Homework | Cycle 29 | Session 19

- Prepare the watermark table and create the pipeline to read watermark and max value from source table

- On basic of emp id we are going to do INCREMENTAL LOAD as we do not have any Timestamp coloumn created in Employees table and we have only empid coloumn.

- Create metadata table or else we will call this table as Watermark table also

```
CREATE TABLE Employees
(
empid int,
empname VARCHAR(100),
empcity VARCHAR(50),
empphonenumber bigint
)

INSERT INTO Employees VALUES(1,'Bhavani','GN',6759804567)
INSERT INTO Employees VALUES(1,'Neelu','HYD',6779804567)
INSERT INTO Employees VALUES(1,'Jag','VJD',6759504567)
```

```sql
CREATE TABLE Employees
(
empid int,
empname VARCHAR(100),
empcity VARCHAR(50),
empphonenumber bigint
)
```

Messages

Commands completed successfully.

Completion time: 2025-06-10T18:24:53.7410312-04:00

```
33
34      INSERT INTO Employees VALUES(1,'Bhavani','GN',6759804567)
35      INSERT INTO Employees VALUES(1,'Neelu','HYD',6779804567)
36      INSERT INTO Employees VALUES(1,'Jag','VJD',6759504567)
37
38
39
```

Messages

```
(1 row affected)

(1 row affected)

(1 row affected)

Completion time: 2025-06-10T18:27:44.8766667-04:00
```

```
37
38
39      SELECT * FROM Employees
40
41
42
43
44
45
```

Results   Messages

	empid	empname	empcity	empphonenumber
1	1	Bhavani	GN	6759804567
2	1	Neelu	HYD	6779804567
3	1	Jag	VJD	6759504567

- Now we have 3 records of employee table normal table
- Now create WATERMARK Table as metadata table

Ex: identity(1,1) incremental, start with 1 and load +1, +1

CREATE TABLE WATERMARK(

id int identity(100,1),

tablename varchar(100),

schemaname varchar(50),

filename varchar(100),

foldername varchar(100),

lastprocessedvalue varchar(100),

incrementalcoloumn varchar(100)

)

Note: We are using lastprocessedvalue varchar(100),

Which means we are considering empid as lastprocessed value because based on empid we are doing incremental load as we dont have timestamp for now.

Empid datatype is int in future to handle all the datatypes now we are taking VARCHAR for lastprocessed value. But we have to handle all scenario and timestamp scenario also right thatsy we took varchar.

So we are doing incremental load on incrementalcoloumn basis

INSERT INTO WATERMARK
VALUES('Employees','dbo','employee','incremental/dbo/employee',0,'empid')

- When we are copying data first time we have to take 0 as lastprocessed value
- Here we took zero as lastprocessed value because we didnt load any data yet so from 1st record we have to load thatsy we have to take 0
- If you are taking 100 as we have inserted only 3 rows into employees table the data is not going to copy if you take 100 or some other value other than 0
- On basis of empid we are going to load.

```
25
26    SELECT * FROM WATERMARK
```

% ▼   ✓ No issues found ◄

Results | 📄 Messages

id	tablename	schemaname	filename	foldername	lastprocessedvalue	incrementalcoloumn

```
27
28    INSERT INTO WATERMARK VALUES('Employees','dbo','employee','incremental/dbo/employee',0,'empid')
```

146 % ▼   ✓ No issues found ◄

⊞ Results | 📄 Messages

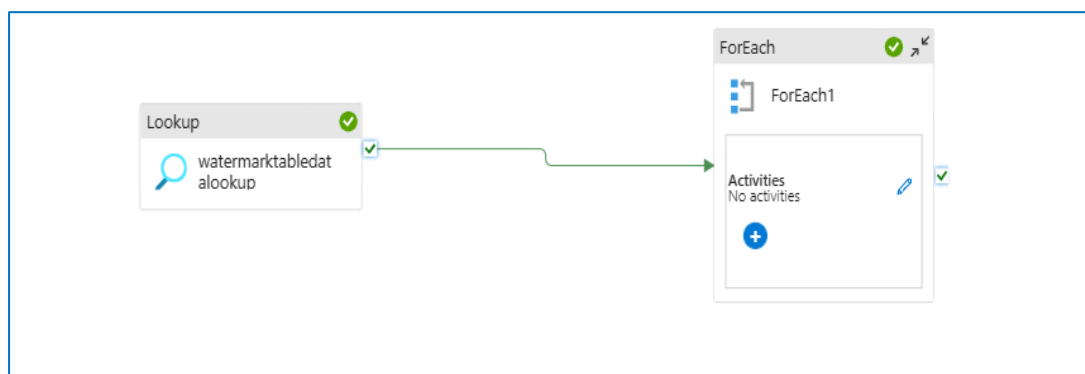|	id	tablename	schemaname	filename	foldername	lastprocessedvalue	incrementalcoloumn
1	100	Employees	dbo	employee	incremental/dbo/employee	0	empid

- Let's design Pipeline to load data for single table.

To load watermark table create a pipeline and choose lookup activity and give watermark table and run pipeline

General **Settings** User properties

Source dataset *   🗃 AzureSqlTable1 ▾   ✏ Open   + New   👓 Preview data   Learn more ⧉

               ⌄ Dataset properties ⓘ

Name	Value
Tablename	watermark
Schemaname	dbo

First row only   ☐

Use query   ⦿ Table   ◯ Query   ◯ Stored procedure

Query timeout (minutes) ⓘ   120

Isolation level ⓘ   Select... ▾

⊞⊞ pl_incremental_load ●   ⊞ AzureSqlTable1   ✕   ⓘ Other users in your workspace may have access to modify this item. Do not use this item unless you tru

    SQL   Azure SQL Database
             AzureSqlTable1

Connection   Schema   **Parameters**

+ New | 🗑 Delete

	Name	Type	Default value	
☐	Tablename	String ▾	Value	🗑
☐	Schemaname	String ▾	Value	🗑

SQL

Azure SQL Database
AzureSqlTable1

Connection   Schema   Parameters

Linked service *   [🗄 AzureSqlDatabase1   ▾]   🔌 Test connection   ✏ Edit   ＋ New   Learn more ⧉

Integration runtime *   [✅ AutoResolveIntegrationRuntime   ▾]   ✏ Edit

Table   [@dataset().Schemaname   ] . [@dataset().Tablename   ]   👓 Preview data
                                   ☑ Enter manually

Lookup ✅
🔍 watermarktabledat
      alookup

Parameters   Variables   Settings   **Output**   ⌃

Pipeline run ID 62fb9ac1-e65b-4f57-8a8a-b8546a13afd4 [📋] ↻ ⓘ   **Pipeline status** ✅ Succeeded   View debug run consumption

[All status ▾] [List ▾]   Monitor in Azure Metrics ⧉   ⤓ Export to CSV | ⌄

Showing 1 - 1 of 1 items

Activity name ↑↓	Activity st... ↑↓	Activit... ↑↓	Run start ↑↓	Duration ↑↓	Integration runtime ↑↓	User prop... ↑↓	Activity run ID ↑↓
watermarktabledatalookup	✅ Succeeded	Lookup	6/10/2025, 8:12:37 PM	15s	AutoResolveIntegrationRuntime (West US 2)		4dc334ad-5e85-4e28-9296-54cc961b

Input

```
{
  "source": {
    "type": "AzureSqlSource",
    "queryTimeout": "02:00:00",
    "partitionOption": "None"
  },
  "dataset": {
    "referenceName": "AzureSqlTable1",
    "type": "DatasetReference",
    "parameters": {
      "Tablename": "watermark",
      "Schemaname": "dbo"
    }
  },
  "firstRowOnly": false
}
```

Output

```
{
    "count": 1,
    "value": [
        {
            "id": 100,
            "tablename": "Employees",
            "schemaname": "dbo",
            "filename": "employee",
            "foldername": "incremental/dbo/employee",
            "lastprocessedvalue": "0",
            "incrementalcoloumn": "empid"
        }
    ],
    "effectiveIntegrationRuntime": "AutoResolveIntegrationRuntime (West US 2)",
    "billingReference": {
        "activityType": "PipelineActivity",
        "billableDuration": [
            {
                "meterType": "AzureIR",
                "duration": 0.016666666666666666,
                "unit": "Hours"
            }
        ],
        "totalBillableDuration": [
            {
                "meterType": "AzureIR",
                "duration": 0.016666666666666666,
                "unit": "Hours"
            }
        ]
    },
    "durationInQueue": {
        "integrationRuntimeQueue": 0
    }
}
```

Now to load multiple tables of data choose for each activity

Pipeline expression builder

Add dynamic content below using any combination of expressions, functions and system variables.

```
@activity('watermarktabledatalookup').output.value
```

Clear contents

Activity outputs   Parameters   System variables   Functions   Variables

🔍 Search

watermarktabledatalookup
watermarktabledatalookup activity output

watermarktabledatalookup count
Count of the rows

watermarktabledatalookup value array
Array of row data

OK   Cancel

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

Sequential ☐

Batch count ⓘ [   ]

Items [@activity('watermarktabledatalooku...]

- Now before going to inside foreach, before loading copydata we have to check what is the maximum value to check we have in in our source table.
- Before load employee table data into adls gen2 we have to store max records value to somewhere for that we have to create stored procedure
- Now i will store in the watermark table in the lastprocessed session instead of 0 now it will update with 3
- Ex: select * from employees where empid>0


```
29
30 | SELECT * FROM Employees where empid>0
```

% ▼   ✓ No issues found

Results   Messages

empid	empname	empcity	empphonenumber
1	Bhavani	GN	6759804567
1	Neelu	HYD	6779804567
1	Jag	VJD	6759504567

- Ex:select * from employees where empid>3

```
29
30     SELECT * FROM Employees where empid>3
```

46 %   ▾   ✓ No issues found   ◄

⊞ Results   🗐 Messages

empid	empname	empcity	empphonenumber

- Now we have execute this "s**elect max(empid) as maxvalue from Employees**" query using lookup activity inside foreach
- Once the data is loaded then i have to update this value to maximum value
- Next day we will compare if the value is greater than>3 or not, if no value we will not load data, if there is a value if there are more than 3 records, if the value is increased than 3 then only i will load the data into gen2
- Jnkfd
- Create a look up activity inside foreach and take lookup activity to get the the maximum value from source and give empty " at the Table name and schema name because we want to write with a query to get max value instead of using names we want to get max value using query.

```
29
30     SELECT * FROM Employees where empid>0
```

%   ▾   ✓ No issues found   ◄

Results   🗐 Messages

empid	empname	empcity	empphonenumber
1	Bhavani	GN	6759804567
1	Neelu	HYD	6779804567
1	Jag	VJD	6759504567

```
35
36     select max(empid) as maxvalue from Employees
```

146 %   ⦿ No issues found   ◁

⊞ Results   📑 Messages

	maxvalue
1	3

AzureSqlTable1   ℹ Other users in your workspace may have access to modify this item. Do not use this item unless you trust all users who may have access to th

✓ Validate   ▷ Debug   ⚡ Add trigger

〰 pl_incremental_load > ⊡ ForEach1

Lookup

🔍 Lookup1

🗑 </> ⧉   ➡

General   **Settings**   User properties

Source dataset *   📓 AzureSqlTable1   ∨   ✎ Open + New 👓 Preview data   Learn more ⬈

                          ∨ Dataset properties ⓘ

                          Name   Value

                          Tablename   [|]
                                          Add dynamic content [Alt+Shift+D]

                          Schemaname   ["]

First row only   ☑

Use query   ⦿ Table   ○ Query   ○ Stored procedure

Query timeout (minutes) ⓘ [120]

Isolation level ⓘ   [Select...   ∨]

To make empid column dynamic, so from the output which we taken earlier from "watermark table data lookup" output incremental column holds empid value .

```
{
    "count": 1,
    "value": [
        {
            "id": 100,
            "tablename": "Employees",
            "schemaname": "dbo",
            "filename": "employee",
            "foldername": "incremental/dbo/employee",
            "lastprocessedvalue": "0",
            "incrementalcoloumn": "empid"
        }
    ],
    "effectiveIntegrationRuntime": "AutoResolveIntegrationRuntime (West US 2)",
    "billingReference": {
        "activityType": "PipelineActivity",
        "billableDuration": [
            {
                "meterType": "AzureIR",
                "duration": 0.016666666666666666,
                "unit": "Hours"
            }
        ],
        "totalBillableDuration": [
            {
                "meterType": "AzureIR",
                "duration": 0.016666666666666666,
                "unit": "Hours"
            }
        ]
    },
    "durationInQueue": {
        "integrationRuntimeQueue": 0
    }
}
```

Query: select max(@{item().incrementalcoloumn}) as maxvalue from @{item().tablename}
max(@{item().incrementalcoloumn}) will give you dynamic value
@{item().tablename} will fetch table name from db as dynamic.

Pipeline expression builder

Add dynamic content below using any combination of expressions, functions and system variables.

```
select max(@{item().incrementalcoloumn}) as maxvalue from @{item
    ().tablename}
```

Clear contents

pl_incremental_load > ForEach1

Lookup

SourceMaxValue

On fail

General	**Settings**	User properties

Source dataset *	AzureSqlTable1	✏️ Open + New 👓 Preview data Learn more 🗗
	∨ Dataset properties ⓘ	

Name	Value
Tablename	"
Schemaname	"

First row only	☑️
Use query	◯ Table ⦿ Query ◯ Stored procedure
Query	select max(@{item().incrementalcolo...
Query timeout (minutes) ⓘ	120

Debug

Input

Copy to clipboard

```
{
  "source": {
    "type": "AzureSqlSource",
    "sqlReaderQuery": "select max(empid) as maxvalue from
Employees",
    "queryTimeout": "02:00:00",
    "partitionOption": "None"
  },
  "dataset": {
```

[@]

SourceMaxValue   ✅ Succeeded   Lookup

It displayed or fetch max value in output through look up

Output

Copy to clipboard

```
{
  "firstRow": {
    "maxvalue": 3
  },
  "effectiveIntegrationRuntime":
"AutoResolveIntegrationRuntime (West US 2)",
  "billingReference": {
    "activityType": "PipelineActivity",
    "billableDuration": [
```

✓ Validate   ▷ Debug   ⚡ Add trigger

{} 🗎 ⋯

Lookup ✓
🔍 watermarktabledat
alookup

ForEach ✓ ↙
▦ ForEach1

Activities
No activities
⊕

🔍
+
−
○
⊡
⊞
↙

Parameters   Variables   Settings   **Output**

Pipeline run ID dd095df0-ae07-49a9-a049-34750ad6add4 ▭ ↻ ⓘ

Pipeline status ✓ Succeeded

View debug run consumption

All status ∨   List ∨

Monitor in Azure Metrics ↗   ⬇ Export to CSV   ∨

Showing 1 - 3 of 3 items

Activity name ↑↓	Activity st... ↑↓	Activit... ↑↓	Run start ↑↓	Duration ↑↓	Integration runtime ↑↓	User prop... ↑↓	Activity run ID ↑↓
SourceMaxValue	✓ Succeeded	Lookup	6/11/2025, 3:23:53 PM	7s	AutoResolveIntegrationRuntime (West US 2)		df92d8b0-5d30-414f-ba06-8e5c5db0
ForEach1	✓ Succeeded	ForEach	6/11/2025, 3:23:53 PM	10s			00a62e95-b91c-4779-a0a9-07b84e4
watermarktabledatalookup	✓ Succeeded	Lookup	6/11/2025, 3:23:34 PM	17s	AutoResolveIntegrationRuntime (West US 2)		88b286f9-67b8-41fd-a8f0-50daeec9