Microsoft

Microsoft
Ready

Microsoft

Building data driven Azure Data Factory pipelines

Joshuha Owen

Have you seen ETL designs like this?

Dozens of connections

Hundreds of objects

Factory Resources

Filter resources by name

▷ Pipelines   0

▲ Datasets   **180**

    AzureSqlTable1
    AzureSqlTable2
    AzureSqlTable3
    AzureSqlTable4
    AzureSqlTable5
    AzureSqlTable6
    AzureSqlTable7
    AzureSqlTable8
    AzureSqlTable9
    AzureSqlTable10
    AzureSqlTable11
    AzureSqlTable12
    AzureSqlTable13

AzureSqlTable1	Connections	

Linked services   Integration runtimes

+ New

Showing 1 - 60 of **60** items

NAME	TYPE
AzureBlobStorage1	Azure Blob Storage
AzureBlobStorage2	Azure Blob Storage
AzureBlobStorage3	Azure Blob Storage
AzureBlobStorage4	Azure Blob Storage
AzureBlobStorage5	Azure Blob Storage
AzureBlobStorage6	Azure Blob Storage
AzureBlobStorage7	Azure Blob Storage
AzureBlobStorage8	Azure Blob Storage
AzureBlobStorage9	Azure Blob Storage
AzureBlobStorage10	Azure Blob Storage
AzureBlobStorage11	Azure Blob Storage
AzureBlobStorage12	Azure Blob Storage
AzureBlobStorage13	Azure Blob Storage
AzureBlobStorage14	Azure Blob Storage
AzureBlobStorage15	Azure Blob Storage
AzureBlobStorage16	Azure Blob Storage
AzureBlobStorage17	Azure Blob Storage
AzureBlobStorage18	Azure Blob Storage
AzureBlobStorage19	Azure Blob Storage
AzureBlobStorage20	Azure Blob Storage
AzureBlobStorage21	Azure Blob Storage
AzureBlobStorage22	Azure Blob Storage

Connection Mana

(project) CSV FIle 1	sourcefile2.csv	sourcefile2.csv 1	sourcefile2.csv 2
sourcefile2.csv 3	sourcefile2.csv 4	sourcefile2.csv 5	sourcefile2.csv 6
sourcefile2.csv 7	sourcefile2.csv 8	sourcefile2.csv 9	sourcefile2.csv 10
sourcefile2.csv 11	sourcefile2.csv 12	sourcefile2.csv 13	sourcefile2.csv 14

adf-dynamicpipelines-sql.database.windows.net.staging.joshuha   sql server 2

sql server 3   adf-dynamicpipelines-sql.database.windows.net.staging.joshuha 3

adf-dynamicpipelines-sql.database.windows.net.staging.joshuha 4   oracle server 1

oracle server 2   adf-dynamicpipelines-sql.database.windows.net.staging.joshuha 7

adf-dynamicpipelines-sql.database.windows.net.staging.joshuha 8

There is a better way!

USE METADATA TO DRIVE THE DESIGN

AZURE DATA FACTORY (ADF) IS BUILT
WITH THIS METADATA IN MIND

THINK IN PATTERNS

What's in it for me?

- Faster development time

- Less objects to maintain

- Make changes without touching the ADF code

- Patterns = Less Mistakes

Session learning objective

At the end of this session, you should be better able to...
- Design and implement data driven Azure Data Factory pipelines

Azure Data Factory terminology review

ADF Key Objects

Pipeline

Activity

Dataset

Linked Service

Parameter

ADF Key Objects

- Logical container
- Holds one or more activities
- Multiple parallel executions

Pipeline

Activity

Dataset

Linked Service

Parameter

ADF Key Objects

Activities

Search activities

◢ **Move & transform**

Copy data

Data flow

◢ **General**

x_+ Append variable

Delete

Execute Pipeline

Execute SSIS package

Get Metadata

Kusto

Lookup

Stored Procedure

$\{x\}$ Set variable

Pipeline

Activity

Dataset

Linked Service

Parameter

ADF Key Objects

- Actions performed on data
- Grouped into:
  - Data Movement Activities
  - Data Transformation Activities
  - Control Activities
- Takes in zero or more datasets
- Produces zero or more datasets

Pipeline

Activity

Dataset

Linked Service

Parameter
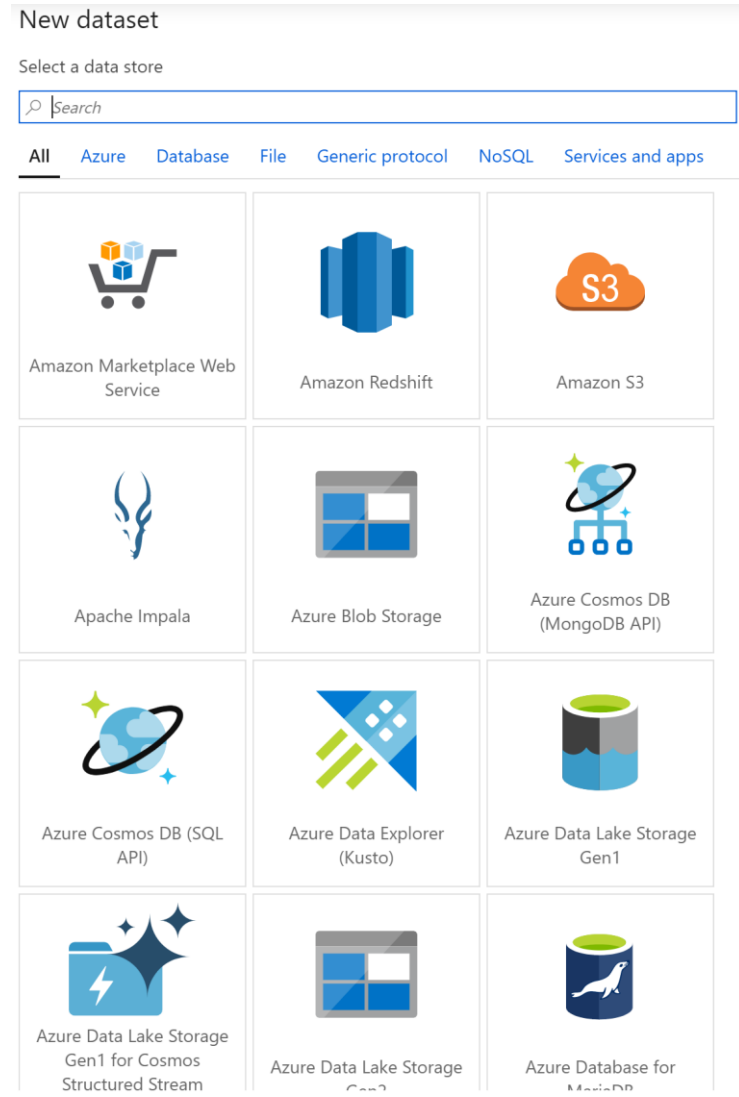
ADF Key Objects

Pipeline

Activity

Dataset

Linked Service

Parameter

ADF Key Objects

- Named reference to data
  - CSV file in Blob
  - Table in SQL Server
  - REST endpoint of Base URL
- Can be source and/or sink
- Can store schema

Pipeline

Activity

Dataset

Linked Service

Parameter
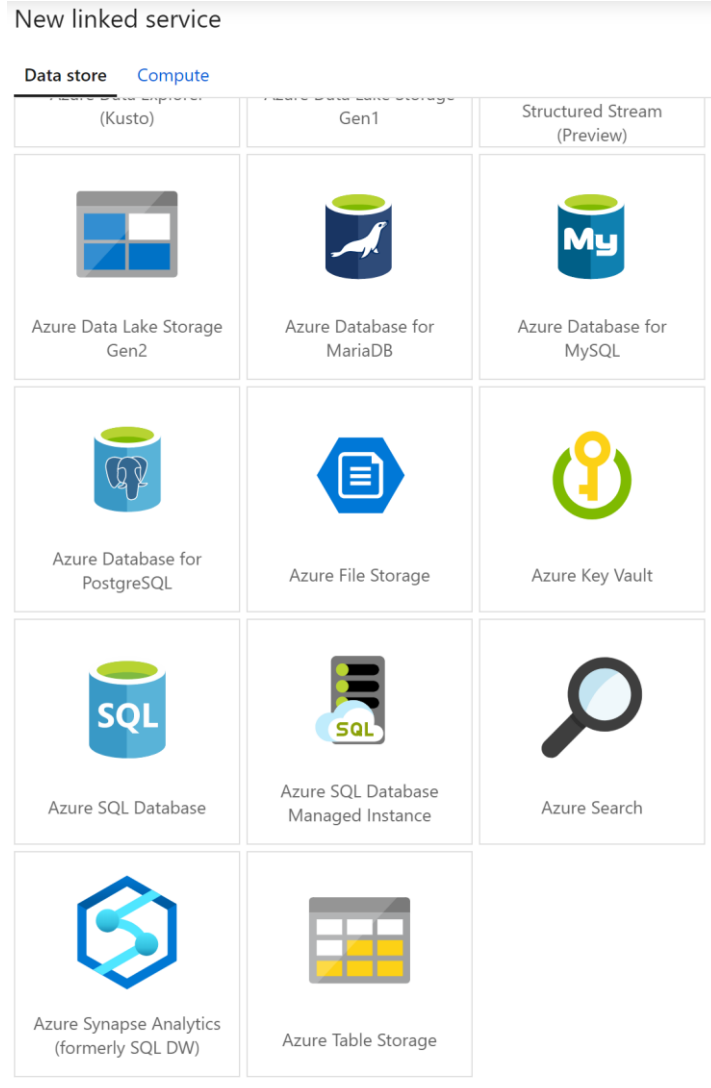
ADF Key Objects

Pipeline

Activity

Dataset

Linked Service

Parameter

ADF Key Objects

- Connections
- Authentication
- Data stores and Compute

Pipeline

Activity

Dataset

Linked Service

Parameter

ADF Key Objects

- Similar to variables
- Evaluated at run-time
- Can be used in
  - Linked Services
  - Datasets
  - Pipelines

Pipeline

Activity

Dataset

Linked Service

Parameter

ADF to SSIS comparisons

ADF

Pipeline ⟷ Package

Activity ⟷ Task

Dataset ⟷ Source/Destination

Linked Service ⟷ Connection Manager

SSIS
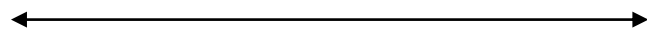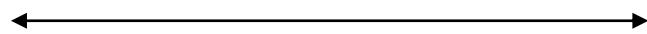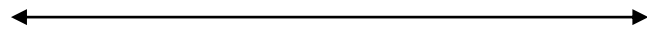
Parametrizing Linked Services

What is Azure Key Vault?

- Vault for storing:
  - Secrets
  - Certificates
  - Keys

- Centralized

- Monitoring

- Azure integration

ADF and Key Vault

- Store connection strings in Key Vault instead of ADF
- Easy to setup
- Promotes simplified DevOps processes
- Developers don't need credential passwords

Key Vault
Secret Example

BlobEndpoint=https://adfdynamicpipelinesblob.blob.core.windows.net/;QueueEndpoint=https://adfdynamicpipelinesblob.queue.core.windows.net/;FileEndpoint=https://adfdynamicpipelinesblob.file.core.windows.net/;TableEndpoint=https://adfdynamicpipelinesblob.table.core.windows.net/;SharedAccessSignature=SOMESASTOKENHERE

Key Vault Linked Service Configuration

…

"name": "LS_AzureKeyVault",

"type":
"Microsoft.DataFactory/factories/linkedservices",

"properties": {

    "type": "AzureKeyVault",

    "typeProperties": {

    "baseUrl": https://KV-DynamicPipelines-JO.vault.azure.net/

…

New linked service (Azure Key Vault)

Name *

AzureKeyVault1

Description

Azure key vault selection method

⦿ From Azure subscription   ◯ Enter manually

Azure subscription

MS Internal Sandbox ▼

Azure key vault name *

KV-DynamicPipelines-JO ▼

Edit key vault

Managed identity name: **adf-dynamicpipelines-complete-jo**
Managed identity object ID:
Managed identity application ID:
Grant Data Factory service managed identity access to your Azure Key Vault.

Annotations

＋ New

▷ Advanced ⓘ

Create   Back   🔌 Test connection   Cancel

Blob Linked Service Configuration

```
…
"name": "LS_AzureBlobStorage",
"properties": {
        "type": "AzureBlobStorage",
        "typeProperties": {
        "connectionString": {
        "type": "AzureKeyVaultSecret",
        "store": {
                "referenceName": "LS_AzureKeyVault",
                "type": "LinkedServiceReference"
        },
        "secretName": "Blob-Source"

…
```

New linked service (Azure Blob Storage)

ℹ️ If the identity you use to access the data store only has permission to subdirectory instead of the entire account, specify the path to test connection. Please make sure your self-hosted integration runtime is higher than version 4.0 if connecting via self-hosted integration runtime.

Name *

LS_AzureBlobStorage

Description

Connect via integration runtime *

AutoResolveIntegrationRuntime ▼

Authentication method

Account key ▼

Connection string	Azure Key Vault

AKV linked service *

LS_AzureKeyVault ▼

Edit connection

Secret name *

Blob-Source

Secret version

Use the latest version if left blank

Test connection

◉ To linked service   ○ To file path

Annotations

＋ New

▷ Advanced ℹ️

✓ Connection successful

| Create | Back | 🔌 Test connection | Cancel |

Blob Linked Service Parametrization

```
…
"name": "LS_AzureBlobStorageParam",
       "properties": {
       "type": "AzureBlobStorage",
       "parameters": {
                "secretName": {
                        "type": "string",
                        "defaultValue": "Blob-Source"
                }
       },
       "typeProperties": {
               "connectionString": {
                       "type": "AzureKeyVaultSecret",
                       "store": {
                               "referenceName": "LS_AzureKeyVault",
                               "type": "LinkedServiceReference"
                       },
               "secretName": "@linkedService().secretName"
…
```

Azure SQL Linked Service Configuration

```
…
"name": "LS_AzureSQLDB",
        "properties": {
        "type": "AzureSqlDatabase",
        "typeProperties": {
                "connectionString": {
                "type": "AzureKeyVaultSecret",
                "store": {
                "referenceName": "LS_AzureKeyVault",
                "type": "LinkedServiceReference"
        },
        "secretName": "SQLServerDB-metadata"
…
```

New linked service (Azure SQL Database)

Name *

AzureSqlDatabase1

Description

Connect via integration runtime *

AutoResolveIntegrationRuntime

| Connection string | Azure Key Vault |

AKV linked service *

LS_AzureKeyVault

Edit connection

Secret name *

SQLServerDB-metadata

Secret version

Use the latest version if left blank

Authentication type *

Sql Authentication or Managed Identity

Managed identity name: **adf-dynamicpipelines-complete-jo**
Managed identity object ID:
Managed identity application ID:
Grant Data Factory service managed identity access to your Azure SQL Database. Details

Annotations

+ New

▷ Parameters

▷ Advanced ⓘ

✓ Connection successful

Create   Back

Test connection   Cancel

Azure SQL Linked Service Parametrization

```
...
"name": "LS_AzureSQLDB",
"properties": {
        "parameters": {
                "secretName": {
                "type": "string",
                "defaultValue": "SQLServerDB-metadata"
                }
        },
"type": "AzureSqlDatabase",
"typeProperties": {
        "connectionString": {
                "type": "AzureKeyVaultSecret",
                "store": {
                "referenceName": "LS_AzureKeyVault",
                "type": "LinkedServiceReference"
                },
        "secretName": {
                "value": "@linkedService().secretName",
                "type": "Expression"
}
...
```

Edit linked service (Azure SQL Database)

LS_AzureSQLDB

Description

Connect via integration runtime *
AutoResolveIntegrationRuntime ▼

| Connection string | Azure Key Vault |

AKV linked service *
LS_AzureKeyVault ▼

Edit connection

Secret name *
@linkedService().secretName 🗑

Secret version
Use the latest version if left blank

Authentication type *
Sql Authentication or Managed Identity ▼

Managed identity name: **adf-dynamicpipelines-complete-jo**
Managed identity object ID:
Managed identity application ID:
Grant Data Factory service managed identity access to your Azure SQL Database. Details

Annotations
+ New

◢ Parameters
+ New | 🗑 Delete

	NAME	TYPE	DEFAULT VALUE
	secretName	String ▼	SQLServerDB-metadata

▷ Advanced ⓘ

Apply   🔌 Test connection   Cancel

Demo

Key Vault, Blob, and SQL Server Linked Services

Parametrizing Datasets

Remember this?

Factory Resources

Filter resources by name

▷ Pipelines   0

▲ Datasets   **180**

  ▦ AzureSqlTable1
  ▦ AzureSqlTable2
  ▦ AzureSqlTable3
  ▦ AzureSqlTable4
  ▦ AzureSqlTable5
  ▦ AzureSqlTable6
  ▦ AzureSqlTable7
  ▦ AzureSqlTable8
  ▦ AzureSqlTable9
  ▦ AzureSqlTable10
  ▦ AzureSqlTable11
  ▦ AzureSqlTable12
  ▦ AzureSqlTable13
  ▦ AzureSqlTable14
  ▦ AzureSqlTable15
  ▦ AzureSqlTable16
  ▦ AzureSqlTable17
  ▦ AzureSqlTable18
  ▦ AzureSqlTable19
  ▦ AzureSqlTable20
  ▦ AzureSqlTable21

✕ **Connections**

Linked services   Integration runtimes

+ New

Showing 1 - 60 of **60** items

NAME ↑↓	TYPE ↑↓
AzureBlobStorage1	Azure Blob Storage
AzureBlobStorage2	Azure Blob Storage
AzureBlobStorage3	Azure Blob Storage
AzureBlobStorage4	Azure Blob Storage
AzureBlobStorage5	Azure Blob Storage
AzureBlobStorage6	Azure Blob Storage
AzureBlobStorage7	Azure Blob Storage
AzureBlobStorage8	Azure Blob Storage
AzureBlobStorage9	Azure Blob Storage
AzureBlobStorage10	Azure Blob Storage
AzureBlobStorage11	Azure Blob Storage
AzureBlobStorage12	Azure Blob Storage
AzureBlobStorage13	Azure Blob Storage
AzureBlobStorage14	Azure Blob Storage
AzureBlobStorage15	Azure Blob Storage
AzureBlobStorage16	Azure Blob Storage
AzureBlobStorage17	Azure Blob Storage
AzureBlobStorage18	Azure Blob Storage
AzureBlobStorage19	Azure Blob Storage
AzureBlobStorage20	Azure Blob Storage
AzureBlobStorage21	Azure Blob Storage
AzureBlobStorage22	Azure Blob Storage

Blob File Dataset Configuration

```
…
"linkedServiceName": {
  "referenceName":
  "LS_AzureBlobStorageParam",
  "type": "LinkedServiceReference",
    "parameters": {
      "secretName": "Blob-Source"
    }
},
…
"typeProperties": {
  "location": {
    "type": "AzureBlobStorageLocation",
    "fileName": "sourcefile1.csv",
    "container": "source"
  },
…
```

General   **Connection**   Schema   Parameters

Linked service *   🖼 LS_AzureBlobStorageParam ▾   🔌 Test connection   ✏ Op

▲ Linked service properties ⓘ

NAME	VALUE
secretName	Blob-Source

File path *   source   /   Directory   /   sourcefile1.csv

Compression type   none ▾

Column delimiter   Comma (,) ▾
                        ☐ Edit

Row delimiter   Auto detect (\r,\n, or \r\n) ▾
                     ☐ Edit

Encoding   Default(UTF-8) ▾

Escape character   Backslash (\) ▾
                        ☐ Edit

Blob File Dataset Parameterization

```
…
"location": {
"type": "AzureBlobStorageLocation",
  "fileName": {
    "value": "@dataset().filename",
    "type": "Expression"
  },
  "folderPath": {
    "value": "@dataset().folderpath",
    "type": "Expression"
  },
  "container": {
    "value": "@dataset().container",
    "type": "Expression"
  }
…
```

General   Connection   Schema   **Parameters**

+ New | 🗑 Delete

	NAME	TYPE	DEFAULT VALUE
☐	container	String ▼	source
	folderpath	String ▼	files
	filename	String ▼	sourcefile1.csv

SQL Dataset Configuration

```json
…
"linkedServiceName": {
  "referenceName": "LS_AzureSQLDB",
  "type": "LinkedServiceReference",
  "parameters": {
    "secretName": "SQLServerDB-
metadata"
  }
}
…
"typeProperties": {
  "schema": "dbo",
  "table": "ConfigADF"
}
…
```

General	Connection	Schema	Parameters

Linked service *   SQL LS_AzureSQLDB ▼

◢ Linked service properties ⓘ

NAME	VALUE
secretName	SQLServerDB-metadata

Table   dbo.ConfigADF ▼

☐ Edit

SQL Dataset Parameterization

```
…
"typeProperties": {
  "schema": {
    "value": "@dataset().schema",
    "type": "Expression"
  },
  "table": {
    "value": "@dataset().tablename",
    "type": "Expression"
  }
}
…
```

Linked service * | SQL | LS_AzureSQLDB ▼ | 🔗 Test connection

▲ Linked service properties ⓘ

NAME	VALUE
secretName	SQLServerDB-metadata

Table   @dataset().schema 🗑 . @dataset().tablename
                ☑ Edit

Demo

Blob Storage CSV and SQL Table Dataset

Data Driven Pipelines

Activity Review

- Gets data from a dataset
- Access from ADF expressions
- Return first or all rows

Lookup

ForEach

Copy data

Activity Review

· Iterate over a collection
· Specify inner activities
· Can run in parallel (up to 50)

Lookup

ForEach

Copy data

Activity Review

- Move data between source and sink
- Source and sink datasets
- Can infer schema

Lookup

ForEach

Copy data

Example Metadata Table

srcContainer	srcFolder	srcFilename	srcSecret	dstSchema	dstTable	dstSecret
source	files	sourcefile1.csv	Blob-Source	dbo	Table1	SQLServerDB-Staging
source	files	sourcefile2.csv	Blob-Source	dbo	Table2	SQLServerDB-staging
source	files	sourcefile3.csv	Blob-Source	dbo	Table3	SQLServerDB-staging

Lookup Activity Configuration

```json
…
"typeProperties": {
  "source": {
  "type": "AzureSqlSource",
  "queryTimeout": "02:00:00"
},
"dataset": {
  "referenceName": "DS_AzSQLTable",
  "type": "DatasetReference",
  "parameters": {
    "schema": "dbo",
    "tablename": "ConfigADF",
    "secretname": "SQLServerDB-metadata"
  }
},
"firstRowOnly": false
…
```

General   **Settings**   User properties

Source dataset *   DS_AzSQLTable

◢ Dataset properties ⓘ

NAME	VALUE
schema	dbo
tablename	ConfigADF
secretname	SQLServerDB-metadata

Use query   ⦿ Table   ◯ Query   ◯ Stored Procedure

Query timeout (minutes)   120   ⓘ

First row only   ☐

Lookup Activity Output

```json
{
  "count": 3,
  "value": [
    {
      "configKey": 1,
      "sourceContainer": "source",
      "sourceFolderPath": "files",
      "sourceFilename": "sourcefile1.csv",
      "sourceConnectionSecret": "Blob-Source",
      "destinationSchema": "dbo",
      "destinationTable": "Table1",
      "destinationConnectionSecret": "SQLServerDB-Staging",
      "isActive": true
    },
    {
      "configKey": 2,
      …
```

ForEach Activity Configuration

```
…
"typeProperties": {
  "items": {
  "value": "@activity('LKP Get Driver
Data').output.value",
  "type": "Expression"
  }
}
…
```

General **Settings** Activities(0) User properties

Sequential ☐

Batch count [ ] 🛈

Items @activity('LKP Get Driver
Data').output.value 🗑

Copy Activity Source Configuration

```
…
"inputs": [
{
  "referenceName": "DS_BlobCSV",
  "type": "DatasetReference",
  "parameters": {
    "container": {
      "value": "@item().sourceContainer",
      "type": "Expression"
    },
    "folderpath": {
      "value": "@item().sourceFolderPath",
      "type": "Expression"
    },
    "filename": {
      "value": "@item().sourceFilename",
      "type": "Expression"
    },
    "secretname": {
      "value": "@item().sourceConnectionSecret",
…
```

Source dataset * | 📄 DS_BlobCSV ▼ | ✏ Open

◢ Dataset properties ⓘ

NAME	VALUE
container	@item().sourceContainer
folderpath	@item().sourceFolderPath
filename	@item().sourceFilename
secretname	@item().sourceConnectionSecret

Copy Activity Sink Configuration

```
…
"outputs": [
{

  "referenceName": "DS_AzSQLTable",
  "type": "DatasetReference",
  "parameters": {
  "schema": {
    "value": "@item().destinationSchema",
    "type": "Expression"
  },
  "tablename": {
    "value": "@item().destinationTable",
    "type": "Expression"
  },
  "secretname": {
    "value": "@item().destinationConnectionSecret",
…
"sink": {
  "type": "AzureSqlSink",
  "preCopyScript": {
    "value": "DROP TABLE IF EXISTS
@{item().destinationSchema}.@{item().destinationTable}"
…
```

Sink dataset * [🗄 DS_AzSQLTable ▼] ✏ Open

◢ Dataset properties ❶

NAME	VALUE
schema	@item().destinationSchema
tablename	@item().destinationTable
secretname	@item().destinationConnectionSecret

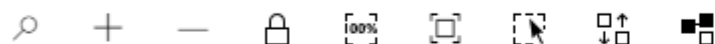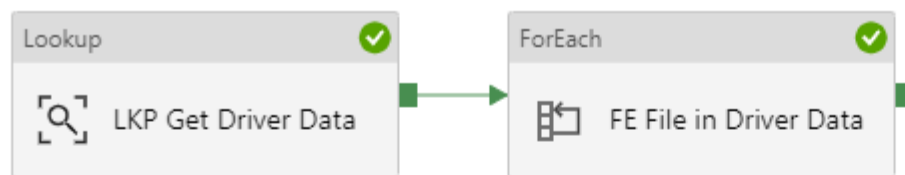Stored procedure name [Select... ▼] ↻ Refresh
☐ Edit ❶

Table option ◯ None ⦿ Auto create table ❶

Pre-copy script
```
DROP TABLE IF EXISTS
@{item().destinationSchema}.@{item().desti
nationTable}
```

Tying It All Together

Demo

Data Driven Pipelines

Session takeaways

· Azure Key Vault and Azure Data Factory make a great pairing
· Parameters and metadata make a great pairing

· Go design your metadata drivers, start simple!
· Build your first dynamic Azure Data Factory pipelines!

Session resources

- Demos and slides:
  - https://github.com/joshuha/DataDrivenADF
- ADF Hands-On Lab:
  - http://aka.ms/adflab2
- Azure Data Platform End-to-End:
  - https://github.com/fabragaMS/ADPE2E
- ADF Youtube:
  - https://www.youtube.com/channel/UC2S0k7NeLcEm5_IhHUwpN0g
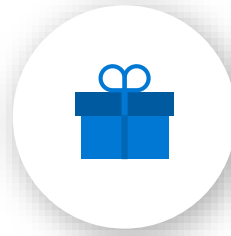- Microsoft Ready content can be found at www.microsoftready.com.
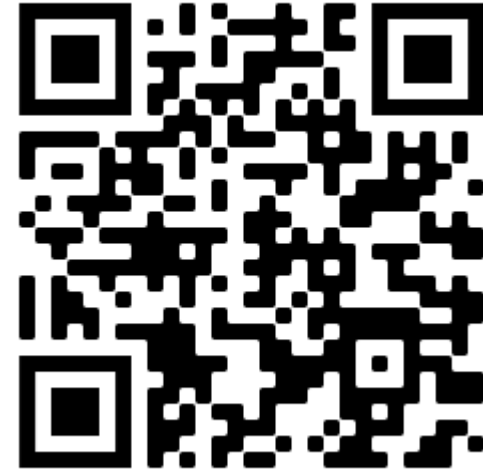
Q&A

If you have questions please proceed to the
Q&A MICROPHONE located in your session room.

Enjoy a session

Fill out an evaluation

Win a prize

Your input matters

Evaluations can be submitted via:

Microsoft Events App – Look for 'My evaluations'

Laptop/Mobile browser – Connect to www.microsoftready.com/evaluations

Download the mobile app at https://myready.microsoft.com/app

Submit to win **prizes** and **VIP experiences** at **Ready Seattle**!
Details at http://aka.ms/ReadyWinners

If you plan to remain in this room for the next session, please be sure to
scan your badge again at the door prior to the session starting