

Azure Synapse Workspace

Hands on Lab Guide

실습 환경 설정

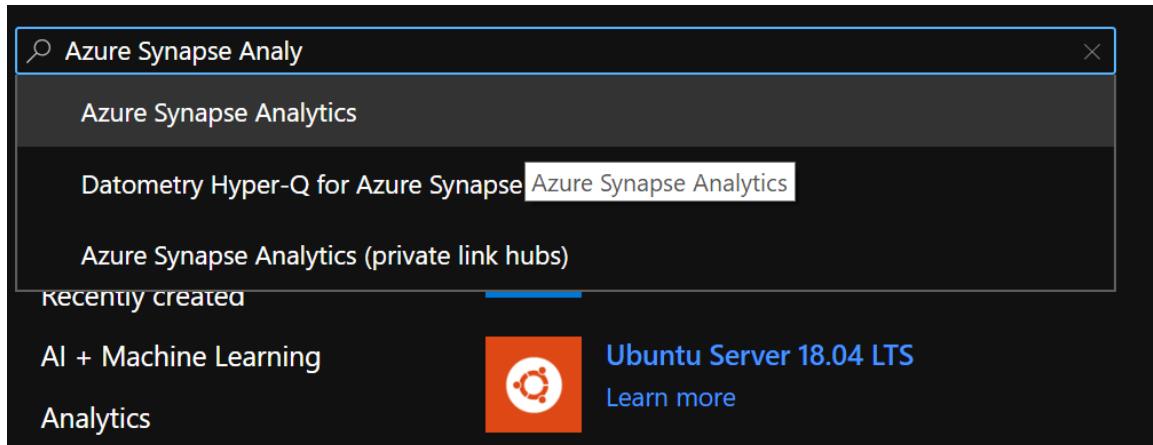
요구 사항

1. Microsoft Azure 구독 (Microsoft 이외의 구독은 유료 구독이어야합니다).
2. **중요** :이 실습의 OAuth 2.0 액세스 구성 요소를 완료하려면 Azure 구독 내에 Azure Active Directory 내에 앱 등록 및 서비스 주체를 만들 수 있는 권한이 있어야합니다.

실습 전에

작업 1 : Azure Synapse Workspace 생성

1. Azure Portal 에 접속하여 Azure Synapse Workspace 를 입력합니다.



2. Region 은 Korea Central 을 선택하고 추가 정보들을 입력합니다.

* Basics * Security Networking Tags Summary

Create a Synapse workspace to develop an enterprise analytics solution in just a few clicks.

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all of your resources.

Subscription * ⓘ Microsoft Azure Internal Consumption

Resource group * ⓘ (New) KyoWon_HoL

[Create new](#)

Workspace details

Name your workspace, select a location, and choose a primary Data Lake Storage Gen2 file system to serve as the default location for logs and job output.

Workspace name * kyowon-synapse

Region * ⓘ **Korea Central**

Select Data Lake Storage Gen2 * ⓘ From subscription Manually via URL

3. Data Lake Storage Gen2 를 새로 만들기 위해 Create New 를 선택하고 필요 정보들을 입력합니다.
4. 파일 시스템 명도 마찬가지로 정해주고 아래와 같이 체크박스에 체크합니다

Select Data Lake Storage Gen2 * ⓘ From subscription Manually via URL

Account name * ⓘ (New) kywonadls

[Create new](#)

File system name * ⓘ (New) synapse

[Create new](#)

Assign myself the Storage Blob Data Contributor role on the Data Lake Storage Gen2 account 'kywonadls'.

i We will automatically grant the workspace identity data access to the specified Data Lake Storage Gen2 account, using the [Storage Blob Data Contributor](#) role. To enable other users to use this storage account after you create your workspace, perform these tasks:

5. Security 탭에서 어드민 계정 정보를 입력합니다.

The screenshot shows the 'Security' tab selected in the top navigation bar. Below it, a sub-section titled 'SQL administrator credentials' is displayed. It includes fields for 'Admin username' (set to 'sqladminuser'), 'Password' (represented by a redacted string), and 'Confirm password' (also represented by a redacted string). A green checkmark icon is positioned next to the 'Confirm password' field. At the bottom right of the screen, there is a large green circular button with a checkmark and the letter 'P'.

Configure security options for your workspace.

SQL administrator credentials

Provide credentials that can be used for administrator access to the workspace's SQL pools. If you don't provide a password, one will be automatically generated. You can change the password later.

Admin username *

sqladminuser

Password

.....

Confirm password

.....

Workspace encryption

⚠ Double encryption configuration cannot be changed after opting into using a customer-managed key at the time of workspace creation.

Choose to encrypt all data at rest in the workspace with a key managed by you (customer-managed key). This will provide double encryption with encryption at the infrastructure layer that uses platform-managed keys. [Learn more](#)

Enable double encryption using a customer-managed key

System assigned managed identity

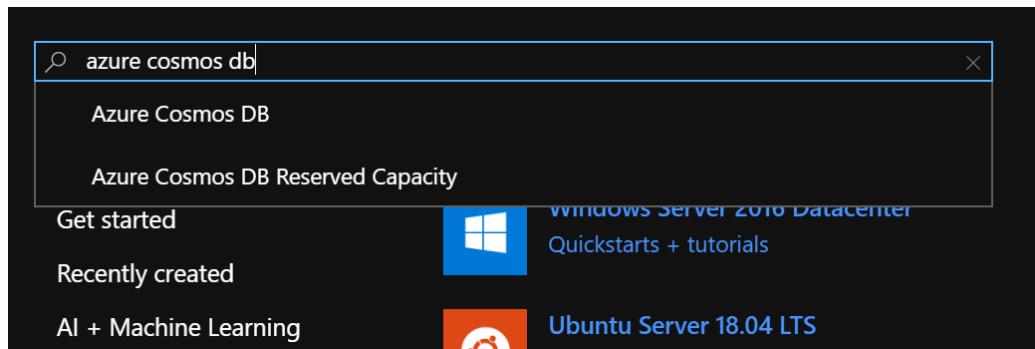
Choose whether you'd like to assign the workspace's system-assigned managed identity CONTROL permissions to SQL pools for pipeline integration. [Learn more](#)

Allow pipelines (running as workspace's system assigned identity) to access SQL pools.

6. Review + Create 하여 생성을 완료합니다.

작업 2 : Azure Cosmos DB 생성

1. Create Resource 를 선택하여 Azure Cosmos DB 를 입력합니다.



2. Azure Cosmos DB 를 선택하고 생성합니다.
3. API 는 Core(SQL)을 선택하고, Region 은 Korea South 혹은 Korea Central 을 선택합니다.

Basics Networking Backup Policy Encryption Tags Review + create

Azure Cosmos DB is a globally distributed, multi-model, fully managed database service. Try it for free, for 30 days with unlimited renewals. Go

Project Details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription * Microsoft Azure Internal Consumption

Resource Group * KyoWon_HoL Create new

Instance Details

Account Name * kyowoncosmosdb

API * ⓘ Core (SQL)

Notebooks (Preview) ⓘ On Off

Location * (Asia Pacific) Korea South

Capacity mode ⓘ Provisioned throughput Serverless (preview)
Learn more about capacity mode

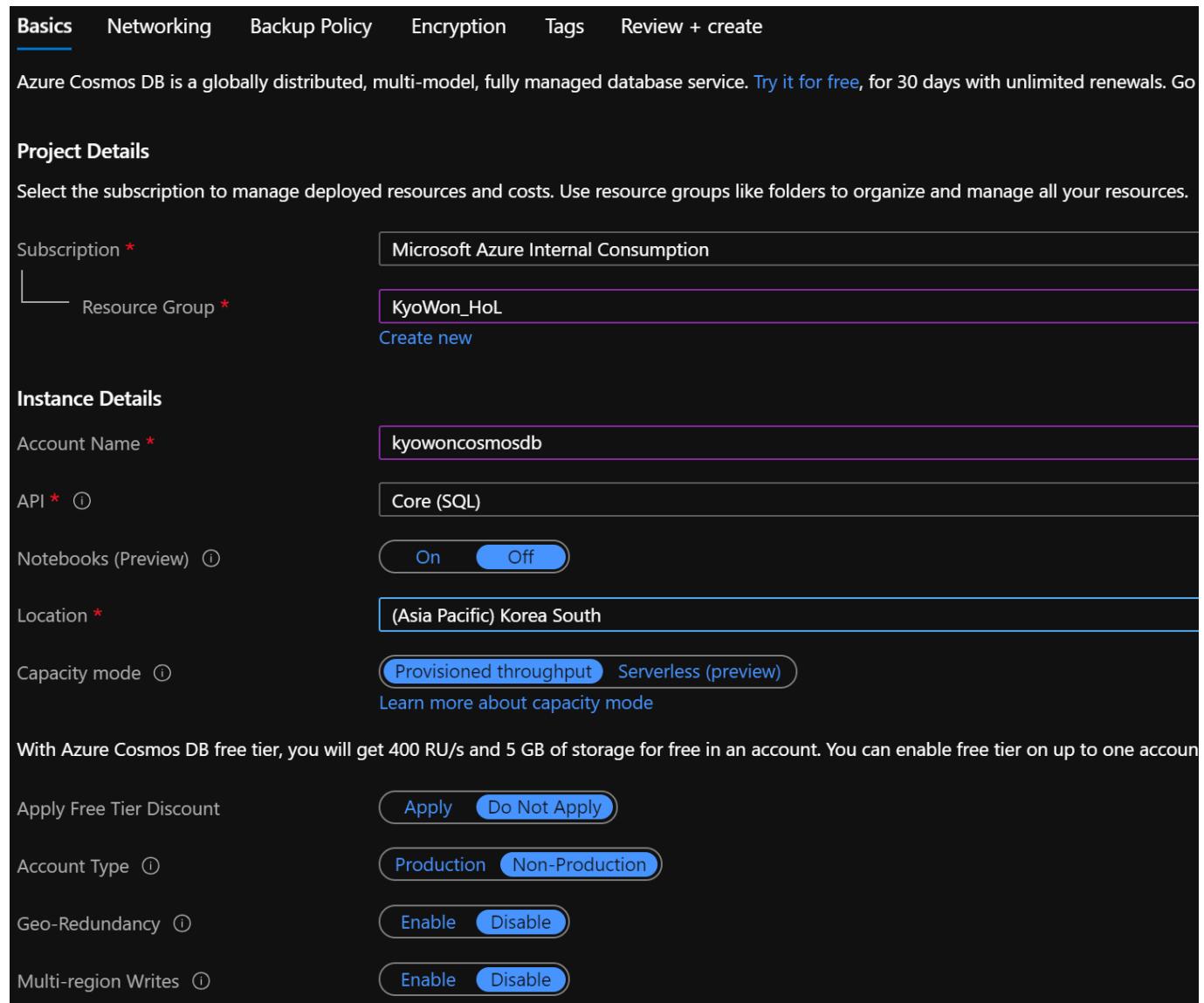
With Azure Cosmos DB free tier, you will get 400 RU/s and 5 GB of storage for free in an account. You can enable free tier on up to one account.

Apply Free Tier Discount Apply Do Not Apply

Account Type ⓘ Production Non-Production

Geo-Redundancy ⓘ Enable Disable

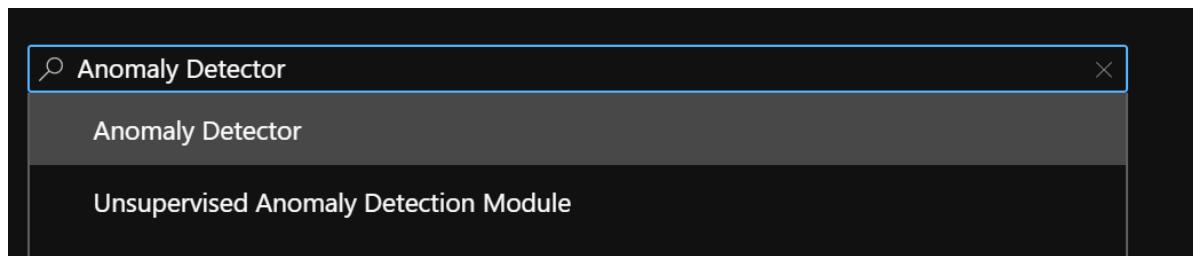
Multi-region Writes ⓘ Enable Disable



4. Review + Create 하여 생성을 완료합니다.

작업 3 : Anomaly Detector (Cognitive Service) 생성

1. Create Resource에서 Anomaly Detector 를 입력하여 설치를 진행합니다.



2. 아래와 같이 적절한 정보를 입력하고 생성을 진행합니다.

Basics Tags Review + create

Easily embed anomaly detection capabilities into your apps so users can quickly identify problems. Through an API, Anomaly Detector ingests time-series data of all types and selects the best-fitting detection model for your data to ensure high accuracy. Customize the service to detect any level of anomaly and deploy it wherever you need it most. Azure is the only major cloud provider that offers anomaly detection as an AI service. No machine learning expertise is required. [Learn more](#)

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription * ⓘ Microsoft Azure Internal Consumption

Resource group * ⓘ KyoWon_HoL Create new

Instance details

Region * ⓘ East Asia

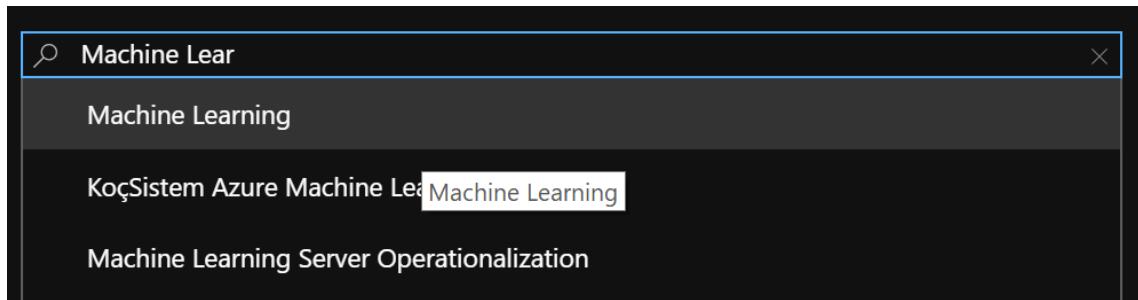
Name * ⓘ kyowonanomalydetect ✓

Pricing tier * ⓘ Free F0 (10 Calls per second, 20K Transactions per month)

[View full pricing details](#)

작업 4 : Azure Machine Learning Workspace 생성

1. Create Resource에서 Machine Learning 를 입력하여 설치를 진행합니다.



2. 아래와 같이 적절한 값을 넣고 생성을 진행합니다.

Basics Networking Advanced Tags Review + create

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription * ⓘ Microsoft Azure Internal Consumption

Resource group * ⓘ KyoWon_HoL

Workspace details

Specify the name and region for the workspace.

Workspace name * ⓘ kyowonmlworkspace

Region * ⓘ Korea Central

Storage account * ⓘ (new) kyowonmlworksp4913577332

Key vault * ⓘ (new) kyowonmlworksp4543140955

Application insights * ⓘ (new) kyowonmlworksp8868320504

Container registry * ⓘ None

실습 랩 가이드

작업 1 : Azure Cosmos DB Database 및 Collection 생성

1. [Azure Portal](#)에서 Azure Cosmos DB 계정으로 이동한 다음 왼쪽 메뉴에서 Data Explorer를 선택합니다.

The screenshot shows the Azure Cosmos DB Data Explorer interface for the account 'kyowoncosmosdb'. The left sidebar contains navigation links: Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Quick start, Notifications, and Data Explorer (which is selected). Below these are Settings sections for Default consistency, Backup & Restore, Firewall and virtual networks, Private Endpoint Connections, CORS, Keys, Add Azure Cognitive Search, and Add Azure Function. The main area displays the 'Welcome to' message and the SQL API. A 'New Container' button is visible at the top right. A 'Start with Sample' button with a planet icon is also present. At the bottom, there are 'Common Tasks' (New Database) and 'Recents' sections.

2. 위의 New Container를 클릭 후 아래와 같이 DB ID를 CosmosDemo로 입력하고 Autoscale을 선택, Container ID를 IoTDeviceInfo로 입력, Partition Key는 /id를 입력합니다.

*** Database id** ⓘ Create new Use existing CosmosDemo Provision database throughput ⓘ*** Throughput (autoscale)** ⓘ Autoscale Manual

Provision maximum RU/s required by this resource. Estimate your required RU/s with [capacity calculator](#).

Max RU/s

 4000

Your database throughput will automatically scale from **400 RU/s (10% of max RU/s) - 4000 RU/s** based on usage.

After the first 40 GB of data stored, the max RU/s will be automatically upgraded based on the new storage value. [Learn more](#).

Estimated monthly cost (USD): **\$35.04 - \$350.40** (1 region, 400 - 4000 RU/s, \$0.00012/RU)

*** Container id** ⓘ IoTDeviceInfo*** Partition key** ⓘ /id My partition key is larger than 100 bytes

3. 이후 아래 Azure Synapse Link 를 Enable 시킵니다. (시간이 조금 소요됩니다)

*** Analytical store** ⓘ On Off

Azure Synapse Link is required for creating an analytical store container. Enable Synapse Link for this Cosmos DB account. [Learn more](#)

4. Analytical Store 에 Radio Button 이 선택된 것을 확인합니다.

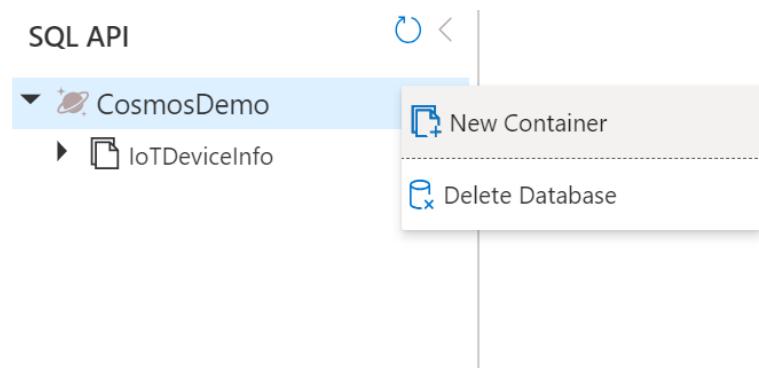
* Analytical store ⓘ

On Off

Unique keys ⓘ

+ Add unique key

5. 다시 CosmosDemo DB 를 선택하고 New Container 를 선택합니다.



6. Collection ID 에 IoTSignals 를 입력, Partition Key 에 마찬가지로 /id 를 입력합니다
Analytical Store 도 마찬가지로 On 을 선택합니다.
7. 계속해서 Collection ID 에 IoTStreamingSignals 를 입력, Partition Key 에 마찬가지로 /id 를 입력하고, 여기서는 Analytical Store 를 Off 를 선택합니다.

* Database id ⓘ
 Create new Use existing
CosmosDemo ▾

* Container id ⓘ
IoTStreamingSignals

* Partition key ⓘ
/id

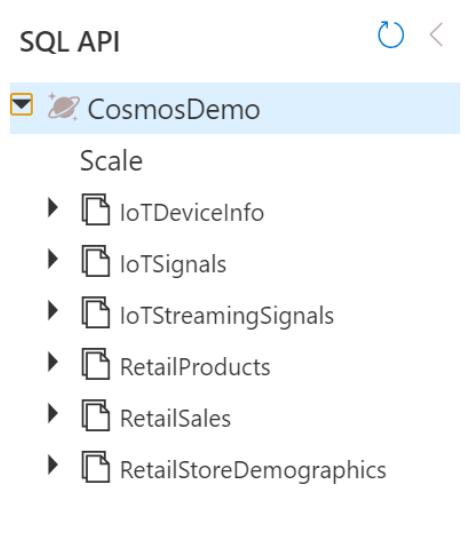
My partition key is larger than 100 bytes

Provision dedicated throughput for this container ⓘ

* Analytical store ⓘ
 On Off

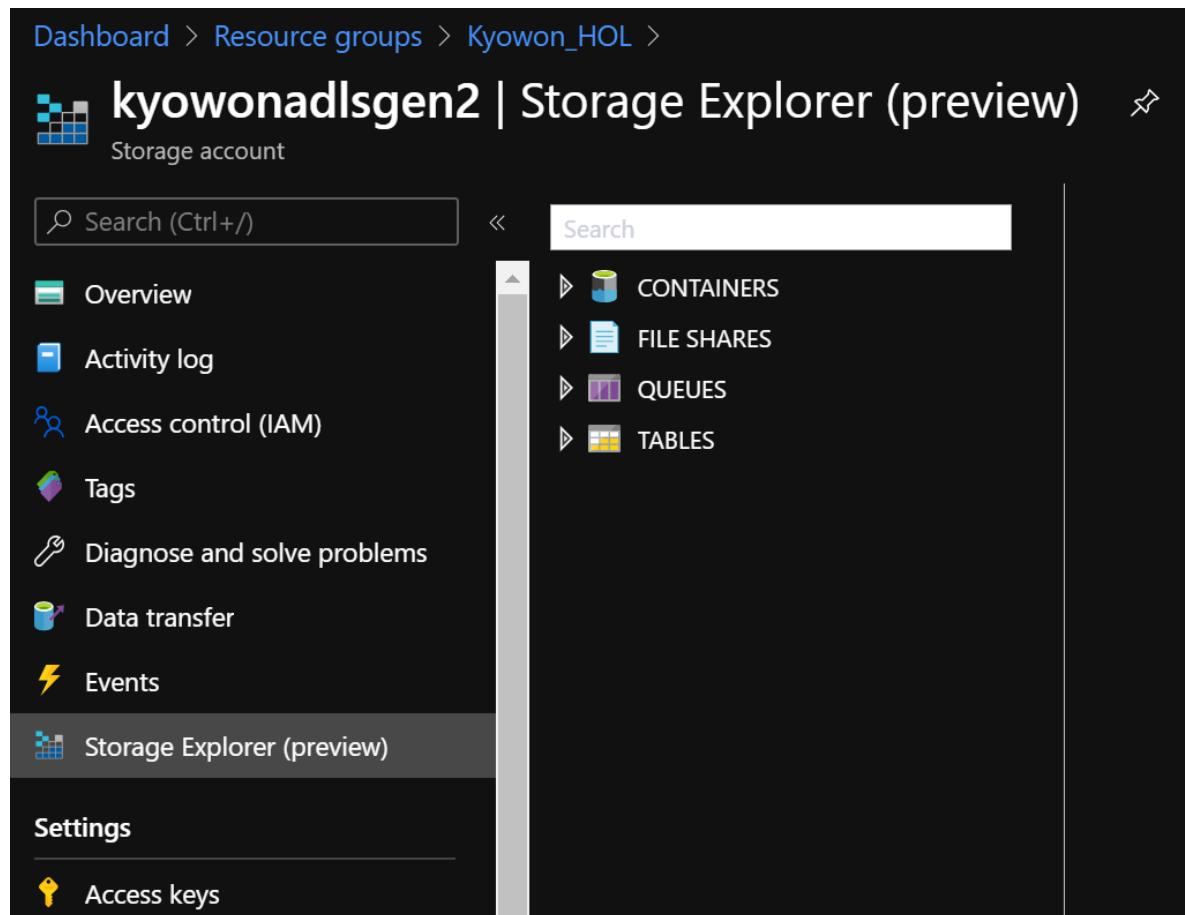
Unique keys ⓘ
+ Add unique key

8. Collection ID 에 RetailProducts 를 입력, Partition Key 에 마찬가지로 /id 를 입력합니다. Analytical Store 를 On 을 선택합니다.
9. Collection ID 에 RetailSales 를 입력, Partition Key 에 마찬가지로 /id 를 입력합니다. Analytical Store 를 On 을 선택합니다.
10. Collection ID 에 RetailStoreDemographics 를 입력, Partition Key 에 마찬가지로 /id 를 입력합니다. Analytical Store 를 On 을 선택합니다.
11. 최종 모습은 아래와 같습니다.

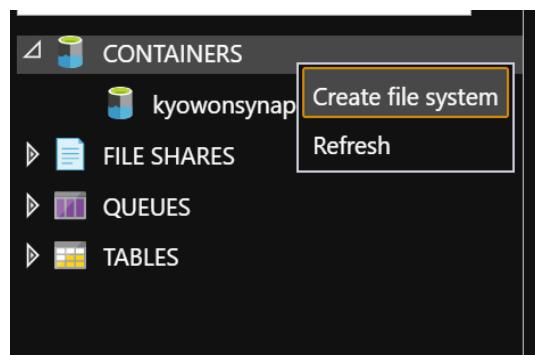


작업 3 : ADLS Gen2

1. [Azure Portal](#)에서 생성된 ADLS Gen2 Storage Account를 선택합니다.
2. 아래와 같이 Storage Explorer를 선택합니다.

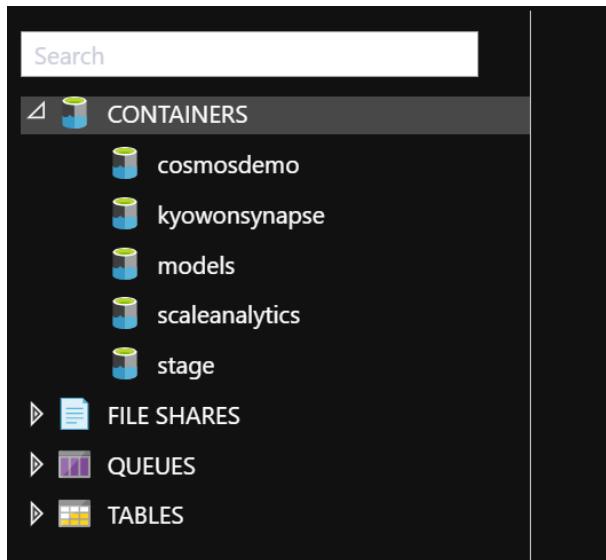


3. Container 를 선택 후 마우스 우클릭 하여 Create File System 을 선택합니다.



4. cosmosdemo 라고 이름을 입력합니다.
5. 마찬가지로 'models' 컨테이너를 생성합니다.
6. 마찬가지로 'scaleanalytics' 컨테이너를 생성합니다.
7. 마찬가지로 'stage' 컨테이너를 생성합니다.

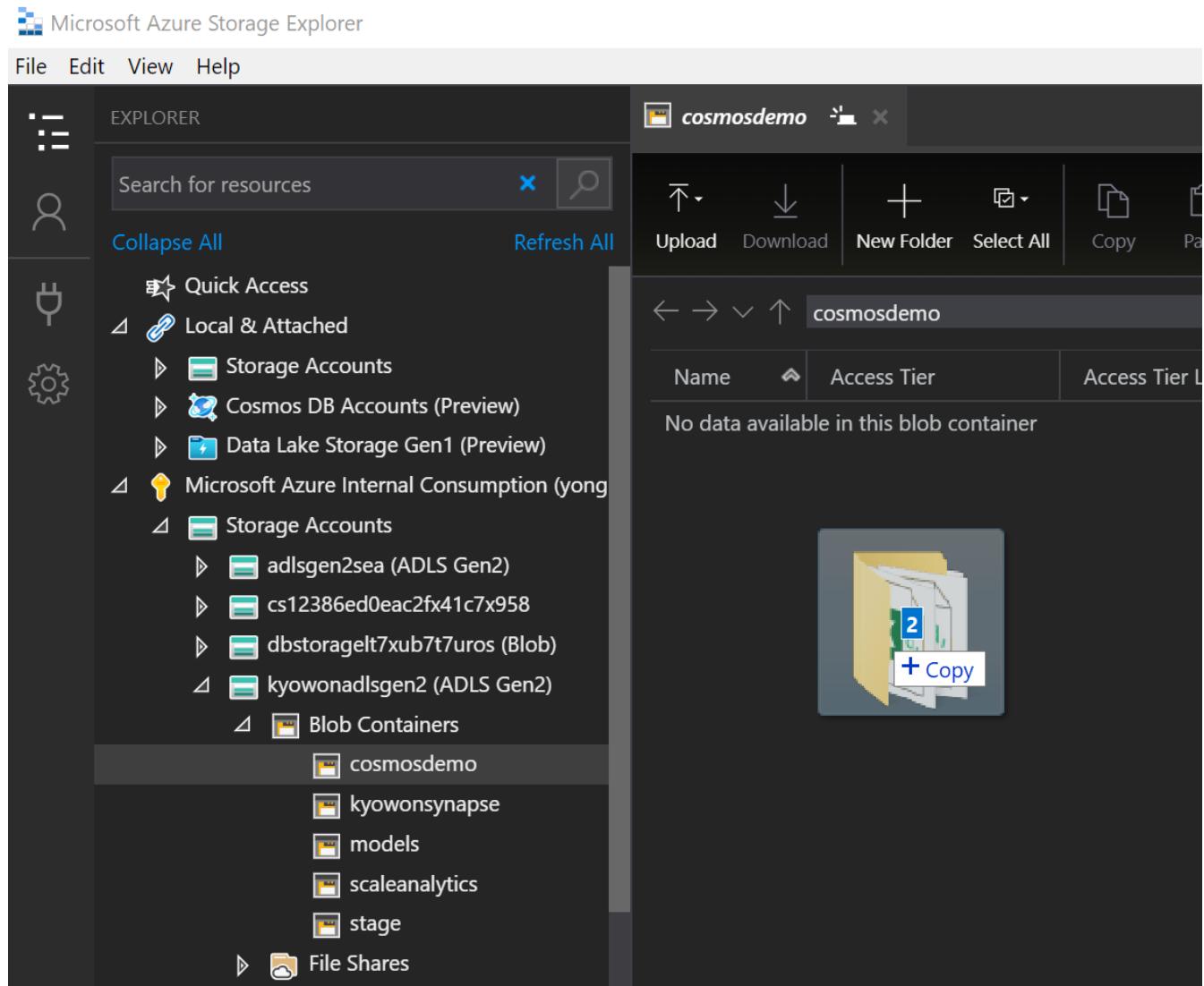
8. 마지막으로 'wwi-02' 컨테이너를 생성합니다.



9. Upload 버튼을 클릭하고 Storage Explorer 를 다운로드 받습니다.



10. Storage Explorer 를 실행하고 계정 연결 후 ADLS Gen2 를 선택합니다. 깃허브에서 다운로드 받은 Synapse_HOL\Synapse_HOL\LabFiles\ADLSGen2\cosmosdemo 안의 모든 파일을 cosmosdemo 컨테이너 안에 올려놓습니다.



11. ScaleAnalytics 와 'wwi-02' 컨테이너에도 마찬가지로

Synapse_HOL\Synapse_HOL\LabFiles\ADLSGen2\scaleanalytics 와 wwi-02 안의 모든 파일을 올려놓습니다

작업 4 : Synapse Workspace

1. 새로 생성한 Azure Synapse Workspace 를 선택합니다.
2. 왼쪽 메뉴에서 sqlpool 을 선택합니다.

The screenshot shows the Azure Synapse Studio interface. On the left, there's a sidebar with options like Overview, Activity log, Access control (IAM), Tags, Settings, SQL Active Directory admin, Properties, Locks, Synapse resources (with SQL pools highlighted in red), Apache Spark pools, Security, Firewalls, and Managed identities. The main panel displays resource group details: Resource group (Kyowon_HOL), Status (Succeeded), Location (Southeast Asia), Subscription (Microsoft Azure Internal Cons...), Subscription ID (2386ed0e-ac2f-41c7-958a-739...), Managed virtual network (No), Managed Identity object id (ecfa88fa-0589-4424-a585-84f...), Workspace web URL (<https://web.azuresynapse.net/>), and Tags (Click here to add tags). Below this is a section for Available resources, which currently shows 'No pools provisioned'.

- +New 를 선택하여 SQL Pool 을 새로 생성합니다. 이름을 넣고, DW100c 로 Performance Level 을 조절하여 줍니다.

The screenshot shows the 'Create SQL pool' wizard. The 'Basics' tab is active. The 'SQL pool name' field contains 'sqlpool'. A slider for 'Performance level' is set to 'DW100c'. Below it, the 'Estimated price' is listed as 'Est. Cost Per Hour: 1349.58 KRW' with a 'View pricing details' link. To the right, there's a list of checked items: Database, The value, Database, and No database.

- 마저 생성을 진행합니다.
- 생성이 완료되면 생성된 sqlpool 을 선택하여 sqlpool 설정으로 들어갑니다.

Analytics pools

Search to filter items...

Name	Type
Built-in	Serverless
sqlpool	Dedicated

Apache Spark pools

No pools provisioned

6. 아래와 같이 Pause 를 클릭하여 잠시 과금되는 것을 방지합니다.

sqlpool (kyowon-synapse/sqlpool) ⚡ ⏪

Dedicated SQL pool

Search (Ctrl+ /) <

Pause Scale Restore + New restore point Delete

Overview

Activity log

Access control (IAM)

Tags

Settings

Workload management

Resource group (change) : KyoWon_HoL

Status : Online

Location : Korea Central

Subscription (change) : Microsoft Azure Internal Consumption

Subscription ID : 2386ed0e-ac2f-41c7-958a-7397006dce05

Tags (change) : Click here to add tags

7. Synapse 리소스로 돌아와서 마찬가지로 왼쪽 메뉴에서 Apache Spark Pool 을 선택하고, +New 를 클릭합니다.

8. 아래와 같이 이름을 기입하고 Performance Level 을 조절합니다.

* Basics * Additional settings Tags Summary

Create a Synapse Analytics Apache Spark pool with your preferred configurations. Complete the Basics tab then go to Review + create to provision with smart defaults, or visit each tab to customize.

Apache Spark pool details

Name your Apache Spark pool and choose its initial settings.

Apache Spark pool name * sparkpool

Node size family MemoryOptimized

Node size * Medium (8 vCPU / 64 GB)

Autoscale * Enabled

Number of nodes * 3 to 5

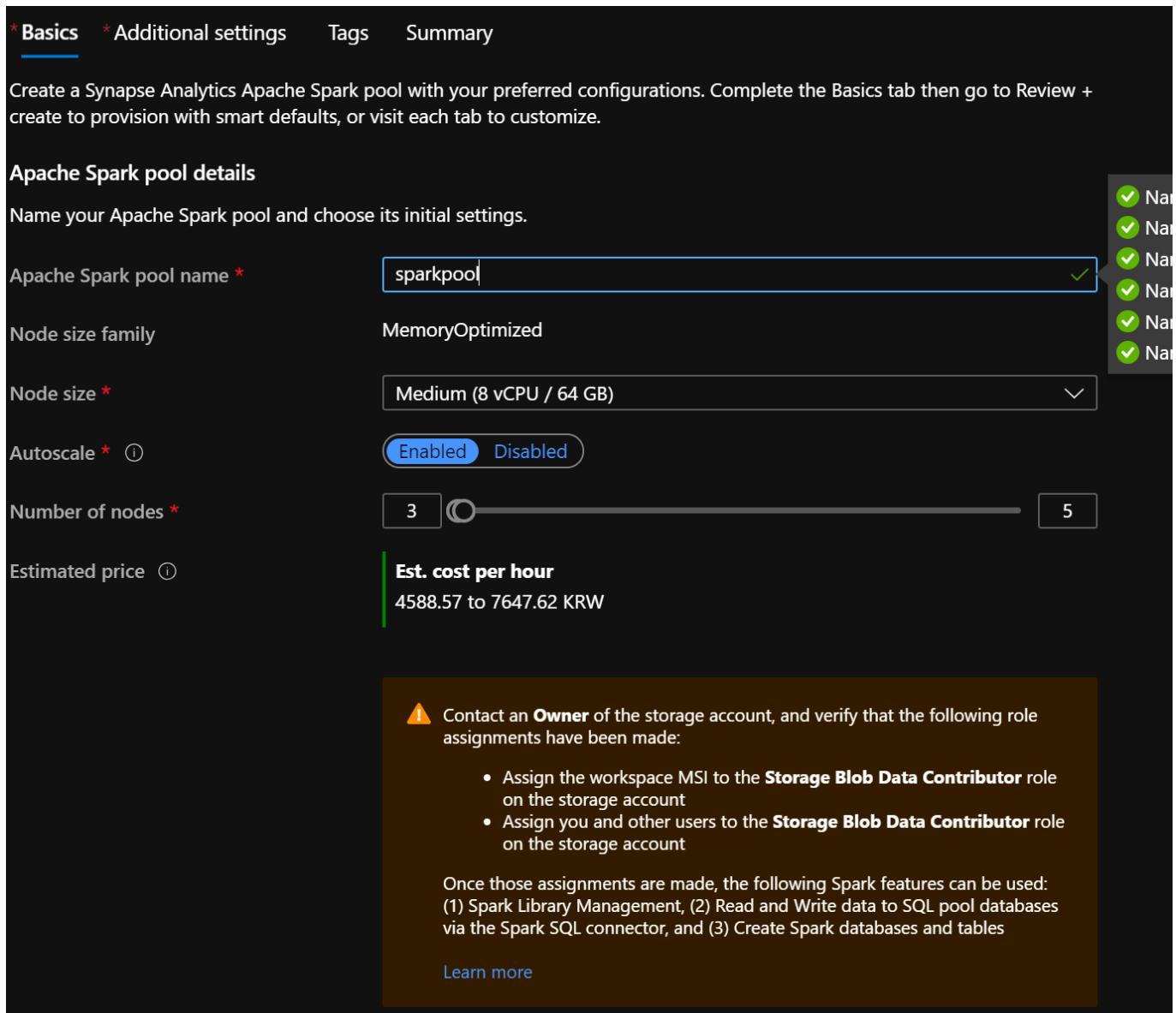
Estimated price ⓘ Est. cost per hour
4588.57 to 7647.62 KRW

⚠️ Contact an Owner of the storage account, and verify that the following role assignments have been made:

- Assign the workspace MSI to the **Storage Blob Data Contributor** role on the storage account
- Assign you and other users to the **Storage Blob Data Contributor** role on the storage account

Once those assignments are made, the following Spark features can be used:
(1) Spark Library Management, (2) Read and Write data to SQL pool databases via the Spark SQL connector, and (3) Create Spark databases and tables

[Learn more](#)



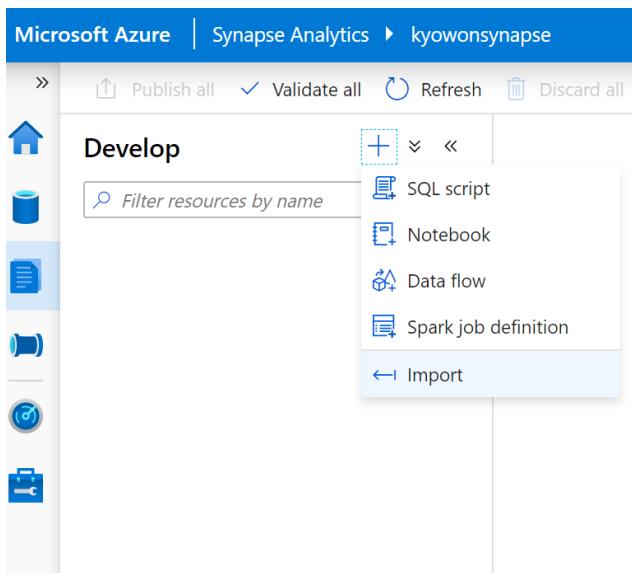
9. Review + Create 하여 마저 생성을 진행합니다.

작업 5 : Synapse Studio SQL Analytics

이번 작업에서는 MySQL Server VM(On-Premise로 가정)에 있는 DW Table 들의 Data 를 (**OLAP Data**) Migration 합니다. Azure Synapse Workspace Overview Page 에서 아래와 같이 Link 를 클릭하여 Synapse Studio 를 실행합니다.

The screenshot shows the Azure Synapse Workspace Overview page for the 'kyowon-synapse' workspace. The left sidebar contains navigation links for Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Settings (SQL Active Directory admin, Properties, Locks), Analytics pools (SQL pools, Apache Spark pools), and Security (Encryption, Firewalls, Managed identities, Private endpoint connections, Approved Azure AD tenants). The main content area displays workspace details: Resource group (KyoWon_HoL), Status (Succeeded), Location (Korea Central), Subscription (Microsoft Azure Internal Consumption), Subscription ID (2386ed0e-ac2f-41c7-958a-7397006dce05), Managed virtual network (No), Managed Identity object (d5e3ee85-c819-4206-a3ec-650c6ce4a3f6), Workspace web URL (https://web.azuresynapse.net?workspace=%2bsubscriptio...), and Tags (Click here to add tags). A red box highlights the 'Getting started' section, which includes a 'Open Synapse Studio' button (with a red border) and a 'Read documentation' link. Below this, there are sections for Analytics pools (Search bar, Name, Type columns) and SQL pools (Search bar, Name, Type columns).

1. 왼쪽 메뉴의 Develop 탭을 선택하고 +를 선택하여 Import 를 실행합니다.



2. 다운로드 받은 Synapse_HOL\Synapse_HOL\LabFiles\Scripts\SQL 안의 SQL Script들을 선택하여 불러옵니다.
3. Publish All 을 클릭하여 Publish 합니다.

Publish all

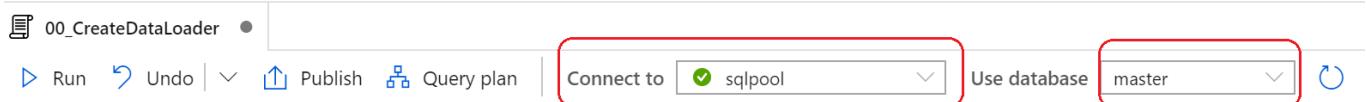
You are about to publish all pending changes to the live environment. [Learn more](#)

Pending changes (7)

NAME	CHANGE	EXISTING
▲ SQL script		
00_CreateDataLoader	(New)	-
01_CreateDWTables	(New)	-
02_CreateCredential	(New)	-
03_CreateExternalTables	(New)	-
04_CreateTransactionTables	(New)	-
05_LoadData	(New)	-
06 OLTP_Table_Count	(New)	-

4. (Optional for SSMS or Azure Data Studio) Data Loader 계정 생성을 위해

00_CreateDataLoader 스크립트를 실행합니다. 먼저 sqlpool 을 연결하고, 실행할 때 위의 두 줄은 master DB 에서 실행하고, 아래는 sqlpool 을 선택하여 실행합니다.



```

1 --Execute this line from the master database
2 CREATE LOGIN dataloader WITH PASSWORD = 'Demo@pass123';
3 CREATE USER dataloader FOR LOGIN dataloader;
4
5 --Execute the remainder of these lines from the sqlpool database
6 CREATE USER dataloader FOR LOGIN dataloader;
7 GRANT CONTROL ON DATABASE::sqlpool TO dataloader;
8 EXEC sp_addrolemember 'largerc', 'dataloader';
9

```

5. 다음으로 MySQL Server VM 의 DW 테이블들의 데이터 이관을 위해 MySQL DW Table 스키마에 맞춰 SQL Pool 안에 테이블을 생성합니다. 01_CreateDWTables 스크립트를 마찬가지로 sqlpool 에 연결하고, sqlpool database 를 선택하고 실행합니다.

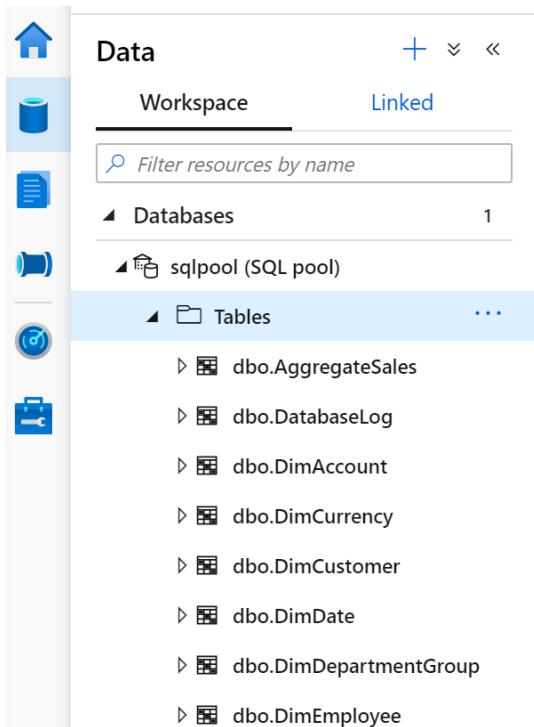


```

1 CREATE TABLE [dbo].[AggregateSales]([SalesAmount] money NOT NULL,[PostalCode] nvarchar(30) NULL,[CustomerIncome]
2 CREATE TABLE [dbo].[DatabaseLog]([DatabaseLogID] int NOT NULL,[PostTime] datetime NOT NULL,[DatabaseUser] nvar
3 CREATE TABLE [dbo].[DimAccount]([AccountKey] int NOT NULL,[ParentAccountKey] int NULL,[AccountCodeAlternateKey]
4 CREATE TABLE [dbo].[DimCurrency]([CurrencyKey] int NOT NULL,[CurrencyAlternateKey] nchar(6) NOT NULL,[Curren
5 CREATE TABLE [dbo].[DimCustomer]([CustomerKey] int NOT NULL,[GeographyKey] int NULL,[CustomerAlternateKey] nva
6 CREATE TABLE [dbo].[DimDate]([DateKey] int NOT NULL,[FullDateAlternateKey] date NOT NULL,[DayNumberOfWeek] tir
7 CREATE TABLE [dbo].[DimDepartmentGroup]([DepartmentGroupKey] int NOT NULL,[ParentDepartmentGroupKey] int NULL,
8 CREATE TABLE [dbo].[DimEmployee]([EmployeeKey] int NOT NULL,[ParentEmployeeKey] int NULL,[EmployeeNationalIDA]
9 CREATE TABLE [dbo].[DimGeography]([GeographyKey] int NOT NULL,[City] nvarchar(60) NULL,[StateProvinceCode] nva
10 CREATE TABLE [dbo].[DimOrganization]([OrganizationKey] int NOT NULL,[ParentOrganizationKey] int NULL,[Percenta
11 CREATE TABLE [dbo].[DimProduct]([ProductKey] int NOT NULL,[ProductAlternateKey] nvarchar(20) NULL,[ProductSubc
12 CREATE TABLE [dbo].[DimProductCategory]([ProductCategoryKey] int NOT NULL,[ProductCategoryAlternateKey] int nu
13 CREATE TABLE [dbo].[DimProductSubcategory]([ProductSubcategoryKey] int NOT NULL,[ProductSubcategoryAlternateKe
14 CREATE TABLE [dbo].[DimPromotion]([PromotionKey] int NOT NULL,[PromotionAlternateKey] int NULL,[EnglishPromoti

```

6. 왼쪽 메뉴에서 Data 탭을 선택하고, Workspace 탭의 sqlpool 에서 테이블이 생성되었는지 확인합니다.



7. 다시 Develop 탭으로 돌아와서, 02_CreateCredential 스크립트를 선택합니다. MySQL DW Table에서 Export 받은 파일들을 ADLS Gen2의 'scaleanalytics' 컨테이너 안에 WoodgroveMySQL 폴더 안에 올려놓았습니다. 여기에 접근하기 위한 Master Key, Credential, Data Source 와 파일 포맷을 정의하여 줍니다.

```

1 CREATE MASTER KEY ENCRYPTION BY PASSWORD = 'Demo@pass123';
2 CREATE DATABASE SCOPED CREDENTIAL MigrationCredential
3 WITH IDENTITY = '<ADLS Gen2 Account Name>' , SECRET = '<ADLS Gen2 Key>'
4
5 CREATE EXTERNAL DATA SOURCE MigrationStor WITH (TYPE = HADOOP,
6 LOCATION=
7 'wasbs://scaleanalytics@<ADLS Gen2 Account Name>.blob.core.windows.net/WoodgroveMySQL',
8 CREDENTIAL = MigrationCredential);
9
10 CREATE EXTERNAL FILE FORMAT MigrationFiles WITH(FORMAT_TYPE = DelimitedText,
11 FORMAT_OPTIONS (FIELD_TERMINATOR = '|'));
12

```

8. ADLS Gen2 Account Name과 Key를 확인하기 위해 Azure Portal의 Storage Account를 선택 후 Access Key 탭에서 확인하여 입력 후 실행합니다.

9. 다음으로, 03_CreateExternalTables 스크립트를 실행하여 Export 받은 파일들을 직접 Link 하는 External Table 들을 생성합니다.
10. 왼쪽 메뉴에서 Data 탭을 선택하고, Workspace 탭의 sqlpool에서 External Table들이 생성되었는지 확인합니다. 일부 테이블에 쿼리를 실행하여 조회가 되는지 확인합니다.
11. 04_LoadData 스크립트를 실행하여, Migration 을 진행합니다.
12. 왼쪽 메뉴에서 Data 탭을 선택하고, Workspace 탭의 sqlpool에서 Table들의 Data가 잘 Migration 되었는지 일부 테이블에 쿼리를 실행하여 확인합니다.

작업 6 : Azure Synapse Ingest Wizard

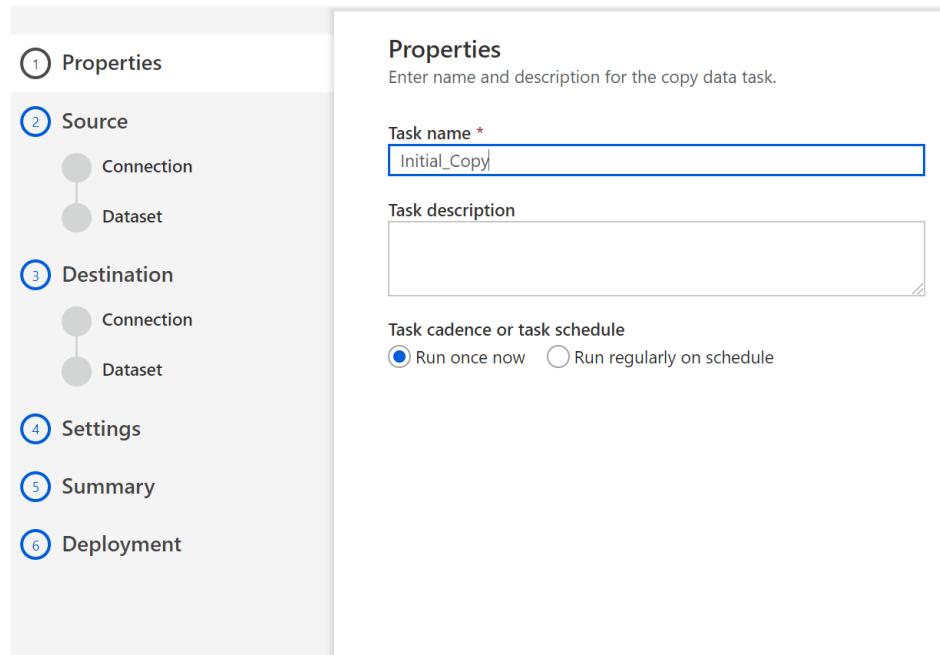
이번 작업에서는 MySQL Server VM에 있는 Transaction Table(**OLTP Data**)을 Pipeline을 통해 Migration 합니다.

- 먼저 05_CreateTransactionTable 스크립트를 실행하여 테이블을 생성합니다.
- Azure Synapse Studio의 Home으로 돌아와 Ingest를 클릭합니다.

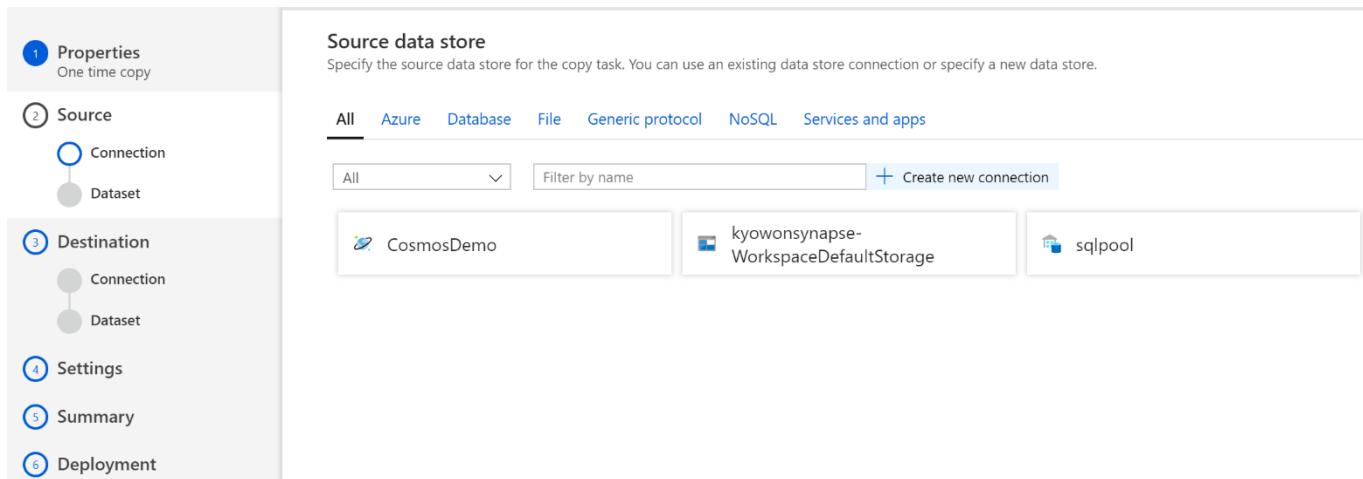
The screenshot shows the Azure Synapse Studio interface. At the top, it says "kyowonsynapse". Below that is a "Now" button. On the left, there are four main navigation icons: "Ingest" (highlighted with a red box), "Explore", "Analyze", and "Visualize". To the right of these icons is a circular graphic with a bar chart and a line graph. Below the navigation icons is a "Resources" section with "Recent" and "Pinned" tabs. Under "Recent", there is a table with columns "NAME" and "LAST OPENED BY YOU". The table lists five recent scripts: 01_CreateDWTables, 02_CreateCredential, 04_LoadData, 03_CreateExternalTables, and 05_CreateTransactionTable. All were opened 6 hours ago except for 05_CreateTransactionTable which was opened 7 hours ago. To the right of the resources is a "Useful links" section with links to "Getting started", "Synapse Analytics overview", "Pricing", "Documentation", and "Give feedback".

NAME	LAST OPENED BY YOU
01_CreateDWTables	6 hours ago
02_CreateCredential	6 hours ago
04_LoadData	6 hours ago
03_CreateExternalTables	6 hours ago
05_CreateTransactionTable	7 hours ago

3. Initial_Copy 라고 Task Name 을 입력합니다. Next.



4. + Create New Connection 을 클릭합니다.



5. MySQL 을 입력하고 아래와 같이 선택합니다. Next

New linked service

The screenshot shows a search interface with a search bar containing 'mysql'. Below the search bar is a navigation bar with tabs: All (selected), Azure, Database, File, Generic protocol, NoSQL, and Services and apps. Two items are listed in the results:

- Azure Database for MySQL**: Represented by a blue cylinder icon with 'My' on it. This item is not selected.
- MySQL**: Represented by the MySQL logo (a stylized fish) and the word 'MySQL'. This item is selected, indicated by a blue border around its box.

6. 아래와 같이 적절하게 입력합니다. (Check with proctor)

- Name: MySQL_Server
- Server Name: 207.46.235.2
- Port: 3306
- Database Name: cohooltp
- User Name: demouser
- Password: Demo@pass123

Name *

MySQL_Server

Description

Connect via integration runtime * ⓘ

AutoResolveIntegrationRuntime

 Connection string Azure Key Vault

Server name *

207.46.235.2

Port

3306

Database name *

cohooltp

User name *

demouser

 Password Azure Key Vault

Password *

.....

SSL mode

Preferred

 Use system trust store Not use system trust store

7. Test Connection 을 클릭하여 연결이 잘 되는지 확인하고 생성합니다.

Annotations

[+ New](#)

Connection successful

[Create](#)[Back](#)

Test connection

[Cancel](#)

8. 생성된 Data Source 를 선택합니다. Next.

The screenshot shows the 'Copy data' blade in the Azure portal. On the left, a navigation pane lists steps: Properties (selected), Source, Destination, Settings, Summary, and Deployment. The 'Source' section is expanded, showing 'Connection' (selected) and 'Dataset'. The 'Destination' section shows 'Connection' (selected) and 'Dataset'. The main area is titled 'Source data store' with the sub-instruction: 'Specify the source data store for the copy task. You can use an existing data store connection or specify a new data store.' Below this is a tab bar with 'All' (selected), 'Azure', 'Database', 'File', 'Generic protocol', 'NoSQL', and 'Services and apps'. A search bar 'Filter by name' and a 'Create new connection' button are present. A list of data stores includes 'CosmosDemo', 'kyowonsynapse-WorkspaceDefaultStorage', 'MySQL_Server' (selected and highlighted in blue), and 'sqlpo'.

9. Cardtransaction 테이블을 선택합니다. Next.

The screenshot shows the Microsoft Data Factory interface in a browser window. On the left, a sidebar lists six steps: 1. Source, 2. Destination, 3. Settings, 4. Summary, 5. Deployment, and 6. Preview. Step 1 is currently selected. The main area is titled "EXISTING TABLES" and contains a list of tables from a database. A search bar at the top says "Filter by name...". Below it is a "Refresh" button with a circular arrow icon. The table list includes:

- Select all
- `address`
- `buildversion`
- `cardtransaction`
- `customer`
- `customeraddress`

Below the table list is a "Preview" section with a table showing transaction data:

transactionID	accountID	transactionAmountUSD	transactionAmount	transactionCui
47F3C9CC- 2948-4069- B7BC- 260361539AAD	A914800993377925	86	86	USD
A949C89B- D346-4C42- AF18- 4F75FA0DD82C	A844427439174139	0	0	USD

10. Apply Filter 항목에서는 그냥 Next로 넘어갑니다.

11. Destination Source 는 sqlpool 을 선택합니다. Next.

The screenshot shows the 'Destination data store' configuration screen in the Azure Data Factory 'Copy data' interface. On the left, a sidebar lists steps 1 through 6: Properties, Source (MySQL), Destination, Settings, Summary, and Deployment. Step 2 (Source) is currently selected. The main area shows a list of destination data stores under the 'All' tab. One item, 'sqlpool', is highlighted in blue, indicating it is selected. Other items include 'CosmosDemo' and 'kyowonsynapse-WorkspaceDefaultStorage'. A 'Create new connection' button is also visible.

12. Table Mapping에서 use existing table 을 선택 후 dbo.CardTransaction 테이블을 선택합니다. Next.

Table mapping

For each table you have selected to copy in the source data store, select a corresponding table in the destination data store or specify the table structure.

The screenshot shows the 'Table mapping' interface. On the left, under 'Source', there is a single entry: 'cardtransaction'. An arrow points from this entry to a dropdown menu labeled '-Select-'. A search bar within this menu contains the placeholder 'Filter...'. Below the search bar is a list of destination tables, with 'dbo.CardTransaction' highlighted in blue. To the right of the highlighted table, a rectangular box contains the text 'dbo.CardTransaction'.

13. Column Mapping에서는 그냥 Next로 넘어갑니다.

14. Setting 화면에서 Staging Account Linked Service 의 +New 를 클릭합니다.

Settings

More options for data movement

Fault tolerance ⓘ

▲ Performance settings

Enable staging ⓘ

▲ Staging settings

Staging account linked service ⓘ + New

Enable Compression ⓘ

▲ Advanced settings

Allow PolyBase ⓘ

Reject type ⓘ

Reject value

Use type default

Data integration unit ⓘ
 Edit

You will be charged # of used DIUs * copy duration * \$0.25/DIU-hour. Local currency subscription type. [Learn more](#)

Degree of copy parallelism ⓘ
 Edit

15. 이름에 StagingBlob 으로 입력하고 ADLS Gen2 를 선택합니다. 마찬가지로 Test Connection 을 클릭하여 연결이 잘 되는지 확인하고 생성을 진행합니다.

Name *

StagingBlob

Description

Type *

Azure Blob Storage



Connect via integration runtime * ⓘ

AutoResolveIntegrationRuntime



Authentication method

Account key



Connection string

Azure Key Vault

Account selection method ⓘ

From Azure subscription Enter manually

Azure subscription ⓘ

Microsoft Azure Internal Consumption (2386ed0e-ac2f-41c7-958a-7397006dce05)



Storage account name *

kyowonadls



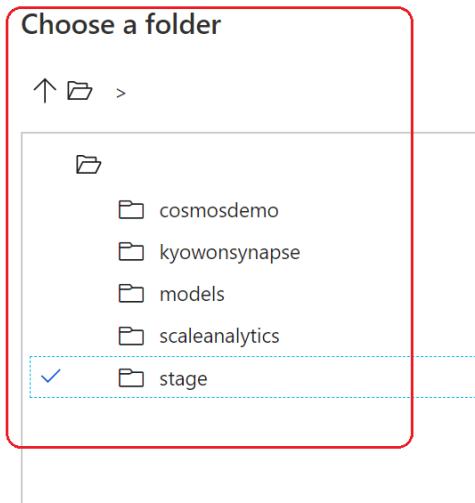
Additional connection properties

+ New

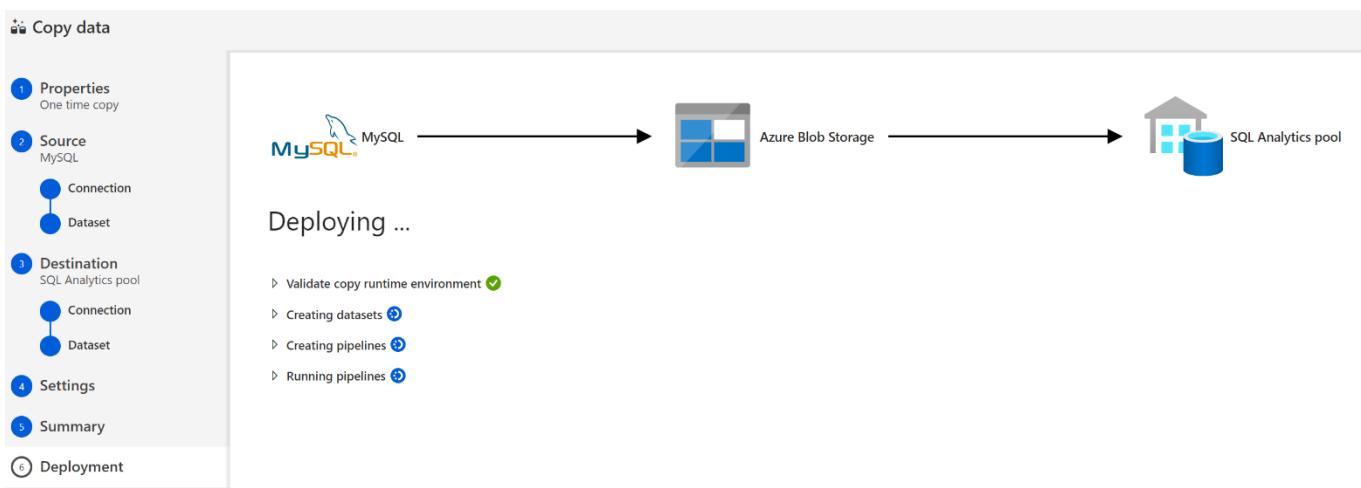
Test connection ⓘ

To linked service To file path

16. 바로 아래에 있는 Storage Path 에서 Browse 를 클릭하여 Stage 를 선택합니다.



17. Next 를 클릭하여 다음으로 넘어갑니다.
18. Summary Page 에서 검토 후 이상업스면 Next 로 넘어갑니다.
19. 아래와 같이 Pipeline 이 Deploy 되고 자동으로 실행됩니다.



20. Monitor 버튼을 클릭하여 모니터 페이지로 와서, 실행된 Pipeline Activity 의 Detail(안경 아이콘)을 클릭합니다.

Initial_Copy

The screenshot shows the Azure Data Factory (ADF) pipeline run interface. At the top, there are navigation tabs for 'List' and 'Gantt'. Below the tabs are several buttons: 'Rerun', 'Rerun from activity', 'Rerun from failed activity', and 'Refresh'. A modal window titled 'Copy data' is open, showing a green checkmark icon and the text 'Copy_fq4'. Below the modal are zoom controls (+, -, 100%, fit). The main area is titled 'Activity runs' and displays a single item: 'Pipeline run ID 583b49a6-0fbb-4ea9-b5e8-d5a8ca61c74e'. Underneath, a dropdown menu shows 'All status' with a downward arrow. It also indicates 'Showing 1 - 1 of 1 items'. A table lists the activity details:

ACTIVITY NAME	ACTIVITY TYPE	RUN START ↑	DURATION	STATUS	INTEGRATION RUNTIME
Copy_fq4	Copy	6/23/20, 6:19:04 AM	00:00:22	✓ Succeeded	DefaultIntegrationRuntime (South...

A red box highlights the 'Details' button at the bottom of the table row.

21. 아래와 같이 Detail 내용을 확인합니다.

Details ↻ Refresh ↗ ↘ ×

Learn more on copy performance details from here. Provide feedback on performance.

Activity run id: 7791bd5c-5c2a-4276-a0ff-1732f28e533d

MySQL → Azure Blob Storage

Succeeded
Azure IR region: Southeast Asia

Azure Blob Storage → Azure Synapse Analytics (formerly SQL DW)

Succeeded
Azure IR region: Southeast Asia

MySQL	Azure Blob Storage	Azure Synapse Analytics (formerly SQL DW)
Data read: 12.134 MB		Data written: 0 byte
Rows read: 31,364		Rows written: 31,364
Peak connections: 1		Throughput: 730.882 KB/s

Copy duration 00:00:17

MySQL → Azure Blob Storage

Start time	Used DIUs	Used parallel copies	Duration	Working duration	Total duration
6/23/20, 6:19:04 AM	4	1	00:00:06	00:00:02	00:00:06
✓ Queue ⓘ				Time to first byte ⓘ 00:00:00	
✓ Transfer ⓘ				Reading from source ⓘ 00:00:01	
				Writing to sink ⓘ 00:00:00	

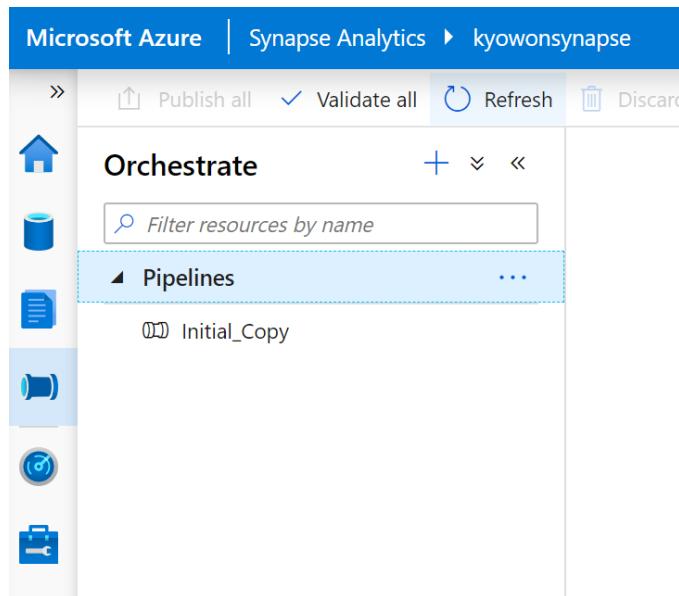
Azure Blob Storage → Azure Synapse Analytics (formerly SQL DW)

22. Develop 탭으로 돌아와서 06_OLTP_Table_Count 스크립트를 실행하여 총 Migration 건수를 확인합니다.

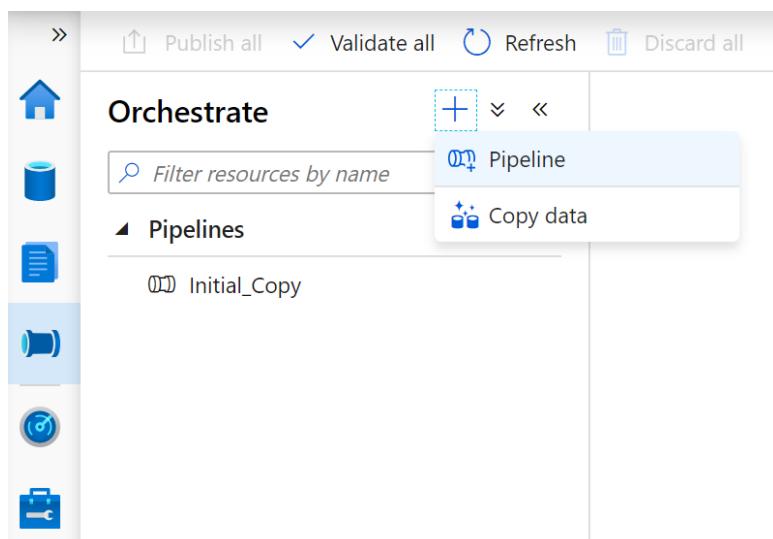
작업 7 : Azure Synapse Pipeline

이번 작업에서는 MySQL Server VM에 있는 Transaction Table의 Incremental Copy를 Pipeline을 통해 실행합니다.

1. Synapse Studio에서 왼쪽에 Integrate 탭을 선택하여 Refresh 합니다. 조금 전에 Ingest Wizard를 통해 생성된 Pipeline이 보일 것입니다.



2. 다음으로 +를 선택하고 Pipeline을 클릭하여 새로운 Pipeline을 생성합니다.



3. Property 탭에 이름에 Incremental_Copy로 입력합니다.

Properties

General

i Choose a name for your pipeline.
This name can be updated at any time until it is published.

Name *

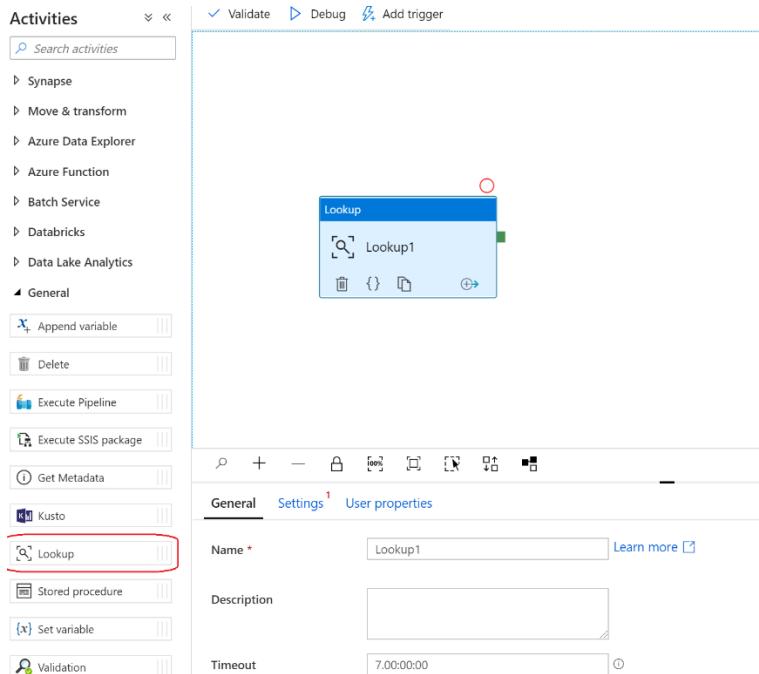
Description

Concurrency

Annotations

+ New

4. 다음으로, Activities 항목안에 General -> LookUp 을 Drag&Drop 하여 끌어옵니다.



5. LookUp1 의 Setting 탭으로 와서, 조금 전 Ingest Wizard 에서 생성된 Destination Data Set(sqlpool)을 선택합니다.

The screenshot shows the 'Settings' tab for a dataset. The 'Source dataset *' dropdown is open, displaying two options: 'Select...' and 'DestinationDataset_fq4'. The 'DestinationDataset_fq4' option is highlighted with a red box. Other tabs like 'General' and 'User properties' are visible at the top.

6. Use Query 에 Query 를 선택하고, 아래 Query 항목에 SELECT MAX([transactionDateTime]) AS transactionDateTime FROM [dbo].[CardTransaction]; 를 입력합니다.

The screenshot shows the 'Settings' tab for a dataset. The 'Source dataset *' dropdown is set to 'DestinationDataset_fq4'. Under the 'Use query' section, the 'Query' radio button is selected. The 'Query' text area contains the following SQL code:

```
SELECT MAX([transactionDateTime]) AS transactionDateTime
FROM [dbo].[CardTransaction];
```

The 'First row only' checkbox is checked.

7. Preview Data 를 클릭하여 값을 확인해봅니다.

Preview data

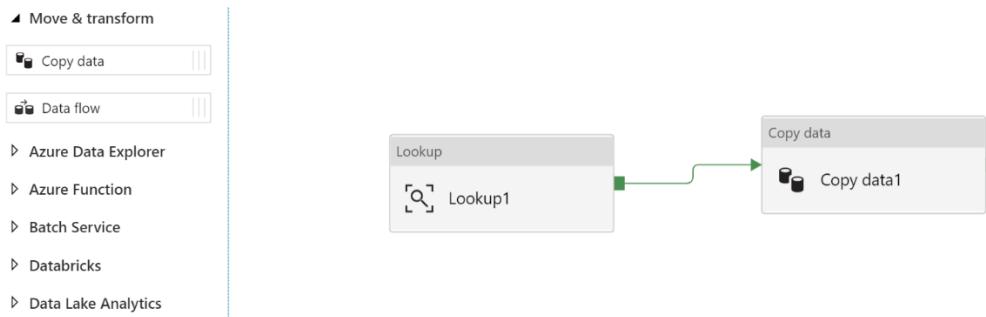
Linked service: sqlpool

Object:

transactionDateTime

2013-05-26T15:44:00

8. 다음은 Activities 항목의 Move & Transform 항목에서 Copy Data 를 Drag & Drop 하여 옮겨놓고, Lookup1 로부터 Output 을 연결시켜줍니다.



9. 다음은 Copy data1 을 선택하여 활성화 하고, Source 탭에서 Ingest Wizard 에서 생성된 SourceDataSet(MySQL)을 선택합니다. 쿼리에는 SELECT * FROM cohoOLTP.CardTransaction WHERE transactionDateTime > '@{activity('Lookup1').output.firstRow.transactionDateTime}'; 로 입력합니다.

This screenshot shows the 'Source' tab of a data activity configuration. At the top, there are various toolbar icons. Below them, tabs include General, Source (which is selected), Sink, Mapping, Settings, and User properties. Under the Source tab, the 'Source dataset' dropdown is set to 'SourceDataset_fq4'. The 'Query' field contains the following SQL code:

```
SELECT * FROM cohoOLTP.CardTransaction WHERE transactionDateTime >
'@{activity('Lookup1').output.firstRow.transactionDateTime}';
```

Below the query, there's a link 'Add dynamic content [Alt+P]'. At the bottom, there are buttons for 'Additional columns' and 'New'.

10. 다음으로 Sink 탭으로 이동하여 목적지를 sqlpool 로 선택합니다.

This screenshot shows the 'Sink' tab of a data activity configuration. The tabs at the top are General, Source, Sink (selected), Mapping, Settings, and User properties. Under the Sink tab, the 'Sink dataset' dropdown is open, showing options like 'Select...', 'DestinationDataset_fq4', and another 'Select...' option. The 'DestinationDataset_fq4' option is highlighted.

11. 마지막으로 Setting 탭에서 Enable Staging에 체크 후 Ingest Wizard에서 했던 과정과 동일하게 설정합니다.

The screenshot shows the 'Settings' tab of a Data integration unit configuration. The 'Enable staging' checkbox is checked and highlighted with a red box. Other settings shown include 'Data integration unit' set to 'Auto', 'Degree of copy parallelism' with a checked 'Edit' button, and 'Fault tolerance'. Below these, under 'Staging settings', there is a section for 'Staging account linked service' (set to 'StagingBlob') with a 'Test connection' button, and a 'Storage Path' input field containing 'stage' with a 'Browse' button.

12. 이제 Publish All 을 클릭하여 Publish 합니다.

13. Add Trigger -> Trigger Now 를 클릭하여 바로 실행합니다.

The screenshot shows the triggers section of the Azure Data Factory interface. The 'Trigger now' option is selected and highlighted with a blue background. Below it is a 'New/Edit' button. At the bottom, there is a visual representation of a pipeline with a 'Lookup' activity followed by a 'Copy data' activity.

14. 왼쪽의 모니터링 탭으로 이동하여 실행중인 파이프라인을 모니터 할 수 있습니다.

The screenshot shows the Microsoft Azure Synapse Analytics interface. On the left, there's a sidebar with icons for Home, Orchestration, Trigger runs, Integration runtimes, Activities (Apache Spark applications and SQL requests), and a briefcase icon. The 'Orchestration' section is expanded, showing 'Pipeline runs' which is selected and highlighted in blue. The main area is titled 'Pipeline runs' and shows the following details:

- Time: Last 24 hours (6/22/20 7:12 AM - 6/23/20 7:12 AM)
- Time zone: Seoul (UTC+9)
- Runs: Latest runs
- Buttons: Rerun, Cancel, Refresh, Edit columns
- Status: All status
- Showing 1 - 2 items
- Table headers: PIPELINE NAME, RUN START, DURATION, TRIGGERED BY, STATUS, PARAMETERS
- Items:
 - Incremental_Copy**: RUN START: 6/23/20, 7:11:43 AM, DURATION: 00:00:28, TRIGGERED BY: Manual trigger, STATUS: In progress
 - Initial_Copy**: RUN START: 6/23/20, 6:19:00 AM, DURATION: 00:00:26, TRIGGERED BY: Manual trigger, STATUS: Succeeded

15. 바로 전 실행한 Incremental_Copy 를 선택하여 Activity 별 Detail 내용을 확인하실 수 있습니다.

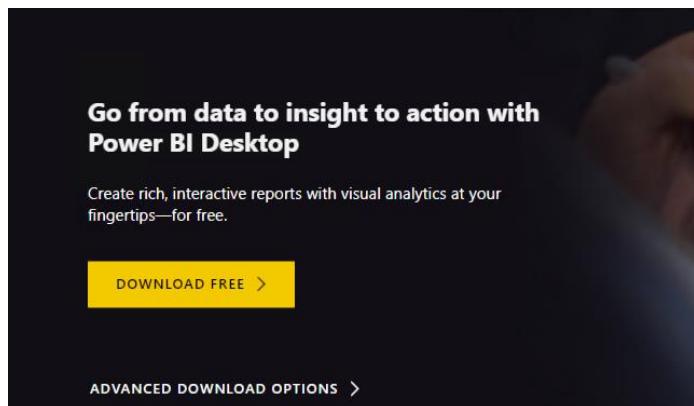
The screenshot shows the Microsoft Azure Synapse Analytics interface. At the top, there's a toolbar with icons for adding (+), subtracting (-), and filtering (magnifying glass). Below it, the title is 'Activity runs' and the pipeline run ID is listed as b86bb0a7-d01a-42f3-b0e1-90537b748469. The sidebar on the left is identical to the previous screenshot. The main area shows the following details:

- All status
- Showing 1 - 2 of 2 items
- Diagram: A flowchart showing a 'Lookup' activity named 'Lookup1' connected by an arrow to a 'Copy data' activity named 'Copy data1'.
- Table headers: ACTIVITY NAME, ACTIVITY TYPE, RUN START, DURATION, STATUS, INTEGRATION RUNTIME, USER PROPERTIES, ERROR, RUN ID
- Items:
 - Copy data1**: ACTIVITY TYPE: Copy, RUN START: 6/23/20, 7:11:50 AM, DURATION: 00:00:22, STATUS: Succeeded, INTEGRATION RUNTIME: DefaultIntegrationRuntime (Southeast Asia);DefaultInteg, RUN ID: df743632-b
 - Lookup1**: ACTIVITY TYPE: Lookup, RUN START: 6/23/20, 7:11:46 AM, DURATION: 00:00:04, STATUS: Succeeded, INTEGRATION RUNTIME: AutoResolveIntegrationRuntime (Southeast Asia), RUN ID: 840b55a9-7

16. Develop 탭으로 돌아와서 06_OLTP_Table_Count 스크립트를 실행하여 총 Migration 건수를 확인합니다.

작업 8 : Power BI 로 Data Visualization.

1. 웹 브라우저에서 Power BI Desktop 다운로드 페이지 (<https://powerbi.microsoft.com/en-us/desktop/>) 로 이동합니다.
2. 페이지 중간에서 Download Free 링크를 선택합니다.



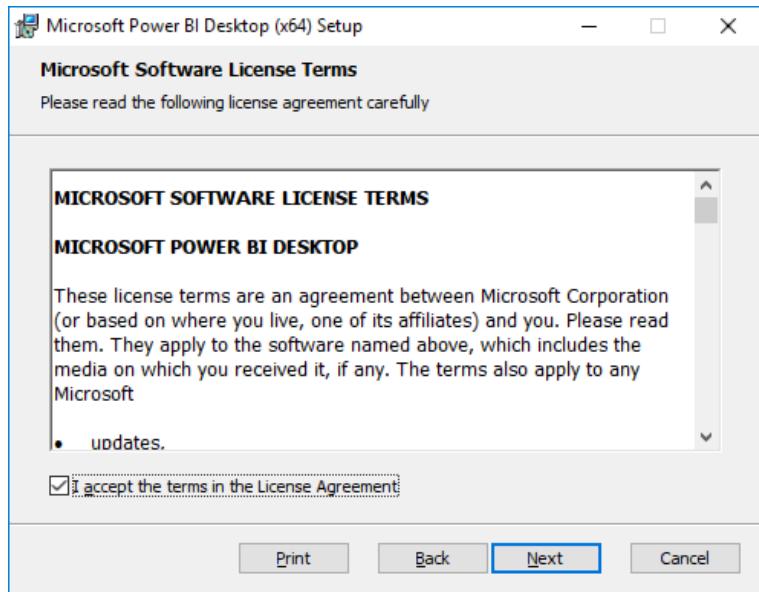
3. 설치파일을 실행합니다.



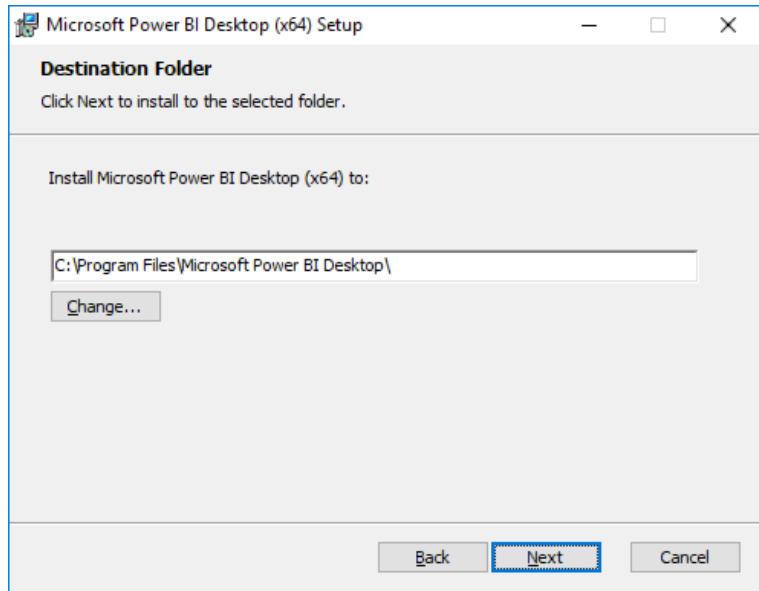
4. Welcome 화면에서 다음을 클릭합니다.



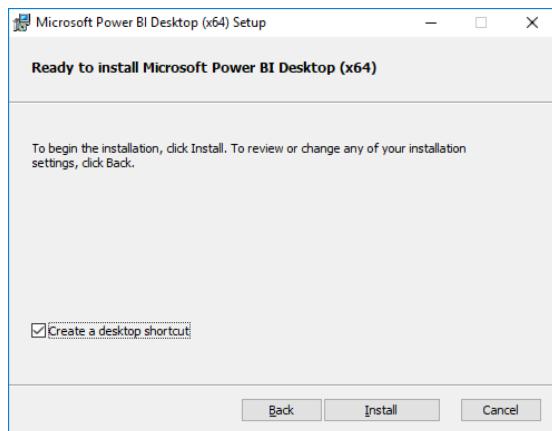
5. 사용권 계약에 동의하고 다음을 선택합니다.



6. 지정 폴더를 디폴트로 두고 다음을 선택합니다.



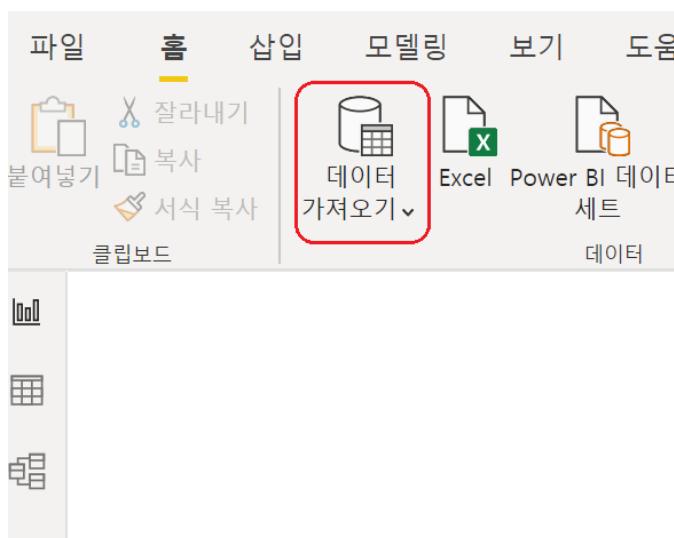
7. 바탕 화면 바로 가기 만들기 상자가 선택되어 있는지 확인하고 설치를 선택합니다.



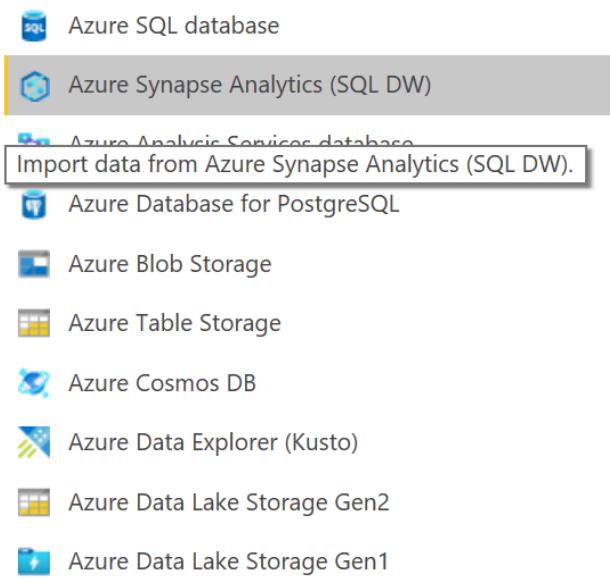
8. Microsoft Power BI Desktop 시작이 선택되어 있는지 확인하고 마침을 선택합니다.



9. Power BI에서 데이터 가져오기를 클릭합니다.



10. Azure 를 선택 후 Azure Synapse Analytics (SQL DW)를 선택합니다. 연결합니다.



11. Azure Portal 의 Synapse Workspace 의 Overview Page 로 가서, SQL Endpoint 를
복사합니다.

Resource group (change) : KyoWon_HoL

Status : Succeeded

Location : Korea Central

Subscription (change) : Microsoft Azure Internal Consumption

Subscription ID : 2386ed0e-ac2f-41c7-958a-7397006dce05

Managed virtual network : No

Managed Identity object ... : d5e3ee85-c819-4206-a3ec-650c6ce4a3f6

Workspace web URL : <https://web.azuresynapse.net?workspace=%2bsubscriptio...>

Tags (change) : Click here to add tags

Firewalls : Show firewall settings

Primary ADLS Gen2 acco... : <https://kyowonadls.dfs.core.windows.net>

Primary ADLS Gen2 file s... : synapse

SQL admin username : sqladminuser

SQL Active Directory ad... : yongl@microsoft.com

Dedicated SQL endpoint : **kyowon-synapse.sql.azuresynapse.net**

Serverless SQL endpoint : kyowon-synapse-ondemand.sql.azuresynapse.net

Development endpoint : <https://kyowon-synapse.dev.azuresynapse.net>

Getting started

- Open Synapse Studio Start building your fully-integrated analytics solution and unlock new insights.
- Read documentation Learn how to be productive quickly. Explore concepts, tutorials, and samples.

12. 복사한 값을 서버 항목에 붙여넣고, Direct Query 를 선택합니다.

SQL Server database

Server ⓘ
kyowon-synapse.sql.azuresynapse.net

Database (optional)

Data Connectivity mode ⓘ
 Import
 DirectQuery

13. Synapse Workspace 때 입력했던 Admin 계정으로 Database Login 을 하거나, Microsoft 계정으로 Login 을 하고 연결합니다.

14. DimGeography 테이블을 선택하고 로드 합니다.

탐색 창

The screenshot shows the Power BI Data Explorer interface. On the left, a tree view lists various dimensions and tables, with 'DimGeography' selected and highlighted with a yellow checkmark. On the right, a table named 'DimGeography' is displayed with columns: GeographyKey, City, StateProvinceCode, and StateProvinceName. The table contains 38 rows of data. At the bottom, there are buttons for '관련 테이블 선택' (Select Related Tables), '로드' (Load), '데이터 변환' (Data Transformation), and '취소' (Cancel).

GeographyKey	City	StateProvinceCode	StateProvinceName
10	Newcastle	NSW	New South Wales
22	East Brisbane	QLD	Queensland
228	Basingstoke Hants	ENG	England
514	Cheektowaga	NY	New York
95	Toronto	ON	Ontario
295	Baldwin Park	CA	California
487	Saint Louis	MO	Missouri
565	La Vergne	TN	Tennessee
119	Grevenbroich	BY	Bayern
475	Branch	MN	Minnesota
3	Darlinghurst	NSW	New South Wales
192	Paris	75	Seine (Paris)
482	Ferguson	MO	Missouri
138	Berlin	HH	Hamburg
293	Alpine	CA	California
507	Santa Fe	NM	New Mexico
12	North Sydney	NSW	New South Wales
352	Palo Alto	CA	California
474	Zeeland	MI	Michigan
569	Nashville	TN	Tennessee
255	London	ENG	England
500	Winston-Salem	NC	North Carolina
387	Englewood	CO	Colorado

15. 오른쪽 필드 항목에서 CountryRegionCode 를 체크합니다. Map 시작화가 자동으로 시작될 것입니다.

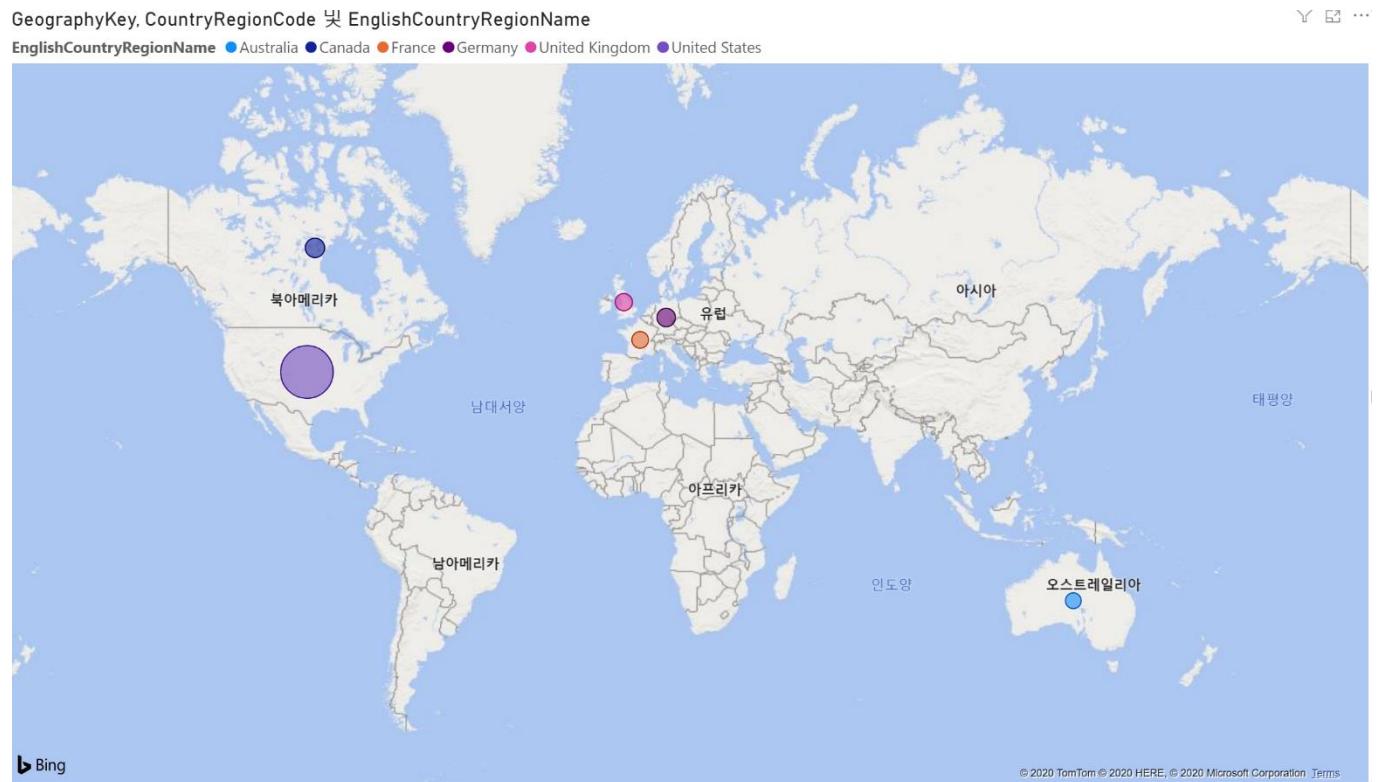
The screenshot shows the Power BI Fields pane. The title is '필드'. Below it is a search bar labeled '검색'. A tree view shows the 'DimGeography' dimension expanded, with its fields listed: City, CountryRegionCode (which is checked), EnglishCountryRegionName, FrenchCountryRegionName, GeographyKey, PostalCode, SalesTerritoryKey, and SpanishCountryRegionName.

16. 조금 더 구체적인 정보를 표시하기 위해, EnglishCountryRegionName 을 범례로, CountryRegion 을 크기로 Drag&Drop 합니다.

The screenshot shows the Power BI Fields pane. On the left, there are several icons representing different data types and operations. Below them, there are sections for '위치' (Location), '범례' (Legend), '위도' (Latitude), '경도' (Longitude), and '크기' (Size). In the '범례' section, there are two items: 'CountryRegionCode' and 'EnglishCountryRegionName'. The 'CountryRegionCode' item is currently selected, indicated by a yellow border around its box. On the right, the '필드' (Fields) pane is open, showing the 'DimGeography' table with the following fields:

- City
- CountryRegionCode
- EnglishCountryRegionName
- FrenchCountryRegionName
- GeographyKey
- PostalCode
- SalesTerritoryKey
- SpanishCountryRegionName
- StateProvinceCode
- StateProvinceName

17. 아래와 같은 시각화를 확인하실 수 있습니다.



작업 9: Create & Populate Customer Information Table

1. 왼쪽 메뉴의 Develop 탭을 선택하고 11_01_Create_Customer_Info_Table 스크립트를 열어 실행합니다.
2. Data 탭으로 이동하여 + 버튼을 누르고 Data Set 을 생성합니다.

The screenshot shows the Azure Synapse Studio interface. The top navigation bar includes 'Synapse live', 'Validate all', and 'Publish all' buttons. Below the navigation is a 'Data' tab section. On the left, there's a vertical sidebar with icons for Home, Databases, Workspaces, Pipelines, and Integration datasets. The main area displays a 'Workspace' section with a search bar and a 'Linked' section containing 'SQL database', 'Connect to external data', and 'Integration dataset'. Under 'Integration datasets', two datasets are listed: 'DestinationDataset_ygh' and 'SourceDataset_ygh'. A context menu is open over the 'Integration datasets' section, showing options like 'Create', 'Generate', 'Preview', 'Browse gallery', and 'Script' (with sub-options 'WITH (DISTINCT, CLUSTERED)', 'GO'). The menu has items numbered 9 through 14.

3. ADLS Gen2 를 선택합니다. Next

New dataset

Select a data store

All Azure Database File Generic protocol NoSQL Services and apps

The screenshot shows a grid of nine data storage options. The first row contains: Azure Blob Storage (with a Windows icon), Azure Cosmos DB (MongoDB API) (with a globe and nodes icon), and Azure Cosmos DB (SQL API) (with a globe and nodes icon). The second row contains: Azure Data Explorer (Kusto) (with a blue and green icon), Azure Data Lake Storage Gen1 (with a cylinder icon), and Azure Data Lake Storage Gen2 (with a Windows icon, highlighted with a blue border). The third row contains three icons that are mostly obscured by shadows.

4. Delimited Text 를 선택합니다. Next

Select format

Choose the format type of your data

The screenshot shows a grid of six data format options. The first row contains: Parquet (with a blue grid icon), DelimitedText (with a CSV icon, highlighted with a red border), and JSON (with a JSON icon). The second row contains: Avro (with an Avro icon), ORC (with an ORC icon), and Binary (with a document icon).

5. 아래와 같이 Property 를 설정합니다.

Field	Value
Name	Enter asamcw_customerinfo_csv

Field	Value
Linked service	Select < ADLS Gen2 Account Name >
File Path - Container	Enter wwi-02
File Path - Directory	Enter customer-info
File Path - File	Enter customerinfo.csv
First row as header	Checked
Import schema	Select From connection/store

6. 다시 + 버튼을 누르고 Data Set 을 생성합니다.
7. Azure 를 선택하고, **Azure Synapse dedicated SQL pool** 을 선택합니다.

Select a data store

The screenshot shows the 'Select a data store' interface. At the top, there is a search bar with the text 'azure sy'. Below the search bar, there are several tabs: 'All', 'Azure', 'Database', 'File', 'Generic protocol', 'NoSQL', and 'Se...'. Under the 'Azure' tab, there are two items listed: 'Azure Synapse Analytics' and 'Azure Synapse dedicated SQL pool'. The 'Azure Synapse dedicated SQL pool' item is highlighted with a red rectangular box around its icon and text.

8. 아래와 같이 Property 를 설정합니다.

Field	Value
Name	Enter asamcw_customerinfo_asa
Linked service	Select WorkspaceDefault
Table name	Select WWICustomerInfo.
Import schema	Select From connection/store

Set properties

i Choose a name for your dataset. This name can be updated at any time before publishing.

Name

Azure Synapse dedicated SQL pool * [\(i\)](#)

Table name

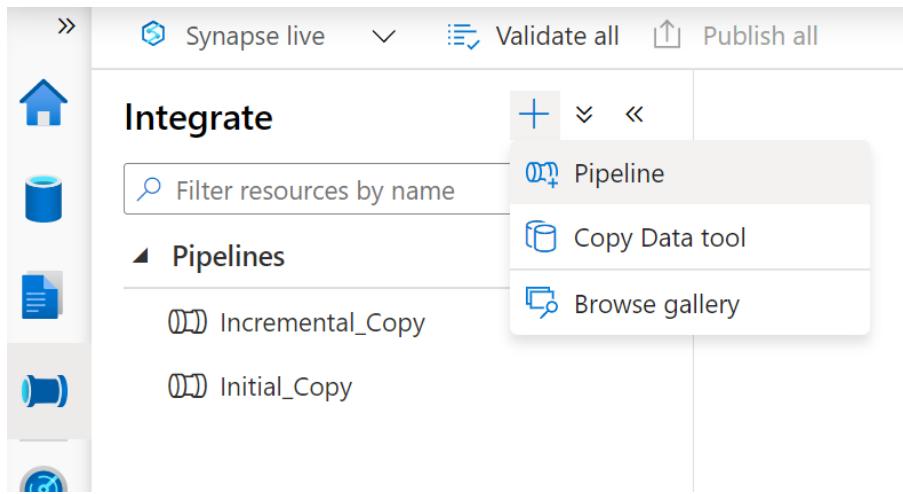
Edit

Import schema

From connection/store None

9. Publish All 하여 새로 생성한 두 Data Set 을 저장합니다.

10. Integrate 탭으로 이동하여 +를 누르고 새로운 Pipeline 을 생성합니다.



11. 이름에 **Copy Customer Information** 을 입력합니다.

Properties

General

i Choose a name for your pipeline.
This name can be updated at any
time until it is published.

Name *

Description

Concurrency

(i)

Annotations

+ New

12. Move & Transform에서 Copy Data 를 Drag & Drop 합니다.

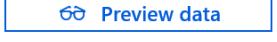
The screenshot shows the 'Activities' list in the Azure Data Factory interface. The 'Move & transform' section is highlighted with a red box, and a red arrow points to the 'Copy data' item within it. Other items listed include 'Data flow' and several other categories like 'Synapse', 'Azure Data Explorer', etc.

13. Copy Data Activity 를 선택하여 General 탭에서 이름을 입력합니다.

The screenshot shows the 'General' tab for a new activity. The 'Name' field is filled with 'Copy Customer Information' and is highlighted with a red box. Other fields visible include 'Description', 'Timeout' (set to 7.00:00:00), 'Retry' (set to 0), 'Retry interval' (set to 30), 'Secure output' (unchecked), and 'Secure input' (unchecked). A 'Learn more' link is also present.

14. Source 탭에서 아래와 같이 미리 생성한 Data Set 을 선택합니다.

General Source Sink Mapping Settings User properties

Source dataset *  Open + New 

File path type File path in dataset Wildcard file path Prefix List of files 

Filter by last modified Start time (UTC) End time (UTC) 

Recursively 

Enable partition discovery

Max concurrent connections 

Skip line count

Additional columns 

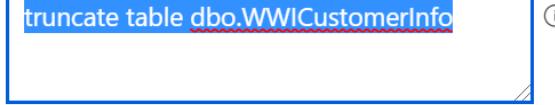
15. Sink 탭에서 마찬가지로 Destination Source 를 선택하고, Bulk Insert 를 선택합니다.
Pre-Copy Script 에는 truncate table dbo.WWICustomerInfo 를 입력합니다.

General Source **Sink** Mapping Settings User properties

Sink dataset *  Open + New

Copy method PolyBase  Copy command (Preview)  Bulk insert

Table option None Auto create table 

Pre-copy script 

Add dynamic content [Alt+P]

16. Mapping 탭으로 이동하여 Import Schema 버튼을 클릭하고 확인합니다.

Source	Type	Destination	Type
UserName	String	UserName	nvarchar
Gender	String	Gender	nvarchar
Phone	String	Phone	nvarchar
Email	String	Email	nvarchar
CreditCard	String	CreditCard	nvarchar

17. Publish All 하고 Trigger 하여 실행합니다.

18. Monitor 탭에서 수행 결과를 확인하고 11_02_Select_Customer_Info_Table 스크립트를 수행하여 결과를 확인합니다.

작업 10: Create & Populate Campaign Analytics Table

1. 왼쪽 메뉴의 Develop 탭을 선택하고 12_01_Create_Campaign_Analytics_Table 스크립트를 열어 실행합니다.

2. Data 탭으로 이동하여 + 버튼을 누르고 Data Set 을 생성합니다.
3. 이전과 동일하게 ADLS Gen2 를 선택 후 Delimited Text 를 선택합니다.
4. 아래와 같이 Property 를 입력합니다.

Field	Value
Name	Enter asamcw_campaignanalytics_csv
Linked service	Select <ADLS Gen2 Account Name>
File Path - Container	Enter wwi-02
File Path - Directory	Enter campaign-analytics
File Path - File	Enter campaignanalytics.csv
First row as header	Unchecked
Import schema	Select From connection/store

5. 다시 +를 누르고 Data Set 을 생성합니다.
6. 바로 전 작업과 동일하게 **Azure Synapse dedicated SQL pool** 를 선택합니다.
7. 아래와 같이 Property 를 입력합니다.

Field	Value
Name	Enter asamcw_campaignanalytics_asa
Linked service	Select WorkspaceDefault
Table name	Select WWICampaignAnalytics
Import schema	Select From connection/store

Name

asamcw_campaignanalytics_asa

Azure Synapse dedicated SQL pool * ⓘ

sqlpool

Table name

dbo.WWICampaignAnalytics

Edit

Import schema

From connection/store None

8. Publish All 하여 저장합니다.

9. Develop 탭으로 돌아와서 +를 누르고 Data Flow 를 생성합니다.

The screenshot shows the Microsoft Azure Synapse Analytics interface. At the top, it says "Microsoft Azure | Synapse Analytics > kyowon-synapse". Below that is a toolbar with icons for "Synapse live", "Validate all", and "Publish all". On the left, there's a sidebar with icons for Home, Databases, Tables, Pipelines, and Jobs. The main area is titled "Develop" and contains a search bar "Filter resources by name". Under "SQL scripts", there are five items: 00_CreateDataLoader, 01_CreateDWTables, 02_CreateCredential, 03_CreateExternalTables, and 04_LoadData. A context menu is open over the "Data flow" item, listing "SQL script", "Notebook", "Data flow" (which is highlighted), "Apache Spark job definition", "Browse gallery", and "Import".

10. 이름에 **Campaign_Analytics_Data** 를 입력합니다.

Properties

General

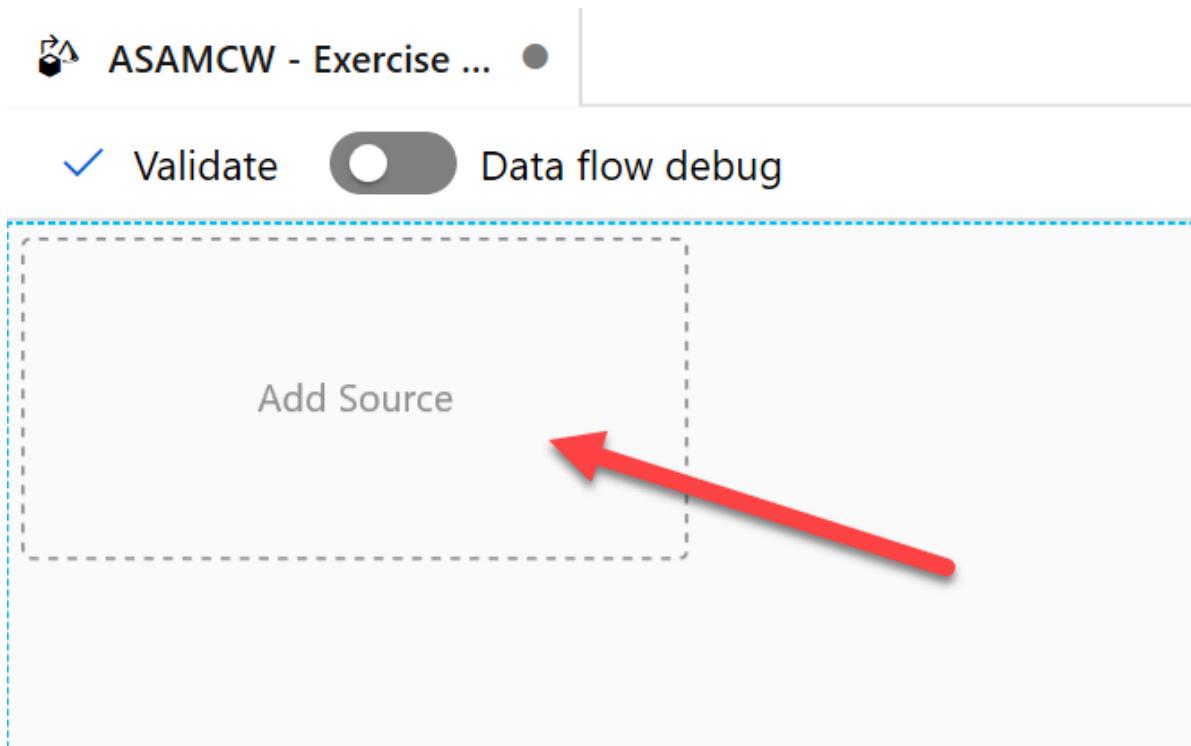
i Choose a name for your data flow.
This name can be updated at any time until it is published.

Name *

Campaign_Analytics_Data

Description

11. 화면에 보이는 Add Source 를 클릭합니다.



12. 아래와 같이 Property 를 입력합니다.

Field	Value
Output stream name	Enter campaignanalyticscsv
Source type	Select Dataset
Dataset	Select asamcw_campaignanalytics_csv .
Skip line count	Enter 1

Source settings Source options Projection Optimize Inspect Data preview

Output stream name * campaignanalyticscsv [Learn more](#)

Source type * Dataset [Test connection](#) [Open](#) [New](#)

Dataset * asamcw_campaignanalytics_csv

Options

- Allow schema drift ⓘ
- Infer drifted column types ⓘ
- Validate schema ⓘ

Skip line count 1

Sampling * Enable Disable ⓘ

13. Data Flow 를 만들 때 데이터 미리보기 및 스키마 가져오기(프로젝션)와 같은 디버그를 설정하여 특정 기능을 사용할 수 있습니다만 시간적, 환경적 제약으로 인해 이런 과정은 Skip 합니다. Data Source 에는 일반적으로 정의 해줘야 하는 Schema 가 있으며, Data Flow 의 Designer Toolbar 도구모음에서도 Script 로 구성이 가능합니다.



14. 스크립트 안의 내용을 아래 스크립트로 수정합니다.

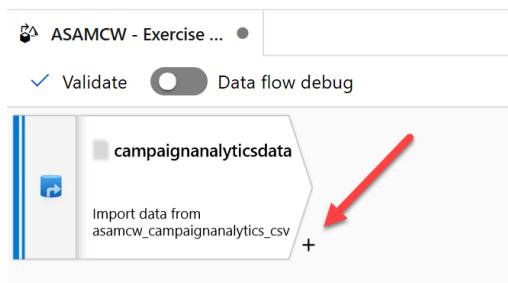
```
source(output(
    {__col0__} as string,
    {__col1__} as string,
    {__col2__} as string,
    {__col3__} as string,
    {__col4__} as string,
    {__col5__} as double,
    {__col6__} as string,
    {__col7__} as double,
    {__col8__} as string,
    {__col9__} as string
),
allowSchemaDrift: true,
validateSchema: false,
skipLines: 1) ~> campaignanalyticscsv
```

Note: 소스파일의 헤더가 잘못되어 매핑을 변경하는 시나리오 입니다.

15. Projection 탭으로 이동하여 아래와 같이 나타나는지 확인합니다.

Source settings	Source options	Projection	Optimize	Inspect	Data preview
		Define default format	Detect data type	Import projection	Reset schema
Column name		Type		Format	
col0		abc string	▼	Specify format	▼
col1		abc string	▼	Specify format	▼
col2		abc string	▼	Specify format	▼
col3		abc string	▼	Specify format	▼
col4		abc string	▼	Specify format	▼
col5		1.2 double	▼	Specify format	▼
col6		abc string	▼	Specify format	▼
col7		1.2 double	▼	Specify format	▼
col8		abc string	▼	Specify format	▼
col9		abc string	▼	Specify format	▼

16. 이제 아래 +를 클릭하여 Select 를 선택합니다.



17. Output Steam Name 에 를 **mapcampaignanalytics** 입력합니다.

18. Name as 아래에 있는 입력란에 아래와 같이 순서대로 입력합니다.

- Region
- Country

- ProductCategory
- CampaignName
- RevenuePart1
- Revenue
- RevenueTargetPart1
- RevenueTarget
- City
- State

The screenshot shows the 'Map' blade in the Azure Data Factory Mapping Editor. The 'Output stream name' is set to 'mapcampaignanalytics'. The 'Incoming stream' is 'campaignanalyticscsv'. Under 'Options', 'Skip duplicate input columns' and 'Skip duplicate output columns' are checked. The 'Input columns' section shows 10 input columns from 'campaignanalyticscsv's column, each mapped to an output column with a red border. The output columns are: Region, Country, ProductCategory, CampaignName, RevenuePart1, Revenue, RevenueTargetPart1, RevenueTarget, City, and State.

Input columns *	Name as
abc _col0_	Region
abc _col1_	Country
abc _col2_	ProductCategory
abc _col3_	CampaignName
abc _col4_	RevenuePart1
1.2 _col5_	Revenue
abc _col6_	RevenueTargetPart1
1.2 _col7_	RevenueTarget
abc _col8_	City
abc _col9_	State

19. 다시 +를 클릭하여 Derived Column을 선택합니다.

20. 이름을 **convertandaddcolumns**로 입력합니다.

21. 아래를 참고하여 입력합니다.

Column	Expression	Description
Revenue	<code>toDecimal(replace(concat(toString(R evenuePart1), toString(Revenue)), '\\", \"'), 10, 2, '\$###,###.##')</code>	Concatenate the RevenuePart1 and Revenue fields, replace the invalid \ character, then convert and format the data to a decimal type.

Column	Expression	Description
RevenueTarget	toDecimal(replace(concat(toString(RevenueTargetPart1), toString(RevenueTarget)), '\\', ''), 10, 2, '\$###,###.##')	Concatenate the RevenueTargetPart1 and RevenueTarget fields, replace the invalid \ character, then convert and format the data to a decimal type.
Analyst	iif(isNull(City), "", replace('DataAnalyst' + City, ' ', ''))	If the city field is null, assign an empty string to the Analyst field, otherwise concatenate DataAnalyst to the City value, removing all spaces.

Derived column's settings [Optimize](#) [Inspect](#) [Data preview](#)

Output stream name * [Learn more](#)

Incoming stream *

Columns * [①](#)

<input type="text" value="Revenue"/>	<input type="text" value="toDecimal(replace(concat(toString(RevenuePart1), toString(Revenue)), '\\', ''), 10, 2, '\$###,###.##')"/>	+ Edit Delete
<input type="text" value="RevenueTarget"/>	<input type="text" value="toDecimal(replace(concat(toString(RevenueTargetPart1), toString(RevenueTarget)), '\\', ''), 10, 2, '\$###,###.##')"/>	+ Edit Delete
<input type="text" value="Analyst"/>	<input ",="" '="" ''))"="" ',="" +="" city,="" replace('dataanalyst'="" type="text" value="iif(isNull(City), "/>	abc + Edit Delete

22. 다시 +를 클릭하고 Select 를 선택합니다.

23. 이름에 **selectcampaignanalyticscolumns** 입력하고 **RevenuePart1** 과 **RevenueTargetPart1** 를 지웁니다.

Select settings Optimize Inspect Data preview

Output stream name * Learn more [?](#)

Incoming stream * [?](#)

Options Skip duplicate input columns [?](#)
 Skip duplicate output columns [?](#)

Input columns * [Auto mapping](#) [Reset](#) [Add mapping](#) [Delete](#)

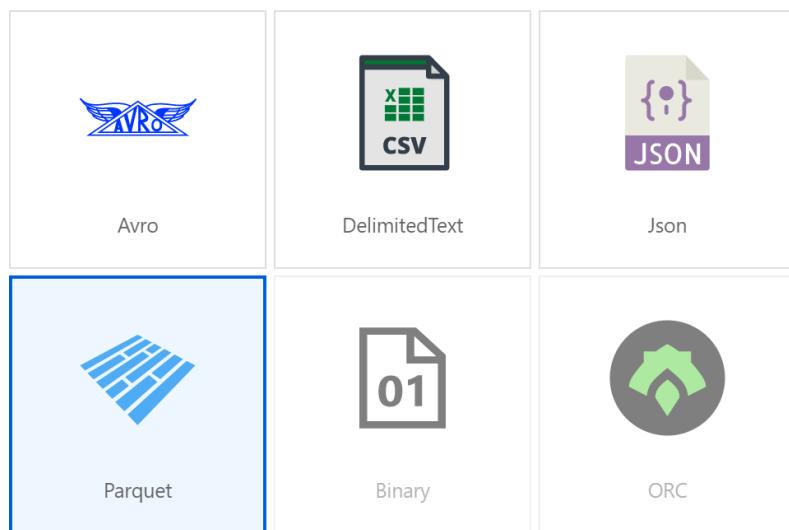
convertandaddcolumns's column	Name as
abc Region	Region
abc Country	Country
abc ProductCategory	ProductCategory
abc CampaignName	CampaignName
e* Revenue	Revenue
e* RevenueTarget	RevenueTarget
abc City	City
abc State	State
abc Analyst	Analyst

24. 마지막으로 +를 클릭하고 Sink 를 선택합니다.

25. 이름에 **campaignanalyticsasa** 를 입력하고 Destination Data Set 을 선택할 때 +New 를 눌러 새로 생성합니다. 이번에는 ADLS Gen2 의 Parquet 을 선택합니다.

Select format

Choose the format type of your data



26. 이름에 **asamcw_campaignanalytics_parquet** 를 입력하고 아래와 같이 Property 값을 넣어줍니다.

Set properties

i Choose a name for your dataset. This name can be updated at any time until it is published.

Name

asamcw_campaignanalytics_parque

Linked service *

kyowonsynapse-WorkspaceDefaultStorage

**File path**

wwi-02

/ campaign-analytics

/ File

**Import schema** From connection/store From sample file None

▷ Advanced

Sink [Settings](#) [Mapping](#) [Optimize](#) [Inspect](#) [Data preview](#)

Output stream name *	campaignanalyticsasa	Learn more
Incoming stream *	selectcampaignanalyticscolumns	
Sink type *	Dataset	
Dataset *	asamcw_campaignanalytics_parquet	Test connection Open New
Options	<input checked="" type="checkbox"/> Allow schema drift <input type="checkbox"/> Validate schema	

27. 위와 같이 된 것을 확인 후 Setting 탭으로 가서 File Name Option에서 **Output to one single file** 을 선택합니다. 파일 명에는 **campaignanalytics.parquet** 를 입력합니다. (Set Single Partition 도 클릭합니다)

Sink Settings Mapping Optimize Inspect Data preview

Clear the folder Add dynamic content [Alt+P]

File name option * Default Pattern Per partition As data in column Output to single file

Output to single file * campaignanalytics.parquet ⓘ

28. Publish All 하여 저장합니다.
29. Integrate 탭에서 +를 클릭 후 Pipeline 을 생성합니다. 이름을 **Copy Campaign Analytics Data** 로 입력합니다.
30. Data Flow 를 Drag & Drop 하고 방금 생성한 Data Flow 를 선택합니다.

ASAMCW - Exercise ... X

Activities ⌂ << Validate ⌂ Debug

Search activities

▶ Synapse

▶ Move & transform

Copy data

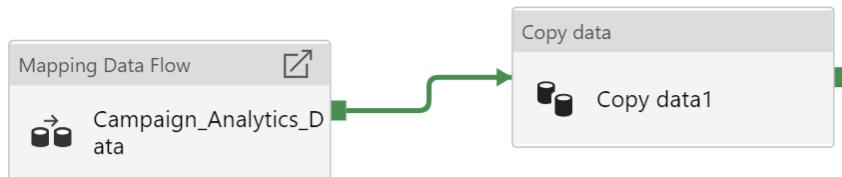
Data flow

▶ Azure Data Explorer

▶ Azure Function

▶ Batch Service

31. Copy Data 를 마찬가지로 Drag & Drop 하여 아래와 같이 이어줍니다.



32. Copy Data 를 선택하고, Source 탭으로 이동하여, **asamcw_campaignanalytics_parquet** 를 Source Data Set 으로 선택합니다. Wildcard file Path 를 선택 후 campaign-analytics 를 입력합니다.

The screenshot shows the 'Source' tab configuration for the 'Copy data' task. The 'Source dataset' dropdown is set to 'asamcw_campaignanalytics_parquet'. The 'File path type' section has 'Wildcard file path' selected. The 'Wildcard paths' field contains 'wwi-02 / campaign-analytics / *.parquet'. Other settings include 'Start time (UTC)' and 'End time (UTC)', both empty; 'Filter by last modified', which is empty; 'Recursively' checked; 'Enable partition discovery' unchecked; and 'Max concurrent connections' empty.

33. Sink 탭으로 이동해서 **asamcw_campaignanalytics_asa** 를 선택하고 Bulk Insert 를 선택합니다.

The screenshot shows the 'Sink' tab configuration for the 'Copy data' task. The 'Sink dataset' dropdown is set to 'asamcw_campaignanalytics_asa'. The 'Copy method' section has 'Bulk insert' selected. The 'Table option' section has 'None' selected. The 'Pre-copy script' section is empty.

34. Publish All 하여 저장합니다.

35. Trigger Now 하여 실행하고 Monitoring 탭에서 확인합니다.
36. 작업이 완료되면 12_02_Count_Campaign_Analytics_Table 스크립트를 실행하여 결과를 확인합니다.

작업 11: Exploring Raw Data

1. 왼쪽 메뉴의 Data 탭을 선택하고 Storage 안에 **sale-small/Year=2010/Quarter=Q4/Month=12/Day=20101231**로 이동하여 파일을 우클릭 후 Select Top 100 rows SQL 스크립트를 실행합니다.

The screenshot shows the Azure Data Explorer interface. On the left, there's a navigation pane with tabs for 'Workspace' and 'Linked'. Under 'Storage accounts', several accounts are listed, including 'kyowonsynapse (Primary - kyowo...)' and 'wwi-02' (which is highlighted with a red box). Under 'Cosmos DB', there's one entry. Under 'Datasets', there are seven entries. In the main area, a folder path is shown: 'wwi-02 > sale-small > Year=2010 > Quarter=Q4 > Month=12 > Day=20101231'. A context menu is open over a file named 'sale-small-20101231'. The menu items include 'New SQL script', 'New notebook', 'Upload', 'Download', 'New folder', 'Select all', 'Select TOP 100 rows' (which is highlighted with a red box), 'Create external table', 'New SQL script and open in new tab', 'Copy ABFSS path', 'Manage access...', 'Rename...', 'Download', 'Delete', and 'Properties...'.

2. SQL on-demand 에 연결된 것을 확인하고 실행합니다.

```

Run Undo Publish Query plan Connect to SQL on-demand Use database master
1 SELECT
2     TOP 100 *
3     FROM
4     OPENROWSET(
5         BULK 'https://kyowonadlsgen2.dfs.core.windows.net/wwi-02/sale-small/Year=2010/Quarter=Q4/Month=12/Day=20101231/sale-small-20101231-snappy.parquet',
6         FORMAT='PARQUET'
7     ) AS [r];
8

```

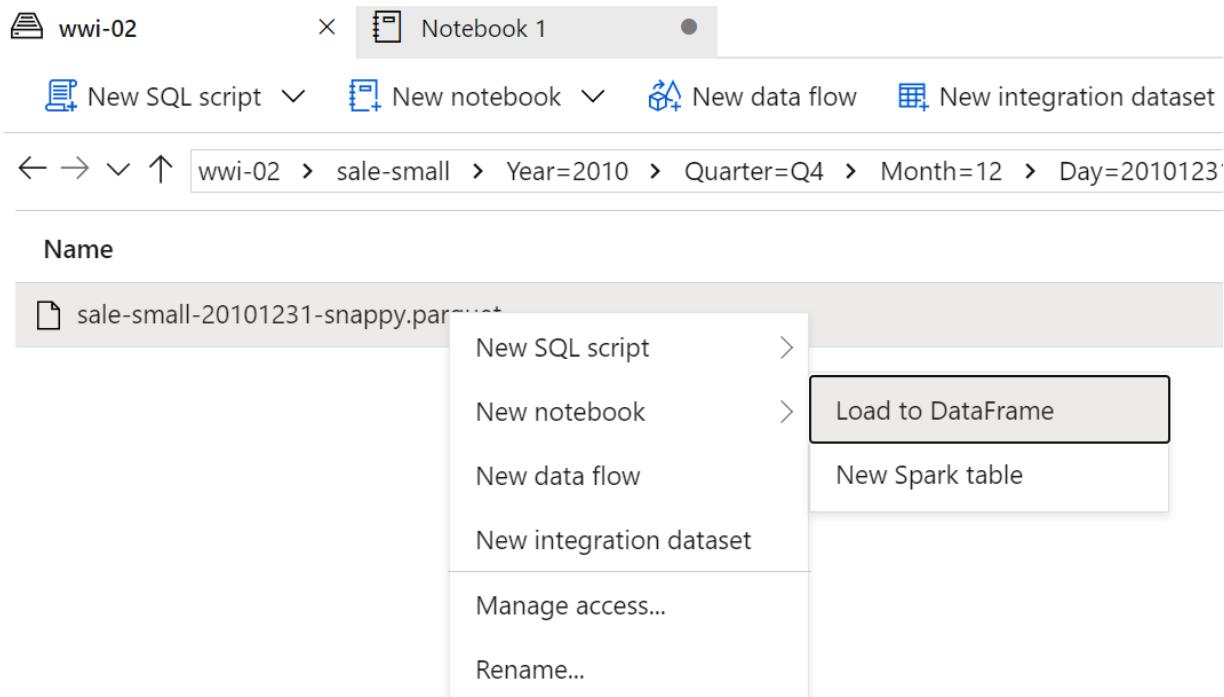
3. 아래와 같이 SQL 을 수정하여 Grouping 과 Aggregation 을 수행합니다.

```

SELECT
    TransactionDate, ProductId,
    CAST(SUM(ProfitAmount) AS decimal(18,2)) AS [(sum) Profit],
    CAST(AVG(ProfitAmount) AS decimal(18,2)) AS [(avg) Profit],
    SUM(Quantity) AS [(sum) Quantity]
FROM
    OPENROWSET(
        BULK 'https://<ADLS Gen2 Account Name>.dfs.core.windows.net/wwi-02/sale-
small/Year=2010/Quarter=Q4/Month=12/Day=20101231/sale-small-20101231-snappy.parquet',
        FORMAT='PARQUET'
    ) AS [r] GROUP BY r.TransactionDate, r.ProductId;

```

4. 다음은 같은파일에서 우클릭하여 새로운 노트북을 실행합니다.



5. Run All 을 수행합니다.

6. Add Code 를 클릭하여 다음 Cell 에 아래와 같이 입력하여 Schema 를 확인합니다.



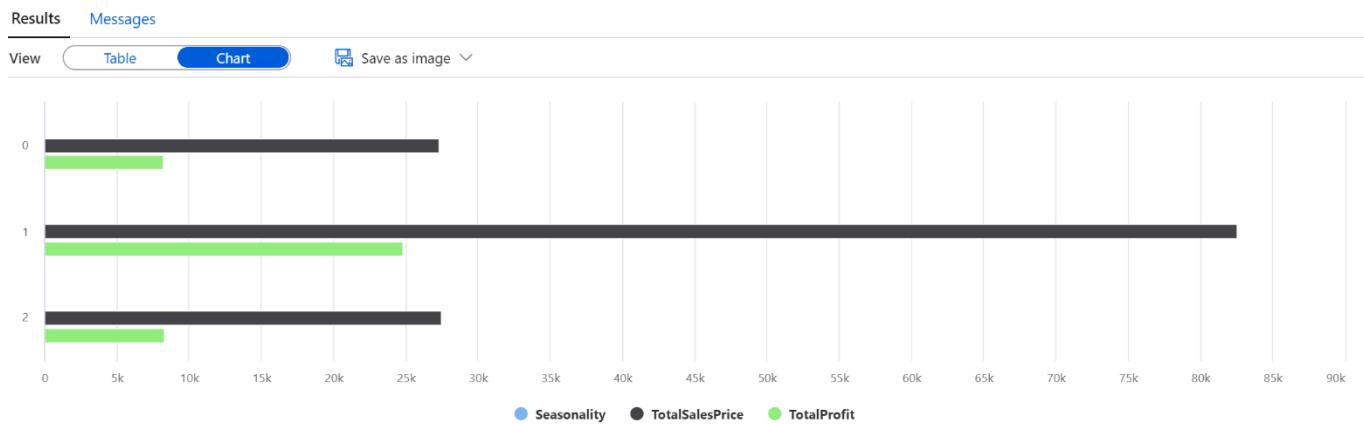
```
data_path.printSchema()
```

7. 다시 Add Code 하여 아래와 같이 입력하여 조금 전 SQL로 수행한 Script 와 결과를 비교해봅니다

```
from pyspark.sql import SparkSession
from pyspark.sql.types import *
from pyspark.sql.functions import *

profitByDateProduct = (data_path.groupBy("TransactionDate", "ProductId")
    .agg(
        sum("ProfitAmount").alias("(sum)ProfitAmount"),
        round(avg("Quantity"), 4).alias("(avg)Quantity"),
        sum("Quantity").alias("(sum)Quantity"))
    .orderBy("TransactionDate"))
profitByDateProduct.show(100)
```

8. Develop 탭의 13_01_Select_From_CSV ~ 13_03_Select_From_JSON SQL 스크립트를 열어 실행합니다.
9. 아래와 같이 Chart로 변경하여 다양하게 시각화를 해봅니다.



작업 12 : Azure Synapse Analytics Security

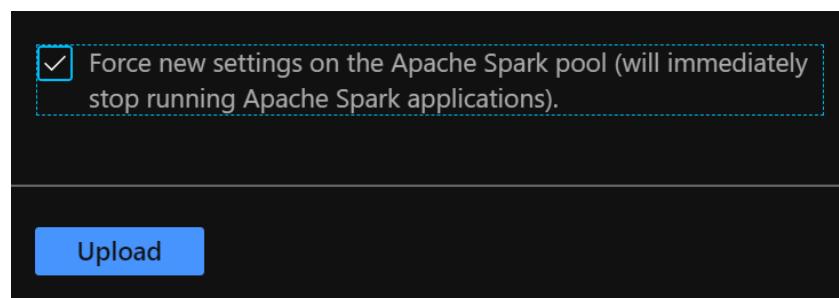
1. Develop 탭에서 14_00_User_Creation SQL 스크립트를 열어 가이드에 따라 Step by Step 단계를 진행합니다.
2. Develop 탭에서 14_01_Column_Level_Security SQL 스크립트를 열어 가이드에 따라 Step by Step 단계를 진행합니다.
3. Develop 탭에서 14_02_Row_Level_Security SQL 스크립트를 열어 가이드에 따라 Step by Step 단계를 진행합니다
4. Develop 탭에서 14_03_Dynamic_Data_Masking SQL 스크립트를 열어 가이드에 따라 Step by Step 단계를 진행합니다

작업 13 : Spark Pool Library Install

5. [Azure Portal](#)로 돌아가서 Synapse Workspace의 sparkpool을 선택합니다.
6. Sparkpool의 왼쪽 Package 탭에서 Upload Environment Config File을 선택합니다.

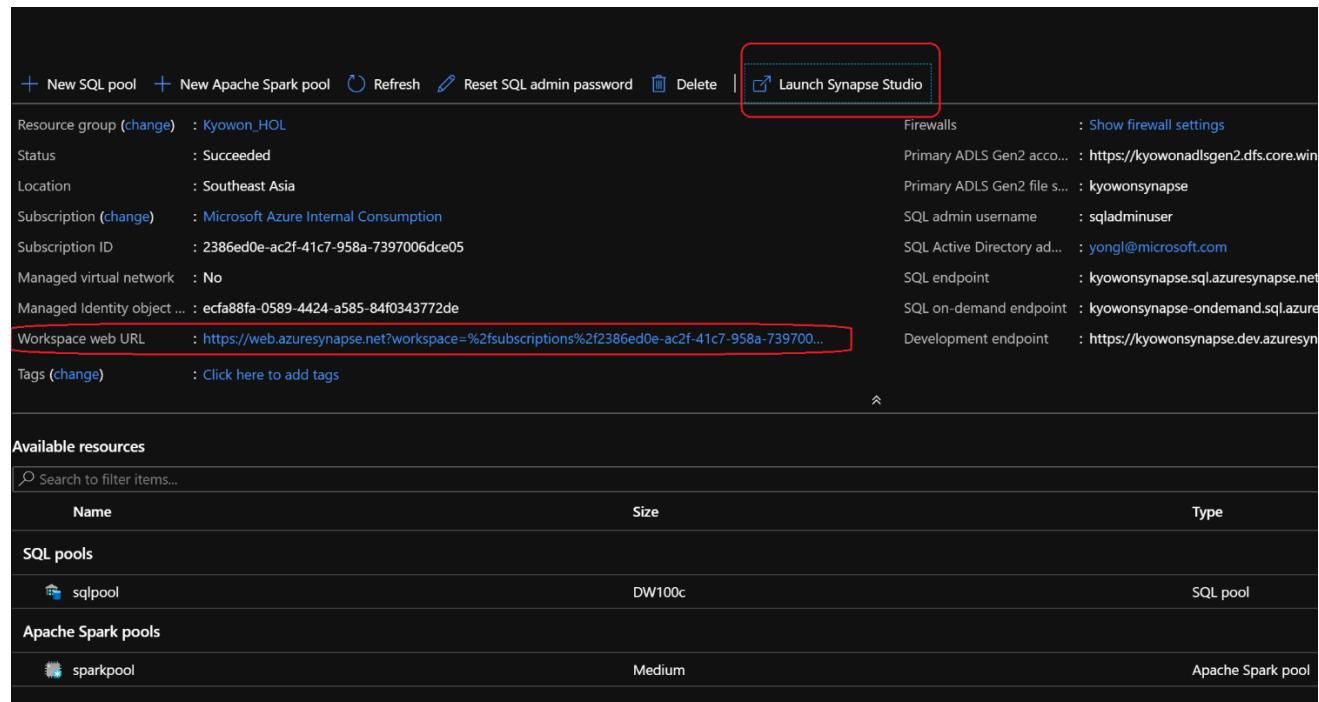
The screenshot shows the 'sparkpool (kyowonsynapse/sparkpool) | Packages' page in the Azure portal. On the left, there's a sidebar with 'Overview', 'Activity log', 'Access control (IAM)', and 'Tags'. Below that is a 'Settings' section with 'Packages' selected, 'Properties', and 'Locks'. The main area has a search bar and a 'Upload environment config file' button. A message says 'User-provisioned packages' and 'No user-provided packages currently uploaded. You can upload "enviro...".'

7. 깃허브에서 다운로드 받은
Synapse_HOL\Synapse_HOL\LabFiles\Scripts\Notebooks\Env_Config 안에 있는
Environment_Config.txt 파일을 선택하고, 아래 체크박스에 체크하고 업로드합니다.



작업 14: Azure Synapse Studio 의 Synapse Link 생성

1. Azure Synapse Workspace Overview Page 에서 아래와 같이 Link 를 클릭하여 Synapse Studio 를 실행합니다.



The screenshot shows the Azure Synapse Workspace Overview Page. At the top right, there is a button labeled "Launch Synapse Studio" with a red box drawn around it. Below the header, there are several configuration settings and resource links. One link, "Workspace web URL : https://web.azure-synapse.net?workspace=%2fsubscriptions%2f2386ed0e-ac2f-41c7-958a-7397006dce05", is also highlighted with a red box. The page also displays sections for "Available resources" and lists "SQL pools" and "Apache Spark pools".

2. Data 탭을 선택하고 + 를 클릭하여 External Data Source 를 연결합니다.

The screenshot shows the Microsoft Azure Synapse Analytics Data hub interface. The top navigation bar includes 'Microsoft Azure | Synapse Analytics > kyowonsynapse'. Below the navigation is a toolbar with 'Publish all', 'Validate all', 'Refresh', and 'Discard all' buttons. On the left, a vertical sidebar lists icons for Home, Workspace, Storage accounts, and others. The main area is titled 'Data' with tabs for 'Workspace' and 'Linked'. A search bar 'Filter resources by name' is present. A context menu is open at the bottom right of the 'Dataset' button, with 'Connect to external data' highlighted.

3. Azure Cosmos DB(SQL API)를 선택합니다.

Connect to external data

Connect to supported data sources to explore them in the Data hub and make them available to pipelines as linked services.



4. 이름에 CosmosDemo로 입력하고 생성한 DB를 선택하여 연결합니다.

New linked service (Azure Cosmos DB (SQL API))

i Choose a name for your linked service. This name cannot be used by another integration runtime.

Name *

Description

Connect via integration runtime *

[Connection string](#) [Azure Key Vault](#)

Account selection method

From Azure subscription Enter manually

Azure subscription

Cosmos DB account name *

Database name *

Additional connection properties

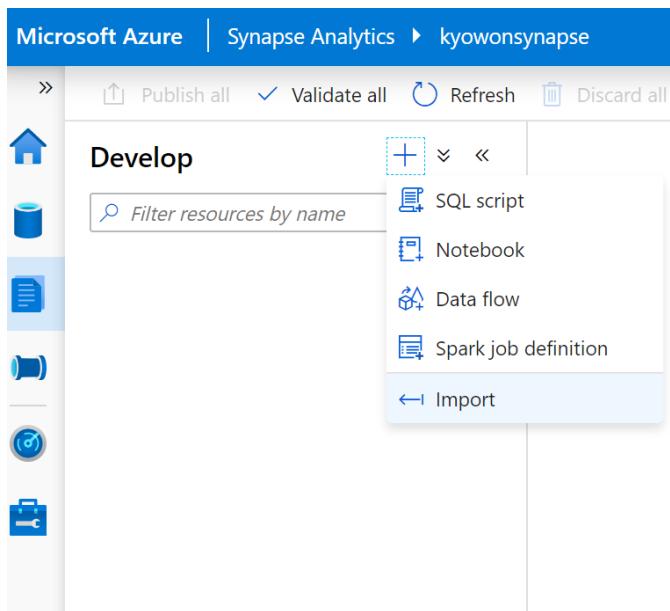
+ New

5. Refresh 하여 생성된 Link 를 확인합니다.

The screenshot shows the 'Data' view in the Azure Synapse Workspace. At the top, there are buttons for 'Publish all', 'Validate all', 'Refresh', and 'Discard all'. Below this, the 'Data' section is divided into 'Workspace' and 'Linked' tabs, with 'Linked' being the active tab. A search bar labeled 'Filter resources by name' is present. Under the 'Linked' tab, there are two main categories: 'Storage accounts' (1 item) and 'Cosmos DB' (1 item). The 'Cosmos DB' category is expanded, showing a database named 'CosmosDemo (CosmosDemo)' with six collections: 'RetailSales', 'IoTSignals', 'IoTDeviceInfo', 'RetailProducts', 'IoTStreamingSignals', and 'RetailStoreDemographics'. To the right of the database name, there are three action buttons: a blue dashed box around the ellipsis ('...'), a grey 'Actions' button, and a yellow 'Copy' button.

작업 15: Azure Synapse Workspace Spark Scenario

10. 왼쪽 메뉴의 Develop 탭을 선택하고 +를 선택하여 Import를 실행합니다.



11. 다운로드 받은 Synapse_HOL\Synapse_HOL\LabFiles\Scripts\Notebooks 안의 노트북 Script들을 선택하여 불러옵니다.

12. Publish All 을 클릭하여 Publish 합니다.

13. 00_CheckLibraries 노트북을 열고 실행하여 설치된 Library들을 확인합니다. (초기 클러스터가 올라올 때 시간이 소요됩니다)

```
+ Cell ▾ ▶ Run all ⏪ Undo | ▾ ▲ Publish | Attach to sparkpool | ⏪ Language PySpark (Python) ▾
Cell 1
1 import pip #needed to use the pip functions
2 for i in pip.get_installed_distributions(local_only=True):print(i)
^
Command executed in 7mins 19s 191ms by yongl on 06-23-2020 09:52:44.072 +09:00
File Cell Kernel Help
zipp 3.1.0
zict 1.0.0
xlwt 1.2.0
XlsxWriter 0.9.6
xlrd 1.0.0
wrapt 1.11.2
widgetsnbextension 3.5.1
wheel 0.30.0
Werkzeug 0.16.1
websocket-client 0.57.0
webencodings 0.5.1
wcwidth 0.2.4
```

14. [Azure Portal](#)에서 오늘 생성한 Resource Group에서 Cosmos DB를 찾아 선택합니다.

Showing 1 to 5 of 5 records. Show hidden types ⓘ

Name ↑↓	Type ↑↓
kyowonadlsgen2	Storage account
kyowoncosmos	Azure Cosmos DB account
kyowonsynapse	Synapse workspace
sparkpool (kyowonsynapse/sparkpool)	Apache Spark pool
sqlpool (kyowonsynapse/sqlpool)	SQL pool

15. Data Explorer 탭을 선택하여 CosmosDemo DB를 열고 각 Collection들의 Item 을 선택하여 Data 를 확인 할 수 있도록 합니다.

kyowoncosmos | Data Explorer

Azure Cosmos DB account

Search (Ctrl+ /) New Container Enable Notebooks (Preview)

Overview Activity log Access control (IAM) Tags Diagnose and solve problems Quick start Notifications Data Explorer

Settings Features Replicate data globally Default consistency

SQL API

CosmosDemo

- Scale
- IoTDeviceInfo
- IoTSignals
- Items
- Settings
- Stored Procedures
- User Defined Functions
- Triggers
- IoTStreamingSignals
- RetailProducts
- RetailSales
- RetailStoreDemographics

16. 이제 Develop 탭에서 01_CosmosDB_Configuration 노트북을 열고 <ADLS Gen2 Account Name> 부분을 수정해준 후 Run All 하지 말고 8 번 Cell 까지 순차적으로 실행합니다.
17. 다시 Cosmos DB 의 Data Explorer로 돌아와서 Data Ingestion 을 확인합니다.
18. 9 번 Cell 과 10 번 Cell 을 수행하여 Streaming Insert 를 준비한 후 다시 Cosmos DB 의 IoTStreamingSignal 컨테이너의 비어있는 Data 를 확인합니다.
19. 11 번 Cell 을 수행하고 바로 Cosmos DB 의 Data 를 확인하여 초당 10 개씩 Insert 가 되는지 확인해 봅니다.
20. 확인이 끝나면 Cancel 하여 Streaming Ingestion 을 중지 합니다.

```

1   {
2     "dateTime": "2020-04-28 17:40:54",
3     "measureValue": "4322",
4     "unit": "Revolutions per Minute",
5     "unitsymbol1": "RPM",
6     "id": "dca4c4d9-d503-4298-a1ce-850469b525d3",
7     "deviceId": "dev-2",
8     "measureType": "Rotation Speed",
9     "_rid": "ePYRAJEonGcFAAAAAAAA==",
10    "_self": " dbs/ePYRAAE==/colls/ePYRAJEonGc=/docs/ePYRAJEonGcFAAAAAAAA==/",
11    "_etag": "\"\\\"1000f824-0000-2500-0000-5ef146c60000\\\""",
12    "_attachments": "attachments/",
13    "_ts": 1592870598
14

```

21. Synapse Studio 의 Develop 탭의 02_IoT_Data_Exploration 스크립트를 열고 Cell 4 까지 실행합니다.

1. Create Spark tables pointing to the Azure Cosmos DB Analytical Store collections using Azure Synapse Link

Cell 2

```
[ ] 1 %%sql
2 create database CosmosDemoIoT
```

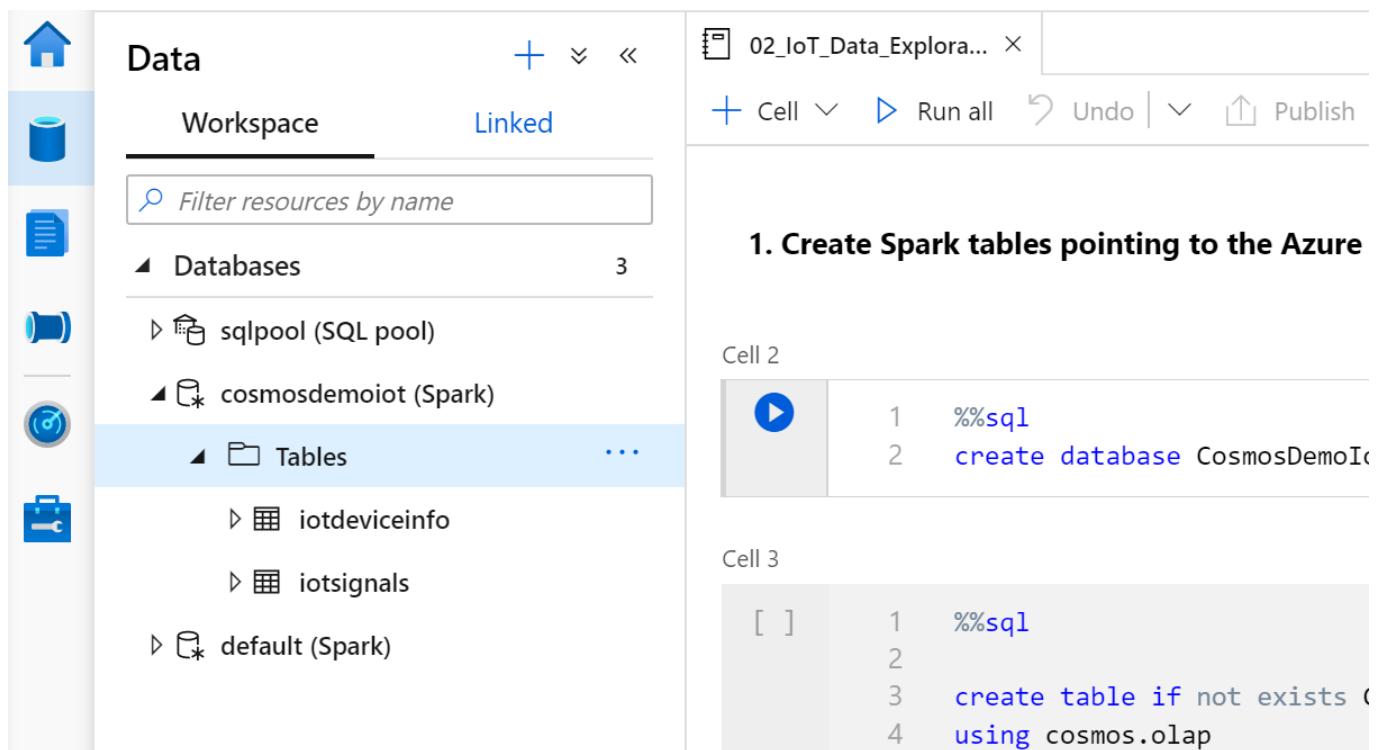
Cell 3

```
[ ] 1 %%sql
2
3 create table if not exists CosmosDemoIoT.IoTSignals
4 using cosmos.olap
5 options(spark.synapse.linkedService 'CosmosDemo',
6 ||     spark.cosmos.container 'IoTSignals')
```

Cell 4

```
 1 %%sql
2
3 create table if not exists CosmosDemoIoT.IoTDeviceInfo
4 using cosmos.olap
5 options(spark.synapse.linkedService 'CosmosDemo',
6 ||     spark.cosmos.container 'IoTDeviceInfo')
```

22. Data 탭을 선택하고 Workspace에서 CosmosDemoloT DB가 생성되었는지, 테이블도 생성되었는지 확인합니다.



The screenshot shows the Azure Synapse Analytics workspace interface. On the left, the Data tab is selected, displaying the Workspace section. It lists databases like sqlpool (SQL pool) and cosmosdemoiot (Spark), which contains tables such as iotdeviceinfo and iotsignals. To the right, a notebook titled "02_IoT_Data_Explora..." is open, showing the history of cells run. Cell 2 contains the SQL command to create the database. Cell 3 contains the SQL command to create the IoTSignals table, which references the newly created database.

```
Cell 2
1 %%sql
2 create database CosmosDemoIoT

Cell 3
[ ] 1 %%sql
2
3 create table if not exists CosmosDemoIoT.IoTSignals
4 using cosmos.olap
```

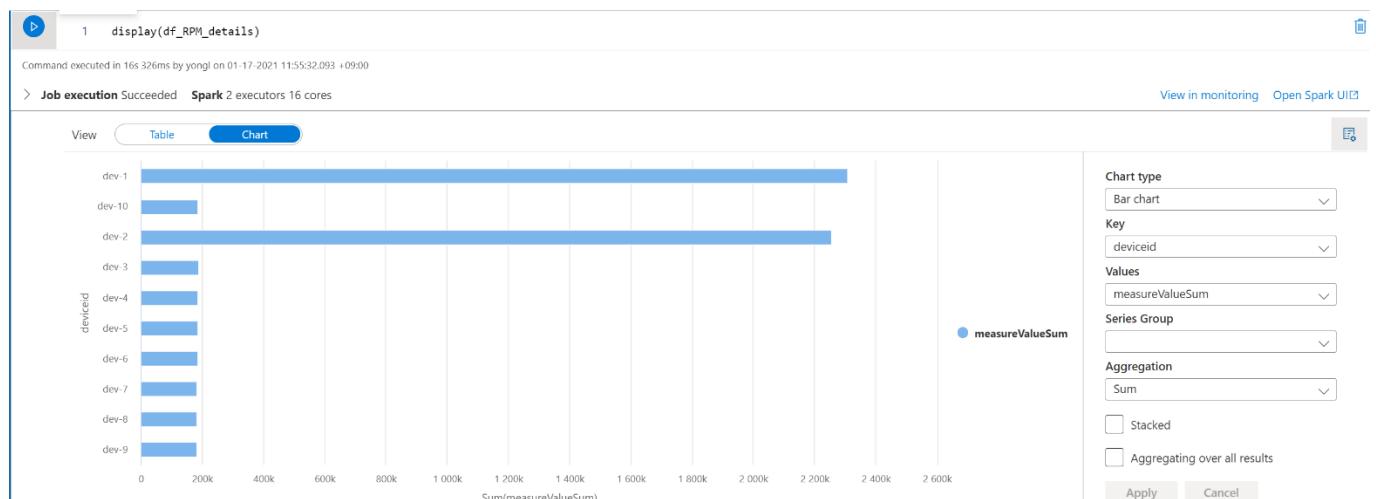
23. Cell 8 번까지 수행 후 결과를 확인합니다. 결과 창에서 Chart 를 선택 후 옵션에서 Value 항목에 latitude 대신에 measurevaluesum 을 선택합니다.

```
Cell 8
1 display(df_RPM_details)

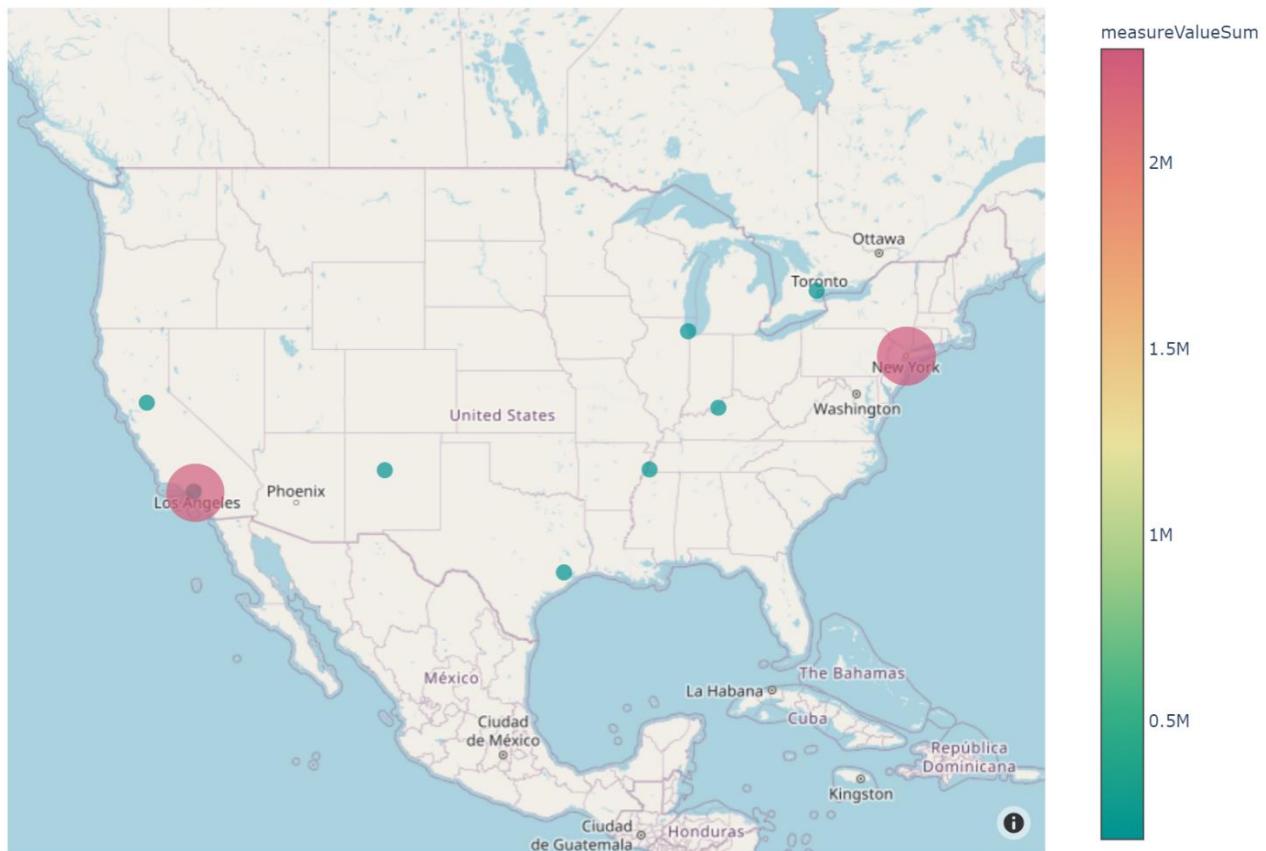
Command executed in 1mins 39s 565ms by yongl on 06-23-2020 10:28:23.613 +0900
> Job execution Succeeded Spark 2 executors 8 cores
View in monitoring Open Spark UI
```

deviceid	devicetype	location	latitude	longitude	measuretype	unitSymbol
dev-10	impulse turbine	Santa Monica, CA	34.011227	-118.499435	Rotation Speed	RPM
dev-3	impulse turbine	Chicago, IL	41.87864	-87.64025	Rotation Speed	RPM
dev-4	impulse turbine	Houston, TX	29.73222	-95.38752	Rotation Speed	RPM
dev-2	impulse-reaction turbine	Los Angeles, CA	33.944237	-118.3976	Rotation Speed	RPM
dev-5	impulse-reaction turbine	Toronto, Ontario	43.731525	-79.60837	Rotation Speed	RPM
dev-7	impulse-reaction turbine	Sacramento, CA	38.47561	-121.43124	Rotation Speed	RPM
dev-8	impulse turbine	Albuquerque, NM	35.107105	-106.57277	Rotation Speed	RPM
dev-1	impulse turbine	Manhattan, NY	40.7142	-74.0064	Rotation Speed	RPM

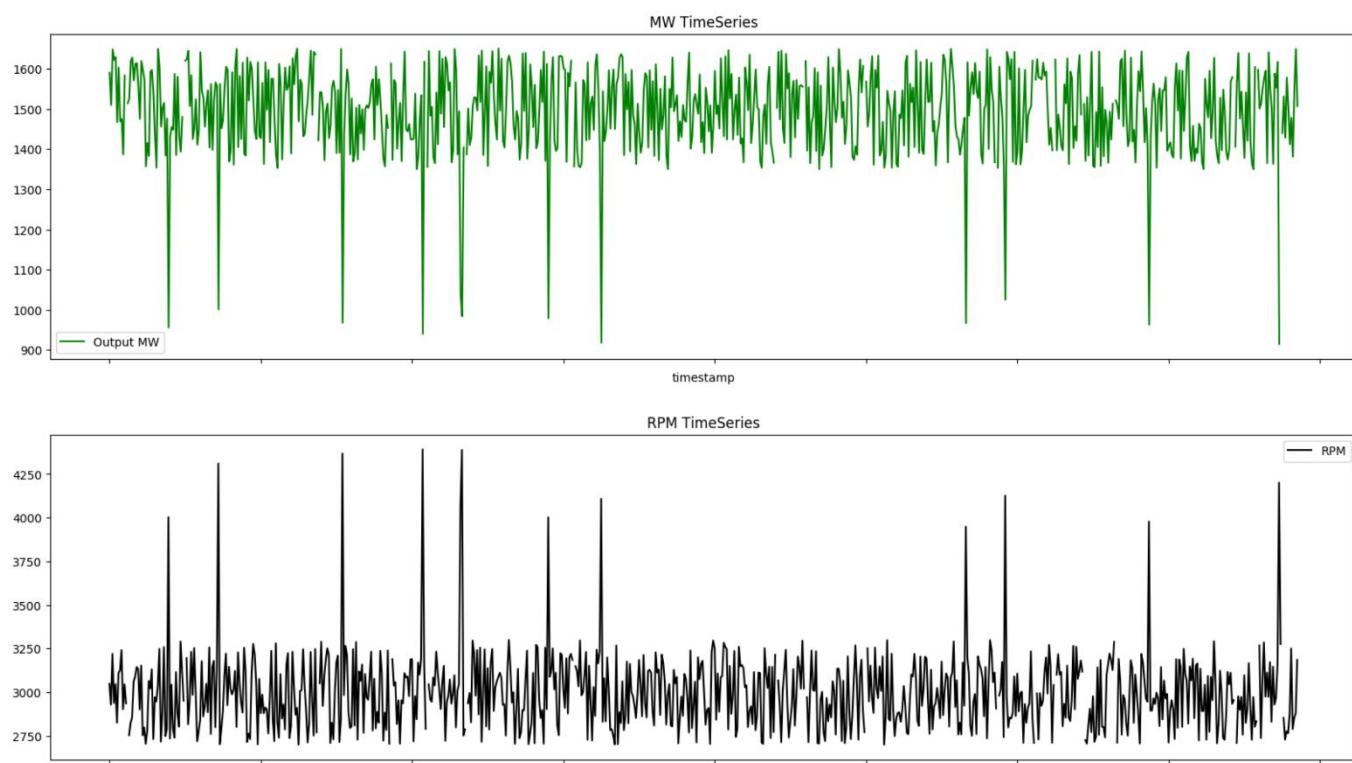
24. 아래와 같이 총 10 개의 IoT Device 중에 1 번과 2 번 Device 에 측정값 합(MeasureValueSum)이 상당히 높은것을 확인할 수 있습니다.



25. Cell 10 번을 수행하여 각 Device 별 Geographic Insight 를 확보합니다.



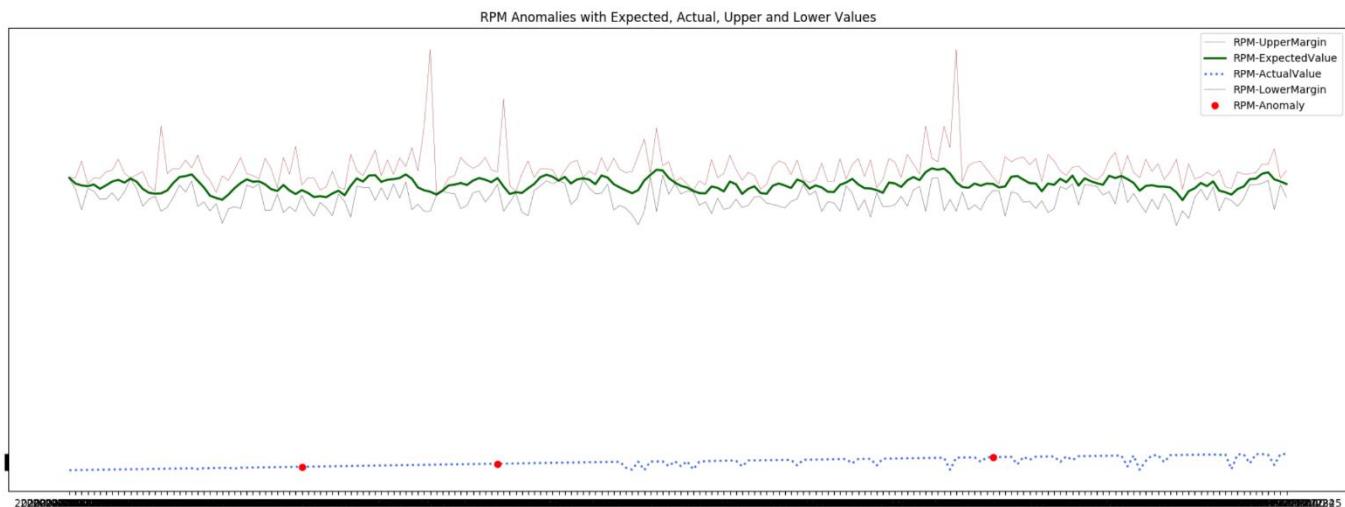
26.03_IoT_Data_AnomallyDetection 를 열고 Cell 4 까지 수행하여, 이상값 Device 중 하나인 Dev-1에 대해 unitSymbol 을 기준으로 MW 와 RPM 에 Data 를 확인합니다.



27. Cell 6 번의 <Azure Anomaly Detector End Point> 부분과 <Azure Anomaly Detector Key> 부분을 적절하게 수정하고, Cell 7 번까지 수행하여 전체 Device 에 대해 unitSymbol 이 RPM 인 Data 에 대해 Anomaly Detector 로부터 받은 내용을 확인합니다.

View						Table	Chart
measureType	.etag	value	grouping	SimpleDetectAnomalies_3e92d...	anomalies		
Rotation Speed	"bd00618e-0000-2600-0000-600...	2964.0	dev-9	null	[false, false, false, 0, 2964.0, 5.409...		
Rotation Speed	"bd00bb84-0000-2600-0000-600...	3014.0	dev-9	null	[false, false, false, 0, 2989.0, 25.25...		
Rotation Speed	"bd00c384-0000-2600-0000-600...	3053.0	dev-9	null	[false, false, false, 0, 3010.333333...		
Rotation Speed	"bd00e186-0000-2600-0000-600...	3101.0	dev-9	null	[false, false, false, 0, 3033.0, 68.68...		
Rotation Speed	"bd00ab8c-0000-2600-0000-600...	2890.0	dev-9	null	[false, false, false, 0, 3004.4, 115.5...		
Rotation Speed	"bd002a86-0000-2600-0000-600...	2798.0	dev-9	null	[false, false, false, 0, 2971.2, 174.9...		
Rotation Speed	"bd002f83-0000-2600-0000-600...	3211.0	dev-9	null	[false, false, false, 0, 3010.6, 202.4...		
Rotation Speed	"bd00bb8e-0000-2600-0000-600...	2791.0	dev-9	null	[false, false, false, 0, 2958.2, 168.8...		

28. Cell 8 번을 수행하여 Data 를 확인하고, Cell 9 번을 수행하여 데이터를 Visualize 합니다.



29. 다음으로 04_Retail_Data_AutoML 노트북을 열어 Cell 4 번까지 실행합니다.

30. Data 탭으로 이동하여 Workspace 에 CosmosDemoRetail DB 와 테이블들이 생성되었는지 확인합니다.

The screenshot shows the Azure Data Explorer interface. On the left, the 'Data' sidebar is open, showing 'Workspace' selected. A search bar at the top says 'Filter resources by name'. Below it, under 'Databases', there are four entries: 'sqlpool (SQL pool)', 'cosmosdemoiot (Spark)', 'cosmosdemoretail (Spark)', and 'default (Spark)'. Under 'cosmosdemoretail (Spark)', there is a 'Tables' section with three entries: 'retailproducts', 'retailsales', and 'retailstoredemographics'. On the right, a Jupyter Notebook cell titled '04_Retail_Data_Auto...' is displayed. The cell content is as follows:

```
[ ] 1 %%sql
2 create database CosmosDemoRetail
```

Cell 3:

```
[ ] 1 %%sql
2
3 create table if not exists CosmosD
4 using cosmos.olap
5 options(spark.synapse.linkedService
6 | | spark.cosmos.container 'Re
```

31. Cell 5 번을 수행하고 Data 를 탐색합니다.
32. Cell 7 번의 필요한 정보들을 입력하고 수행하여 Azure ML Workspace 의 환경설정을 진행합니다.
33. Cell 9 번을 수행하여 Data Prep.을 진행합니다.
34. Cell 11 번을 수행하여 AutoML 을 수행하고 결과를 확인합니다. 30 분 가량 진행됩니다.

```

Running on local machine
Parent Run ID: AutoML_141283ae-566c-4259-b588-191b3292f3c4

Current status: DatasetFeaturization. Beginning to featurize the dataset.
Current status: DatasetFeaturizationCompleted. Completed featurizing the dataset.
Current status: DatasetCrossValidationSplit. Generating individually featurized CV splits.
Current status: DatasetFeaturization. Beginning to featurize the CV split.
Current status: DatasetFeaturizationCompleted. Completed featurizing the CV split.
Current status: DatasetFeaturization. Beginning to featurize the CV split.
Current status: DatasetFeaturizationCompleted. Completed featurizing the CV split.
Current status: DatasetFeaturization. Beginning to featurize the CV split.
Current status: DatasetFeaturizationCompleted. Completed featurizing the CV split.

*****
DATA GUARDRAILS:

TYPE: Frequency detection
STATUS: PASSED
DESCRIPTION: The time series was analyzed, all data points are aligned with detected frequency.

TYPE: Missing feature values imputation
STATUS: PASSED
DESCRIPTION: No feature missing values were detected in the training data.
Learn more about missing value imputation: https://aka.ms/AutomatedMLFeaturization

*****
Current status: ModelSelection. Beginning model selection.

*****
ITERATION: The iteration being evaluated.
PIPELINE: A summary description of the pipeline being evaluated.
DURATION: Time taken for the current iteration.
METRIC: The result of computing score on the fitted pipeline.
BEST: The best observed score thus far.

*****
```

ITERATION	PIPELINE	DURATION	METRIC	BEST
0	StandardScalerWrapper ElasticNet	0:00:20	0.0310	0.0310
1	StandardScalerWrapper ElasticNet	0:00:17	0.0324	0.0310
2	StandardScalerWrapper ElasticNet	0:00:19	0.0308	0.0308
3	StandardScalerWrapper ElasticNet	0:00:15	0.0310	0.0308
4	StandardScalerWrapper LassoLars	0:00:19	0.0323	0.0308
5	StandardScalerWrapper ElasticNet	0:00:21	0.0304	0.0304
6	StandardScalerWrapper RandomForest	0:01:37	0.0262	0.0262
7	StandardScalerWrapper ExtremeRandomTrees	0:00:14	0.0184	0.0184
8	StandardScalerWrapper DecisionTree	0:00:14	0.0040	0.0040
9	MinMaxScaler GradientBoosting	0:00:14	0.0215	0.0040
10	MinMaxScaler DecisionTree	0:00:14	0.0078	0.0040
11	RobustScaler LassoLars	0:00:15	0.0325	0.0040
12	MaxAbsScaler LightGBM	0:00:15	0.0121	0.0040
13	StandardScalerWrapper ElasticNet	0:00:14	0.0305	0.0040
14	MinMaxScaler RandomForest	0:00:23	0.0024	0.0024
15	StandardScalerWrapper LassoLars	0:01:41	0.0321	0.0024
16	StandardScalerWrapper ElasticNet	0:00:19	0.0325	0.0024
17	StandardScalerWrapper DecisionTree	0:00:14	0.0048	0.0024
18	MinMaxScaler DecisionTree	0:00:14	0.0157	0.0024
19	RobustScaler DecisionTree	0:00:19	0.0061	0.0024
20	StandardScalerWrapper RandomForest	0:00:31	0.0097	0.0024
21	RobustScaler RandomForest	0:00:23	0.0057	0.0024
22	MinMaxScaler LightGBM	0:00:32	0.0176	0.0024
23	MaxAbsScaler GradientBoosting	0:00:29	0.0115	0.0024
24	RobustScaler ExtremeRandomTrees	0:00:35	0.0106	0.0024
25	RobustScaler RandomForest	0:00:21	0.0160	0.0024
26	MaxAbsScaler LightGBM	0:00:20	0.0220	0.0024
27	StandardScalerWrapper ExtremeRandomTrees	0:00:20	0.0167	0.0024
28	MaxAbsScaler LightGBM	0:00:15	0.0214	0.0024
29	StandardScalerWrapper LightGBM	0:00:20	0.0102	0.0024
30	VotingEnsemble	0:03:12	0.0024	0.0024

35. Cell 13 번을 수행하여 AutoML 수행결과를 받아오고 Test Data로 Prediction 을 수행합니다.

36. Cell15 를 수행하여 결과를 Plotting 합니다.

[Test data scores]

```
root_mean_squared_error: 11.958
normalized_root_mean_squared_error: 0.007
normalized_root_mean_squared_log_error: 0.003
r2_score: 0.998
root_mean_squared_log_error: 0.015
explained_variance: 0.998
median_absolute_error: 0.426
mean_absolute_error: 3.304
mean_absolute_percentage_error: 0.862
normalized_mean_absolute_error: 0.002
spearman_correlation: 1.000
normalized_median_absolute_error: 0.000
```

