



# Day 3

## **Azure Data Factory Masterclass**



# About Me



**Alpa Buddhhabhatti**



/alpaBuddhabhatti/



/alpabuddhabhatti/



/AlpaB7



/@alpabuddhabhatti



/@meetalpha



**Microsoft Certified:  
Azure Data Engineer  
Associate**

Microsoft



**Microsoft Certified:  
Azure Data Scientist  
Associate**

Microsoft



**Microsoft Certified:  
Azure Developer  
Associate**

Microsoft



**Microsoft Certified  
Trainer 2021-2022**

Microsoft



# Agenda

1. *Day 2 Recap*
2. *Expressions & Functions*
3. *Azure Data Factory more Activities*
  - foreach loop activity*
  - Lookup Activity*
  - if activity*
  - Data flow Activity*
2. *Labs*
  - Creating pipeline with If activity*
  - Creating pipeline with Look up activity*
  - Creating pipeline with Data Flow activity*
3. *Conclusions*



## *Day 2 Recap*

Dynamic solutions



# *Expressions & Functions*

Expressions start with '@'

@(toUpper(pipeline.parameters.BO\_NAME)

@(concat(pipeline.parameters.BO\_NAME, '-in')

@(equals(pipeline.parameters.BO\_NAME, 'bo1')



# *Expressions & Functions*

## Option 1 – 'concat' Function

```
@concat(  
    ' TRUNCATE TABLE dbo. ',  
    pipeline().parameters.TableName  
)
```

## Option 2 – String Interpolation

```
TRUNCATE TABLE dbo.  
@{pipeline().parameters.TableName}
```



# *Expressions & Functions*

or()	
and()	&&
equals()	==
greater()	>
greaterOrEqual()	>=
lesser()	<
lesserOrEqual()	<=
notEquals()	!=



## *ADF – Foreach Loop Activity*

Use a ForEach loop to execute the same set of activities or pipelines multiple times with different values each time

The ForEach loop iterates over a *collection*:

- Array
- JSON Object





## *ADF – Foreach Loop Activity for Array*

Reference the current iteration using **@item()**

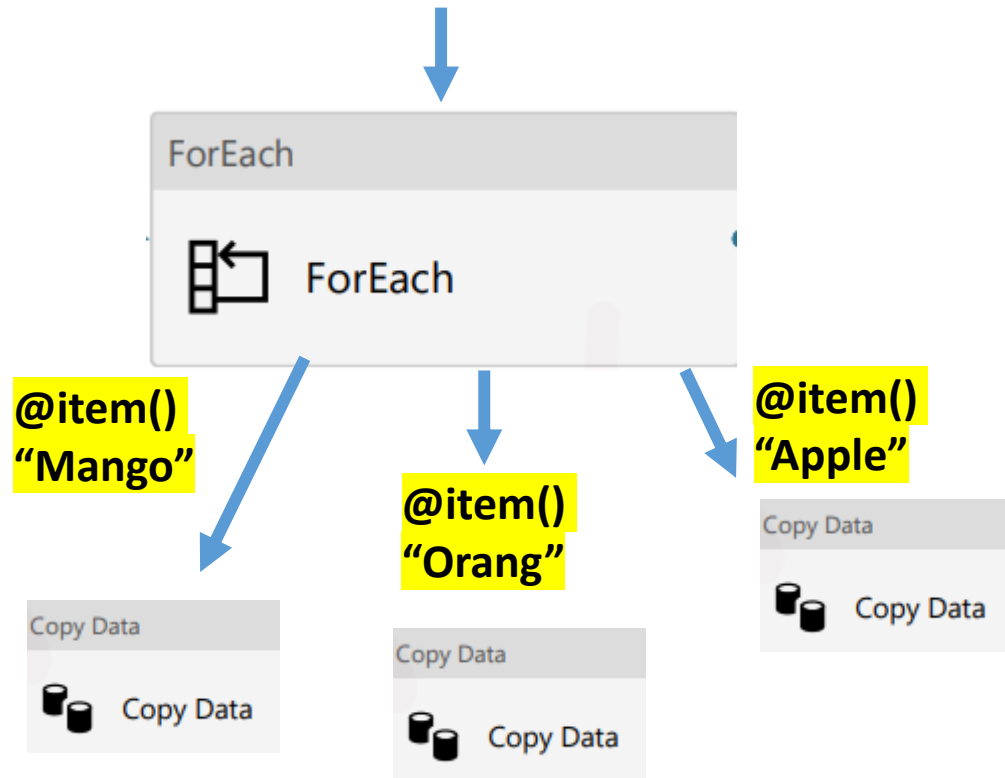
**Array Items : ["Mango", "Orange", "Apple"]**



**@item()**

# ADF – Foreach Loop Activity for Array Item

["Mango", "Orange", "Apple"]



## *ADF – Foreach Loop Activity for Object*

### JSON Object:

```
{ "Emp": [  
    {"Name" : "Alpa", "EmailId" : "abc@gmail.com"} @item()  
    {"Name" : "TestABC", "EmailId" : "testabc@gmail.com"}  
  ] }
```



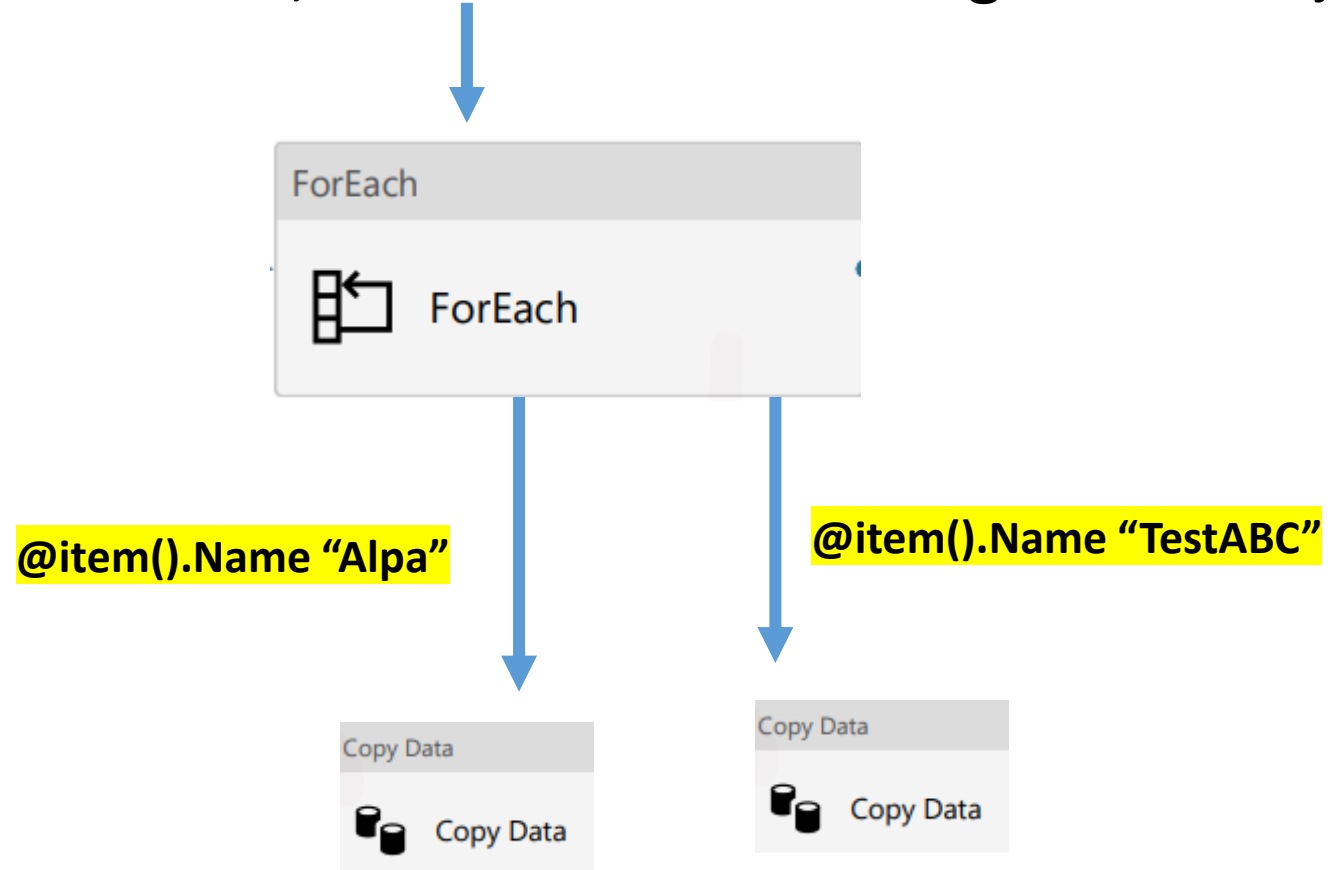
@item().Name

```
{ "Emp": [
```

```
  {"Name" : "Alpa", "EmailId" : "abc@gmail.com"} @item()
```

```
  {"Name" : "TestABC", "EmailId" : "testabc@gmail.com"}  
]
```

```
}
```



## Foreach loop limitations

Nested ForEach Loops are not supported

Maximum number of items: **100 000**

Maximum number of parallel executions: **50**

Default number of parallel executions: **20**



## *ADF – Lookup Activity*

Retrieve dataset from:

- Configuration File / Table
- Query
- Stored Procedure

Use to dynamically get objects to operate on in subsequent activity



## *ADF – Lookup Activity*


Two types of outputs:

- First Row Only
- All Rows



## *ADF – Lookup Activity With First Row only*

```
{  
  "firstRow" :  
  {  
    "Column1" : "Value"  
    "Column2" : "Value"  
  }  
}
```



```
@{activity('Lookup')  
  .output.firstRow  
  .Column1}
```

Use to pass single values to another activity

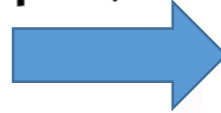
```
@{activity('Lookup').output.firstRow.Column1}
```





## *ADF – Lookup Activity with as output All rows*

```
@{activity('Lookup')  
  .output.value}
```



```
{  
  "count" : "2",  
  "value" : [  
    {  
      "Column1" : "Value",  
      "Column2" : "Value"  
    },  
    {  
      "Column1" : "Value",  
      "Column2" : "Value"  
    }  
  ]  
}
```

Use to pass entire collection to Foreach Loop

```
@{activity('Lookup').output.value}
```



## *ADF – Lookup Activity - All rows*

@{activity( 'Lookup' ).output.**value[1]**}



```
{  
  "count" : "2",  
  "value" : [  
    {  
      "Column1" : "Value",  
      "Column2" : "Value"  
    },  
    {  
      "Column1" : "Value",  
      "Column2" : "Value"  
    }  
  ]  
}
```

**Second row**





## *ADF – Lookup Activity - All rows*

@{activity( 'Lookup' ).output.value[1].Column1}



```
{  
  "count" : "2",  
  "value" : [  
    {  
      "Column1" : "Value",  
      "Column2" : "Value"  
    },  
    {  
      "Column1" : "Value",  
      "Column2" : "Value"  
    }  
  ]  
}
```

2nd row's column1



## *ADF – If Activity*

- It gives two options true and false
- When condition is true, you can add activities related to true condition
- When condition is false, you can add activities related to false condition
- Nested If is not possible using “if activity” (“If activity” inside “if activity”)
- You can replace “if activity” with expression & Functions

# Get Metadata Activity

It is similar to Lookup Activity.

Based on Dataset you have used in activity, you will get output with list of folder OR list of file OR Rows from a particular file

It often use with foreach loop

E.g

- Get list of files from a specific container (DS is pointing to container)

- Get rows from a specific file (DS is pointing to specific file)



## *ADF – Dataflow activity*

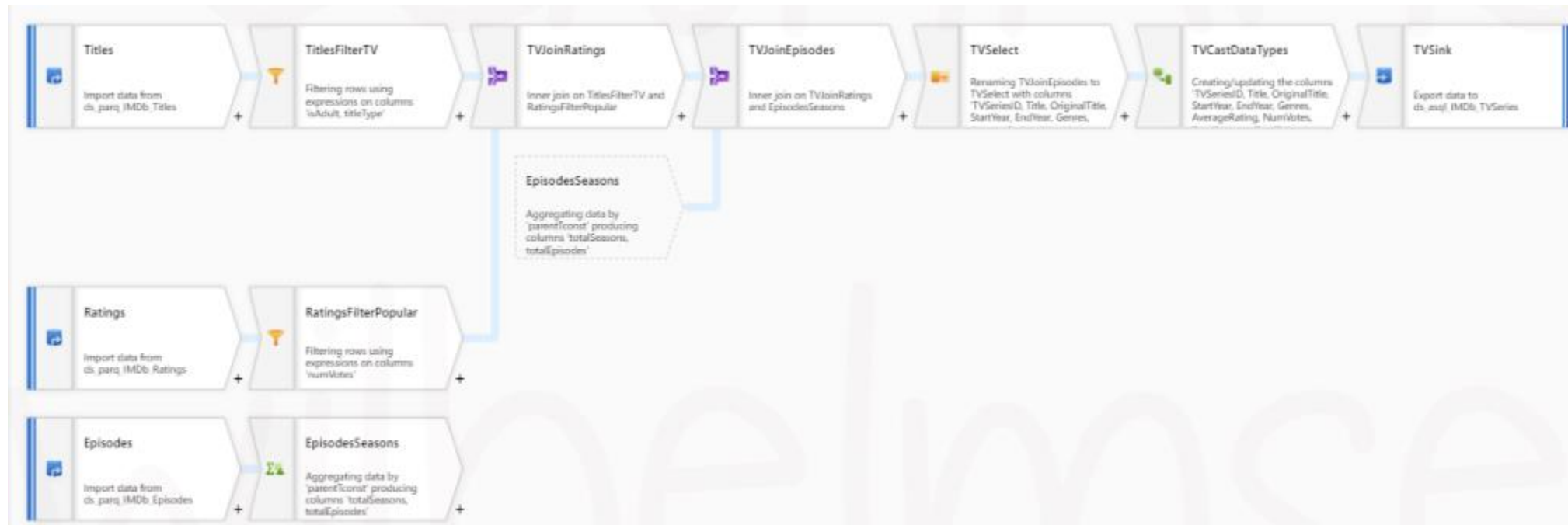
Data transformation at scale

Runs on Azure Databricks

Visual editor, no-code experience



# ADF – Dataflow activity





## *ADF – Dataflow – Why to use?*

Data Cleansing

Data Conversion

Data Preparation

Data Profiling

Data Exploration

Aggregation





## *ADF – Dataflow – When to use?*

### Big Data and Analytics

- Data De-Duplication
- Fuzzy Lookups

### Modern Data Warehouse

- Star Schemas
- Slowly Changing Dimensions



## *ADF – Dataflow – What it is?*

Transformations are steps in the data flow

Visualized from left (*source*) to right (*sink*)

Currently **18** types of transformations



*Labs*