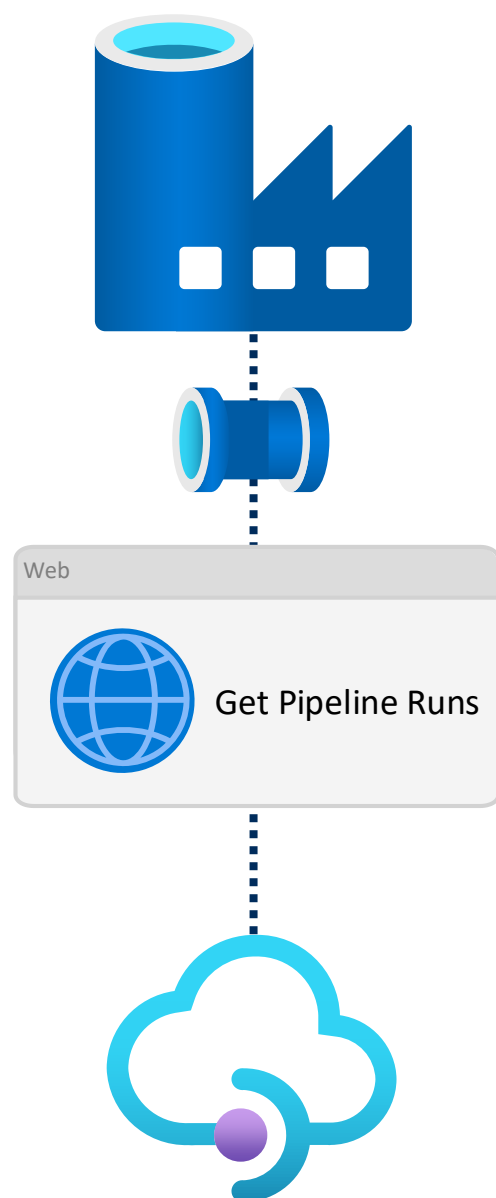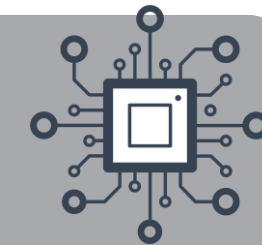
Build your own REST interface <u>as an</u> Azure Resource

REST

Orchestrate Everything

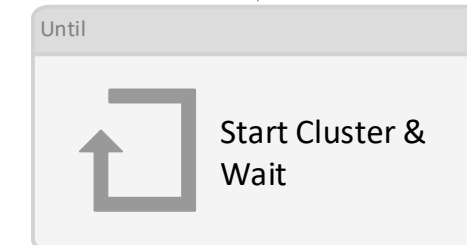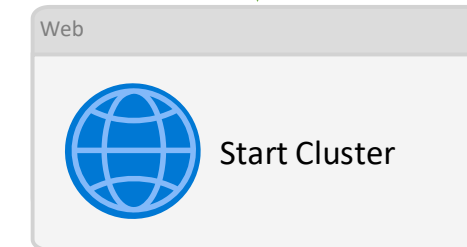**URL** https://management.azure.com/subscriptions/
@{**variables**('SubscriptionId')}/
resourceGroups/
@{**variables**('ResourceGroupName')}/
providers/
Microsoft.DataFactory/
factories/
@{**pipeline**().DataFactory}/
queryPipelineRuns?
api-version=2018-06-01

Get Pipeline Runs (Web)

Body
```
{
    "lastUpdatedAfter": "@{adddays(utcnow(),int(variables('QueryRunDays')))}",
    "lastUpdatedBefore": "@{utcnow()}",
    "filters": [
        {
            "operand": "PipelineName",
            "operator": "Equals",
            "values": [
                "@{pipeline().parameters.PipelineName}"
            ]
        }
    ]
}
```

Auth https://management.core.windows.net
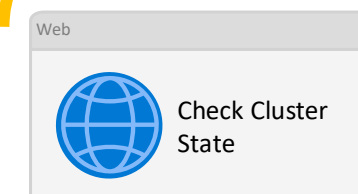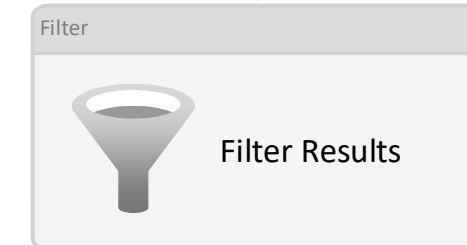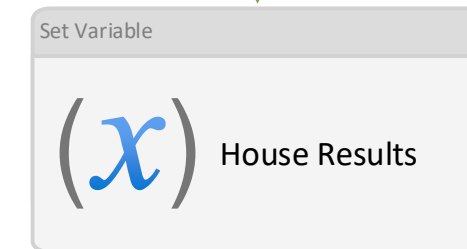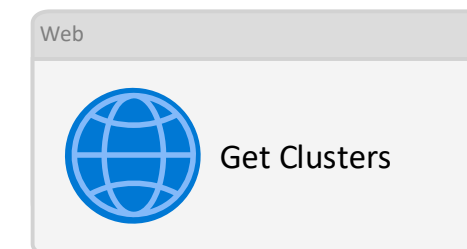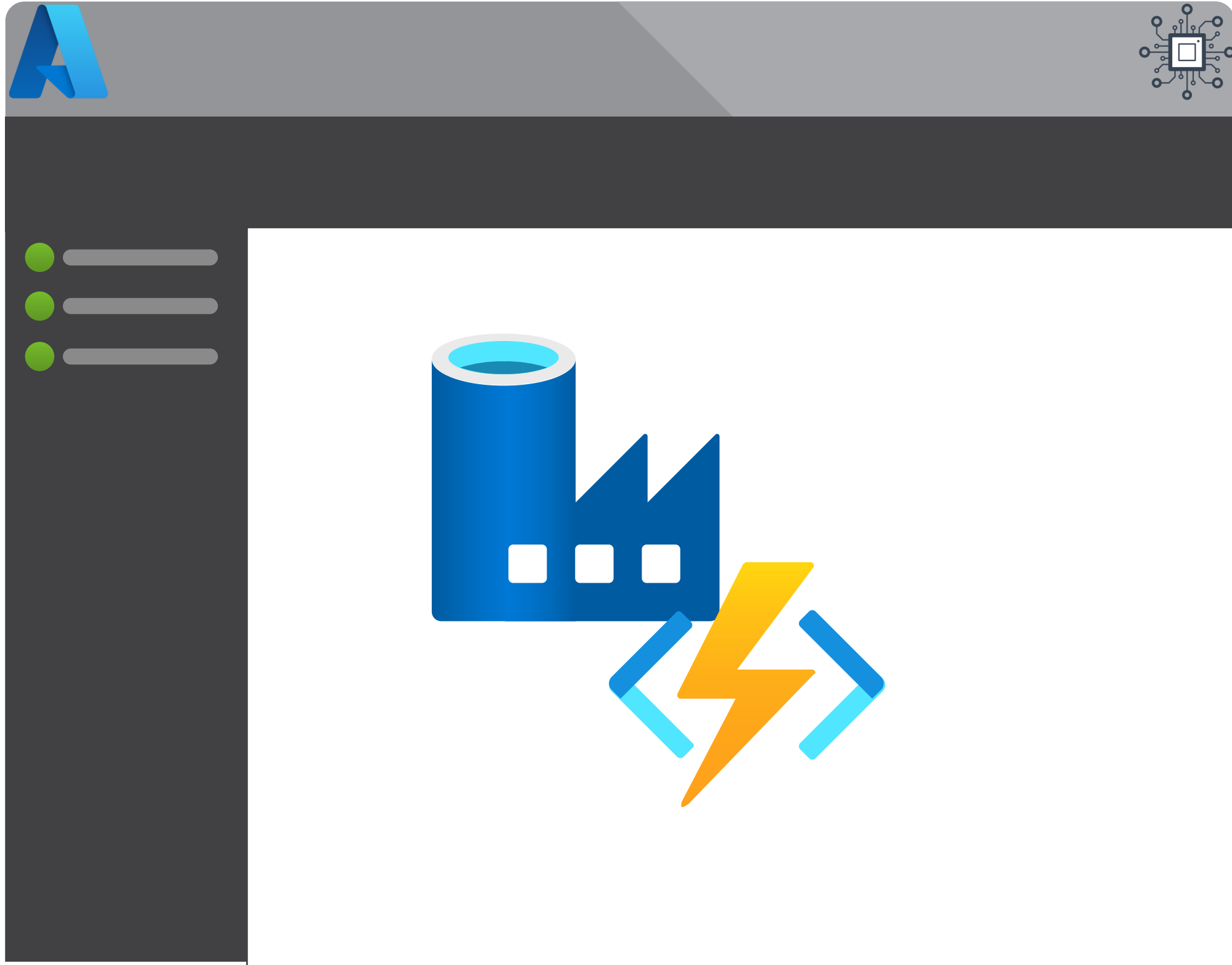
Managed Identity

Get Clusters (Web)

House Results (Set Variable)

Filter Results (Filter)

Start Cluster (Web)

Start Cluster & Wait (Until)

Check Cluster State (Web)

If Cluster Started – Break (If Condition)

Module 5

Pipeline Extensibility

- Azure Batch Service
  - Tasks
  - Compute Pools
  - Scaling
- Pipeline Custom Activities
- Azure Management API
- Azure Functions

Azure Functions App
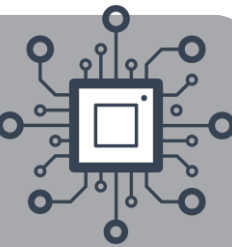
Serverless Scalable Bespoke Executions

Functions

Functions App

POSTMAN

C# .NET Core

.NET

Managed Identity Support

App Service Plan

VNet Integration

EXE   EXE   EXE

CPU CPU   CPU CPU   CPU CPU

CPU   CPU   CPU

App Service

Hosts

Scale Up & Scale Out

Azure Function

Azure Function

Do Stuff

Local Batch Data to Third Party Event Stream

Requirements:
- HR Sensitive Data
- JSON Message Stream
- Data Not Persisted
- No New Local Infrastructure

Hourly Trigger

HR System

Lookup
Get Row ID's

ForEach
Get and Send Data

Lookup Limit:
5,000 Rows or 4 MB

Lookup Output as
Function Body

{;} JSON

Lookup
Get Row Data

SELECT .... WHERE =
@item().RowID

Azure Function
Send Event Message

Azure Function
Send Email Alert

Third Party Absence Scoring Model

Traditional ETL Stack with Semantic Model

Daily Trigger

Requirements:
- Runtime Compute Scaling
- Model Partition Handling (Within Service Tier Limits)

Set Variable
(x) Set Run Date Values

Azure Function
Scale Up Database

Lookup
Get Fact Tables

ForEach
Process Facts

Azure Function
Scale Down Database

Users - Data Refreshed

Azure Function
Send Email Alert

Azure Function
Create Fact Partition

Azure Function
Process Partition

Azure Function
Delete Partition

Azure Function
Send Email Alert

Table Failure

SQL

Module 5

Pipeline Extensibility

```sql
SELECT
    [Contents]
FROM
    [Training]
WHERE
    [Module] = '5';

END; --module, fetch next
```
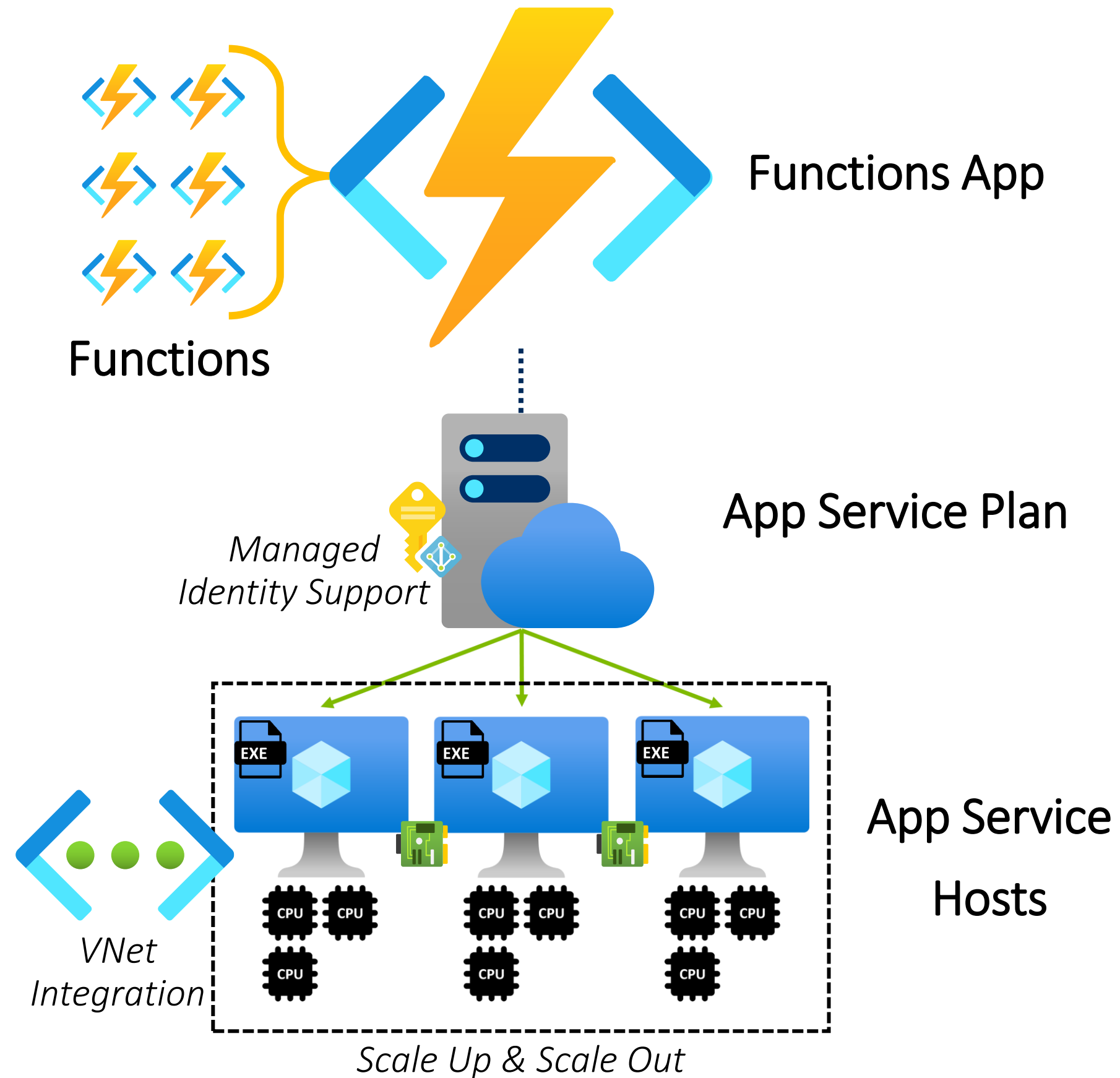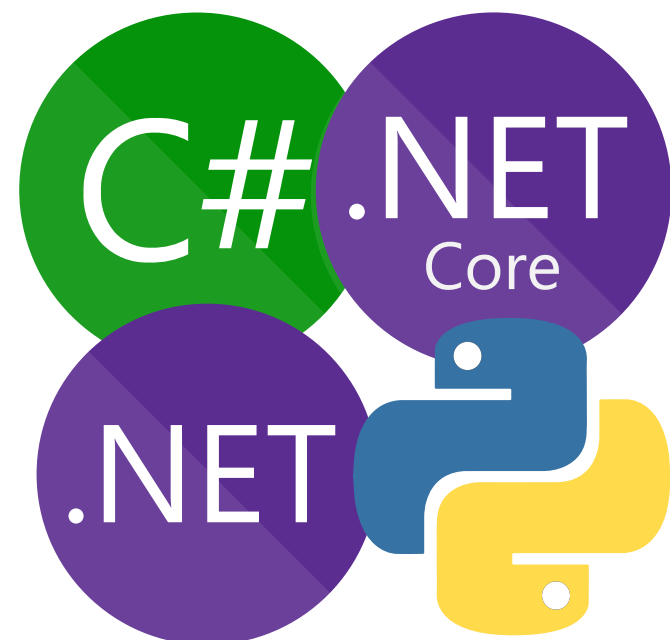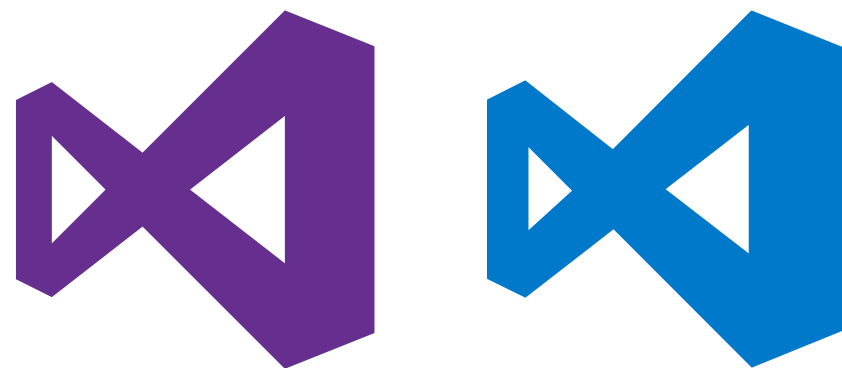
- Azure Batch Service
  - Tasks
  - Compute Pools
  - Scaling
- Pipeline Custom Activities
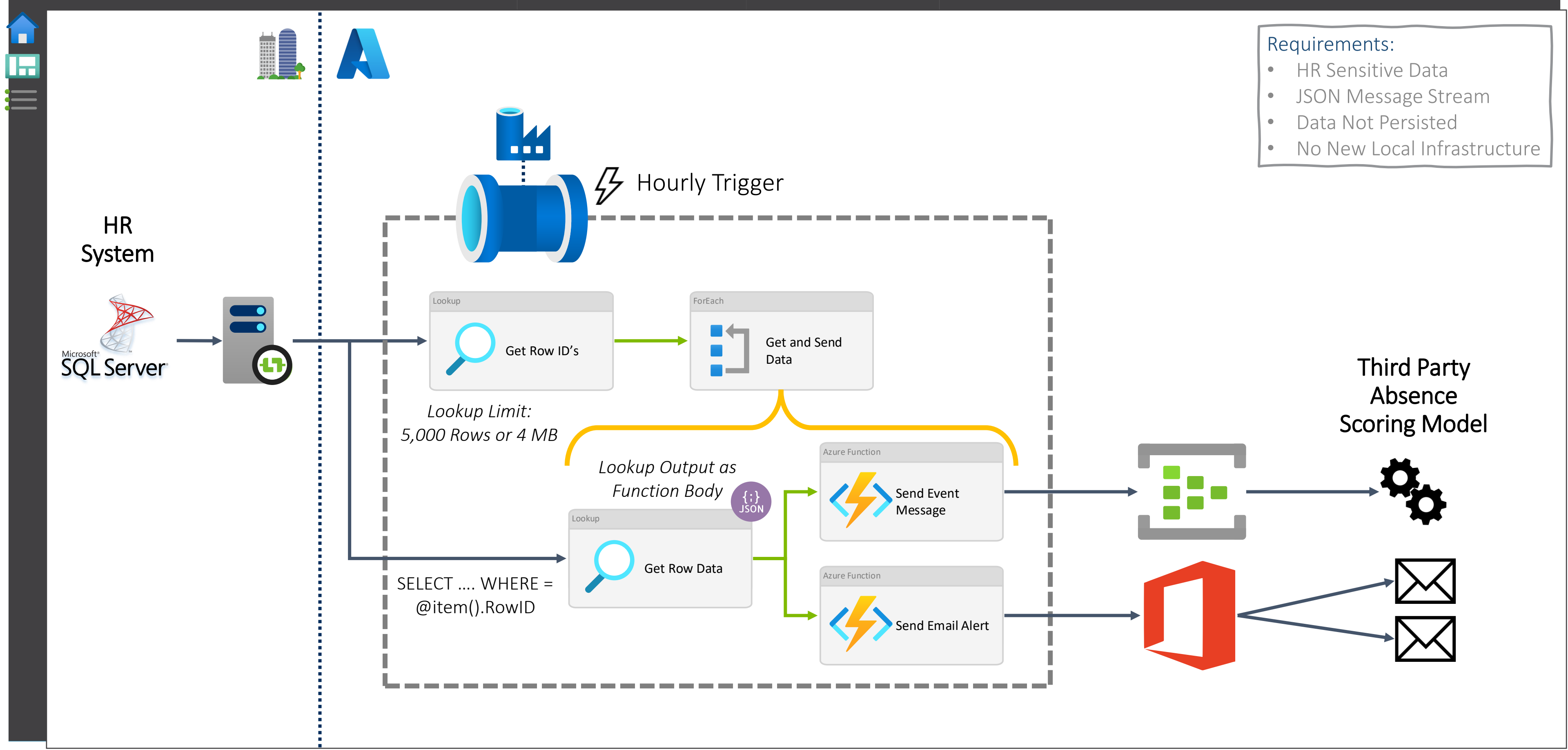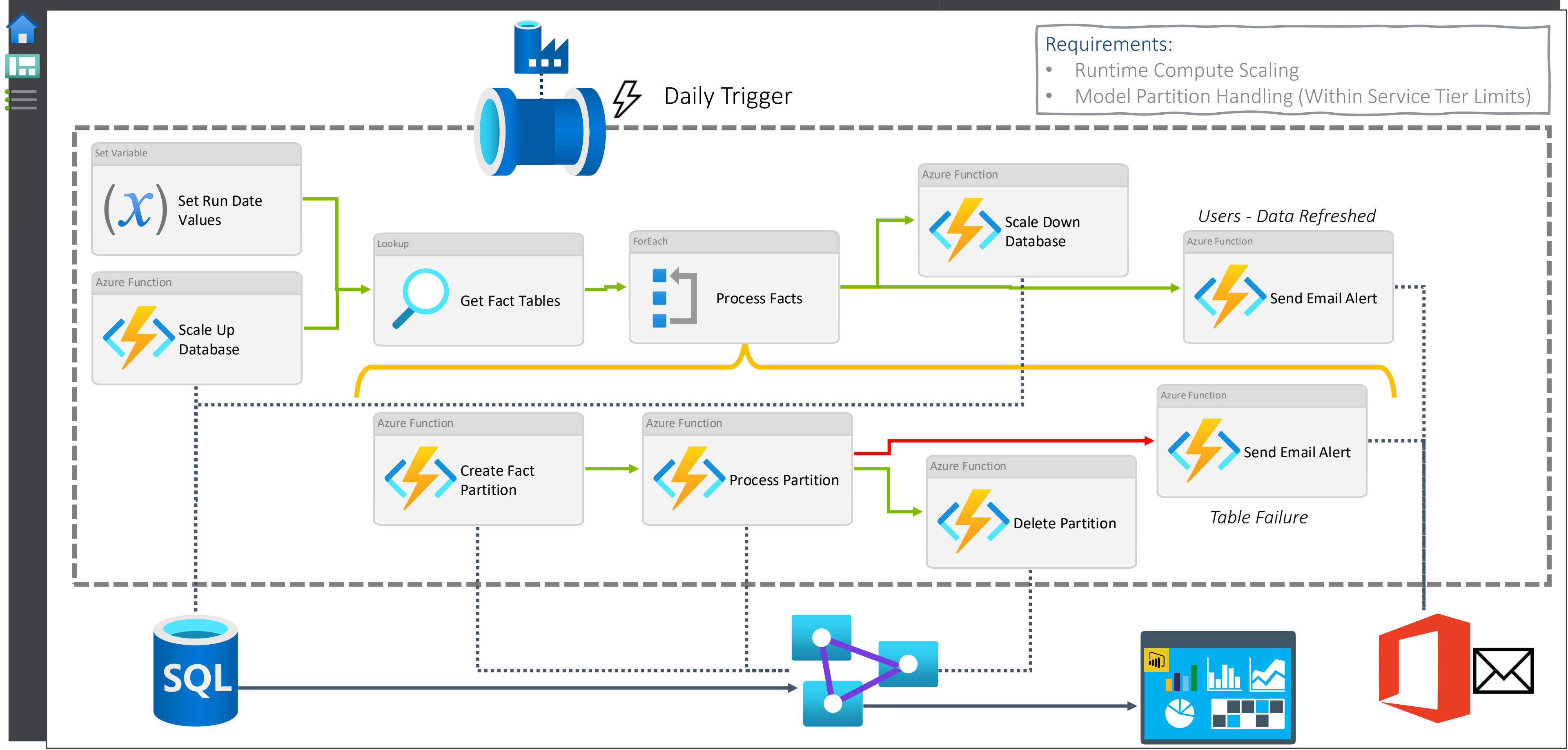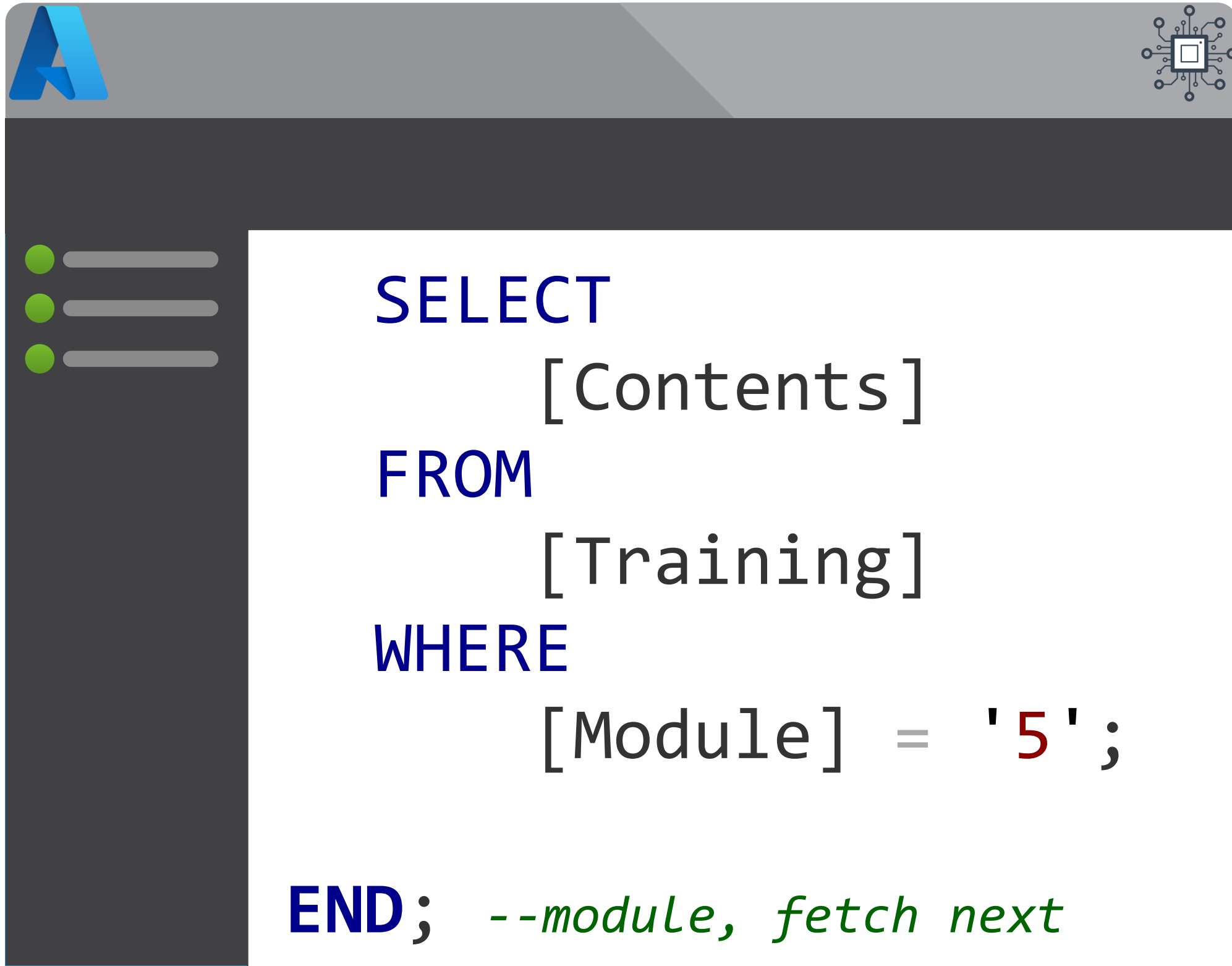- Azure Management API
- Azure Functions