Azure Service

**ASO Hands on Lab**

Contents

[1. 사전 준비 2](#_Toc33518808)

[1.1. 시나리오 개요 2](#_Toc33518809)

[1.2. Azure 서비스 및 관련 제품 2](#_Toc33518810)

[1.3. 필수 조건 2](#_Toc33518811)

[2. 학습 목표 3](#_Toc33518812)

[2.1. 솔루션 아키텍처 3](#_Toc33518813)

[2.2. 대상 환경 정보 3](#_Toc33518814)

[3. Exercise 1: Configure Azure Services 4](#_Toc33518815)

[3.1. Task 1: Create an Azure Storage Account 4](#_Toc33518816)

[3.2. Task 2: Create an Azure SQL Data Warehouse 8](#_Toc33518817)

[3.3. Task 3: Prepare Azure SQL Data Warehouse 11](#_Toc33518818)

[3.4. Task 4: Create an Azure Data Factory v2 15](#_Toc33518819)

[4. Exercise 2: Orchestrate with Azure Data Factory 17](#_Toc33518820)

[4.1. Task 1: Amazon S3 to Azure Gen2 17](#_Toc33518821)

[4.2. Task 2: MySQL to Azure Synapse Analytics 26](#_Toc33518822)

[4.3. Task 3: Azure ML Execute 43](#_Toc33518823)

[4.4. Task 4: PreProcessing and DataExport 45](#_Toc33518824)

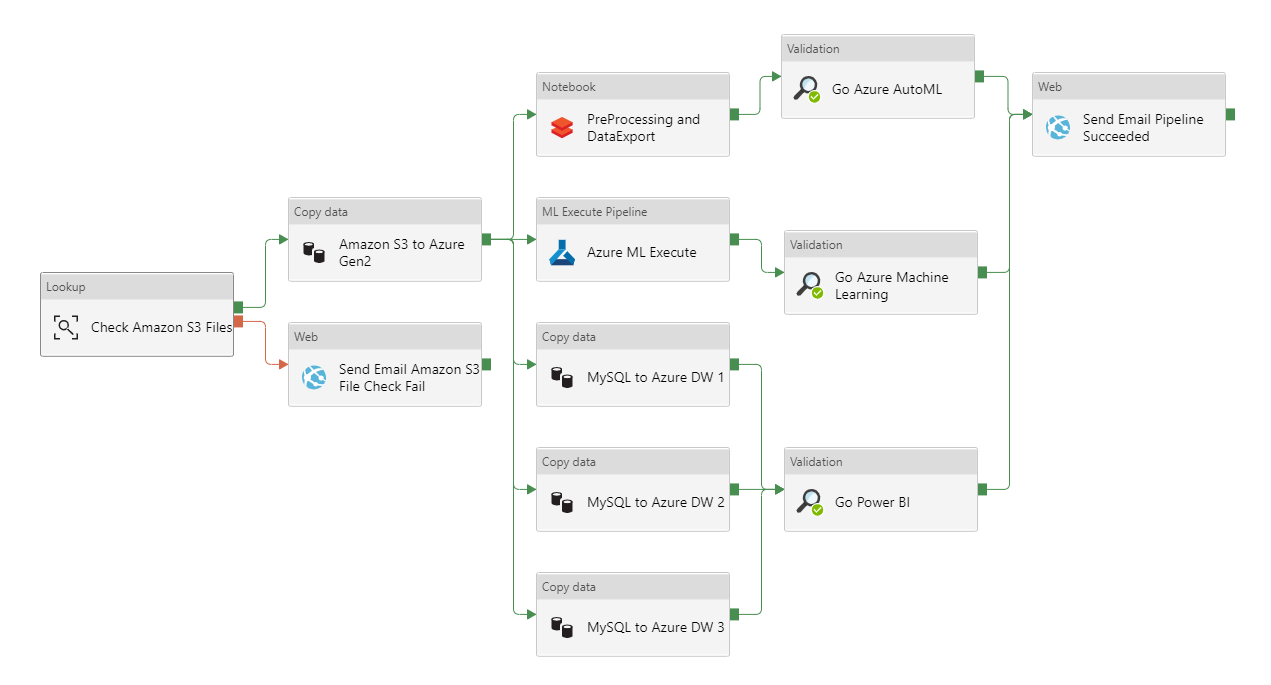
[4.5. Task 5: Send Email 47](#_Toc33518825)

[5. Exercise 3: Visualize data with Power BI Desktop 50](#_Toc33518826)

[5.1. Task 1: Install Power BI Desktop 50](#_Toc33518827)

[5.2. Task 2: Query data with Power BI Desktop 53](#_Toc33518828)

1. 사전 준비
   1. 시나리오 개요

이 실습은 Azure Data factory를 이용하여 다양한 데이터를 Data Lake Storage Gen2에 수집, DW Data는 Azure Data Warehouse에 적재하여 Power BI를 통한 시각화를 진행하고, 정형 데이터와 비정형 데이터는 Databricks를 통한 데이터 전 처리를 통해 머신 러닝에 적합한 상태로 변경하여 변경한 데이터로 Azure Machine Learning 서비스에 연결하여 처리합니다. 

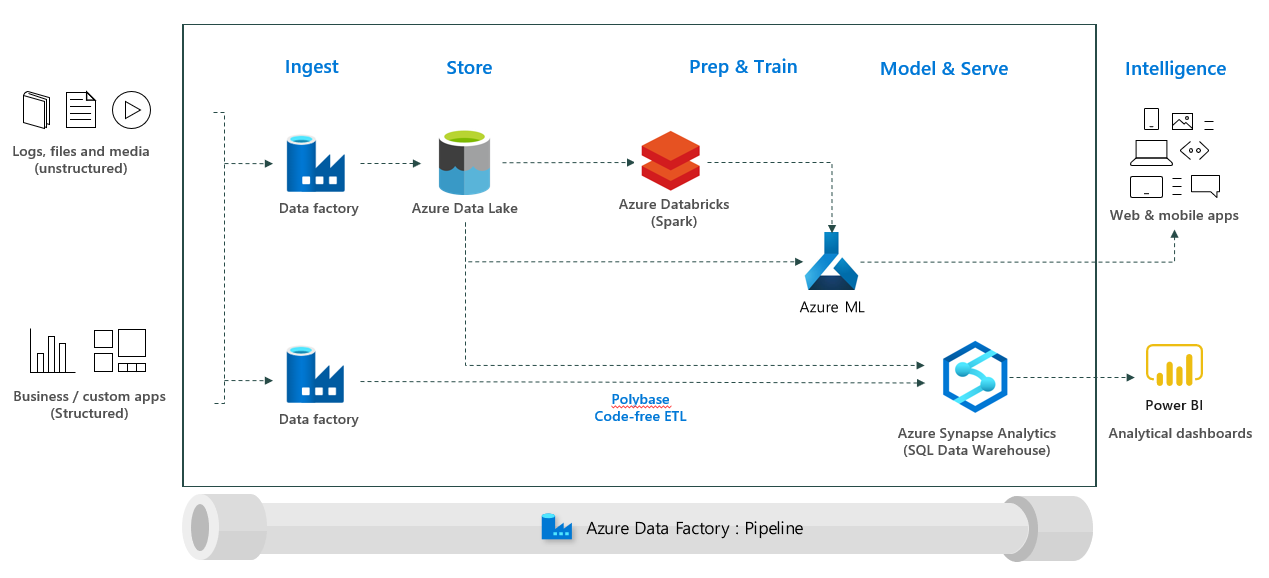
* 1. Azure 서비스 및 관련 제품
* Azure Data Factory
* Azure Storage account
* Azure SQL Data Warehouse
* Azure Databricks Service
* Machine Learning
* Power BI
  1. 필수 조건
* Azure Subscription

1. 학습 목표

이 실습을 통해 Azure Data factory를 이용하여 File 및 DB를 이관하는 방법과 다른 Azure 서버스와의 연결을 통해 Data를 처리하는 방법을 알아보아 Azure Data factory로 전체 Data의 흐름을 관리하는 방법을 연습해 봅니다.

* 1. 솔루션 아키텍처

.



* 1. 대상 환경 정보

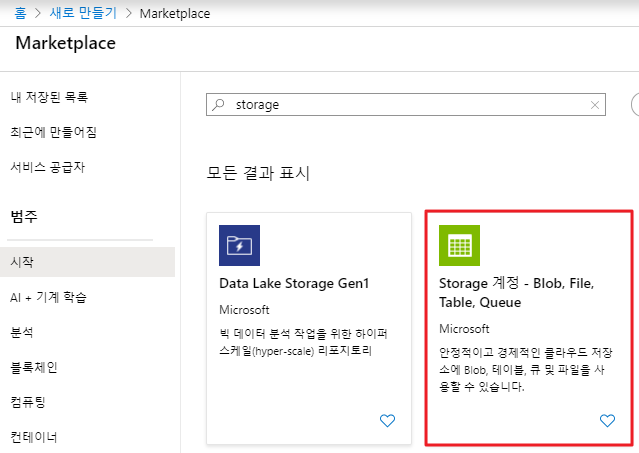
이 연습에서는 이 실습을 위한 소스 환경을 배포합니다.

1. Azure Portal (<https://portal.azure.com>)에서 이 랩에 사용할 구독으로 로그인 했는지 확인합니다.
2. Exercise 1: Configure Azure Services

이 실습에서는 Azure Storage Account, Azure SQL Data Warehouse 및 Azure Data Factory V2를 생성하고 구성합니다. 이러한 서비스를 사용하여 기존 Data Warehouse 및 File을 Azure 상으로 이관할 수 있도록 합니다.

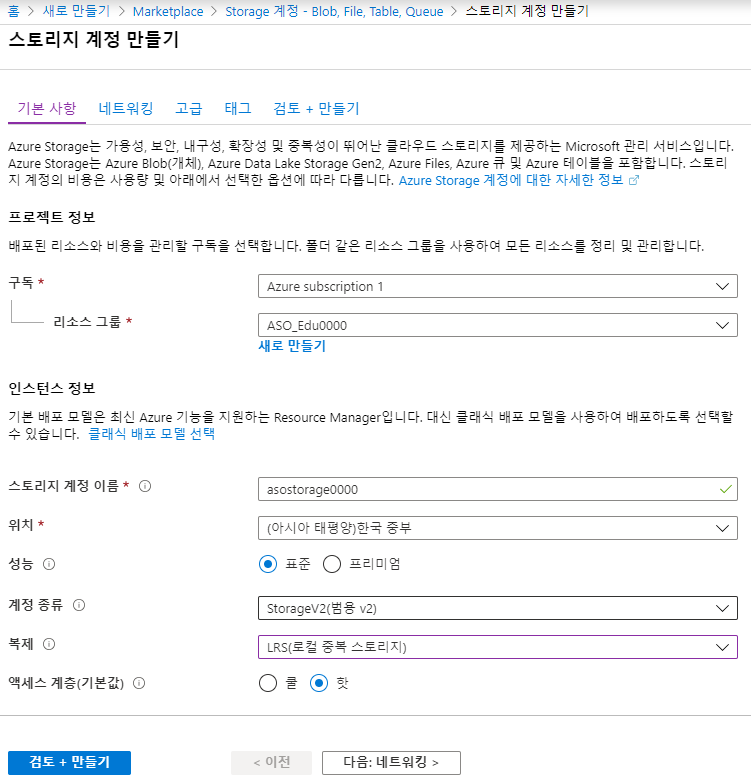
* 1. Task 1: Create an Azure Storage Account

1. Azure 포탈로 이동하여 https://portal.azure.com/에서 로그인합니다.
2. **홈 화면 상단의 리소스 만들기를** 선택하여, **storage**를 검색하고 **Storage 계정**을 선택합니다.



1. 저장소 계정 블레이드에서 **만들기**를 클릭합니다. 다음 정보를 입력합니다.

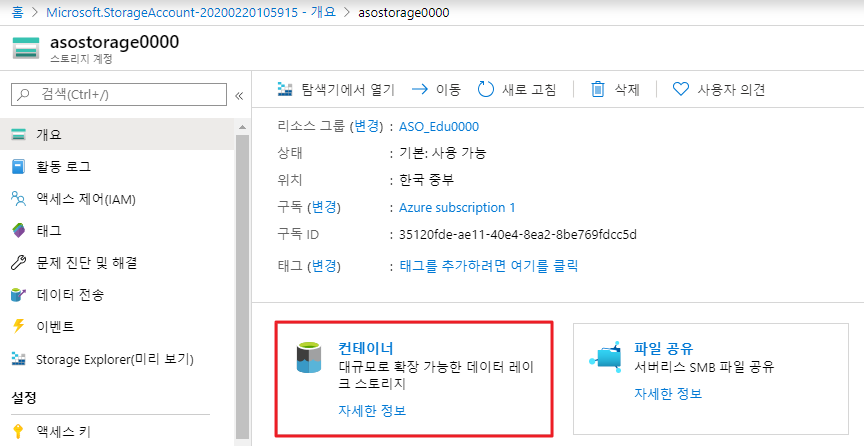
* 구독: **본인의 구독 선택**
* 리소스 그룹: **이전에 생성한 리소스 선택 ->** **ASO\_EduXXXX**
* 스토리지 계정 이름: **asostorageXXXX**
* 위치: **한국 중부(Korea Central)**.
* 성능: **표준**
* 계정 종류: **StorageV2 (범용 v2)**
* Replication: **LRS(로컬 중복 스토리지)**
* 액세스 계층: **핫**



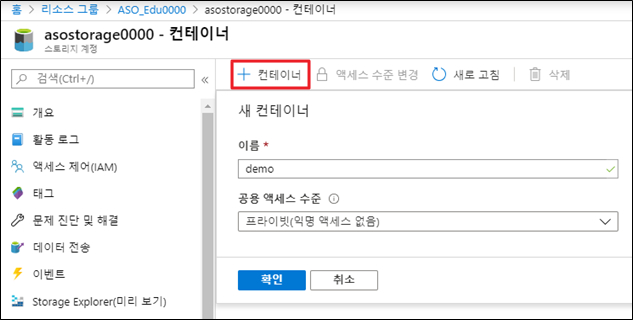
1. 상단의 **고급**을 클릭하고 **Data Lake Storage Gen2**를 **사용**으로 변경합니다.



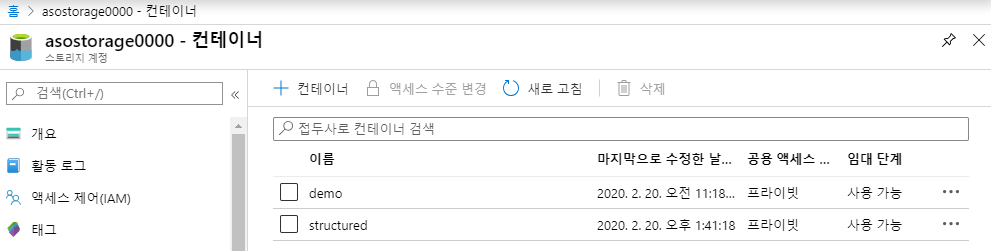
1. **검토+만들기**를 클릭하고 구성 선택 사항을 확인한 후 **만들기**를 선택합니다.
2. 새 스토리지 계정이 완성되면 리소스로 이동하여 **컨테이너**를 클릭합니다.



1. **+컨테이너**를 클릭한 후 **demo를 입력하고** **확인 버튼**을 클릭합니다.

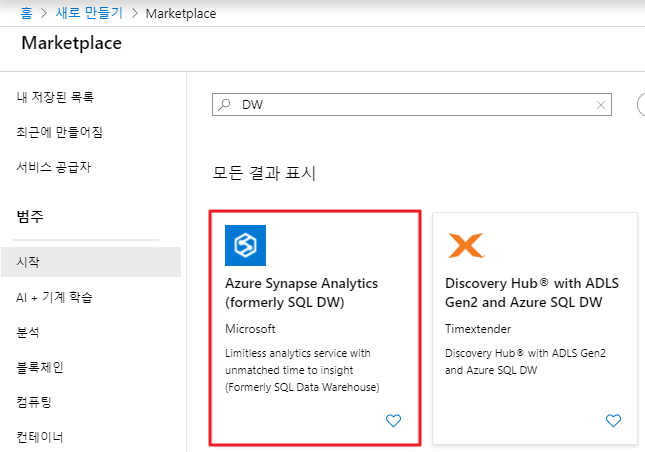


1. 위와 같은 방식으로 **+컨테이너**를 클릭 하여 **structured를 입력하고 확인 버튼**을 클릭합니다.



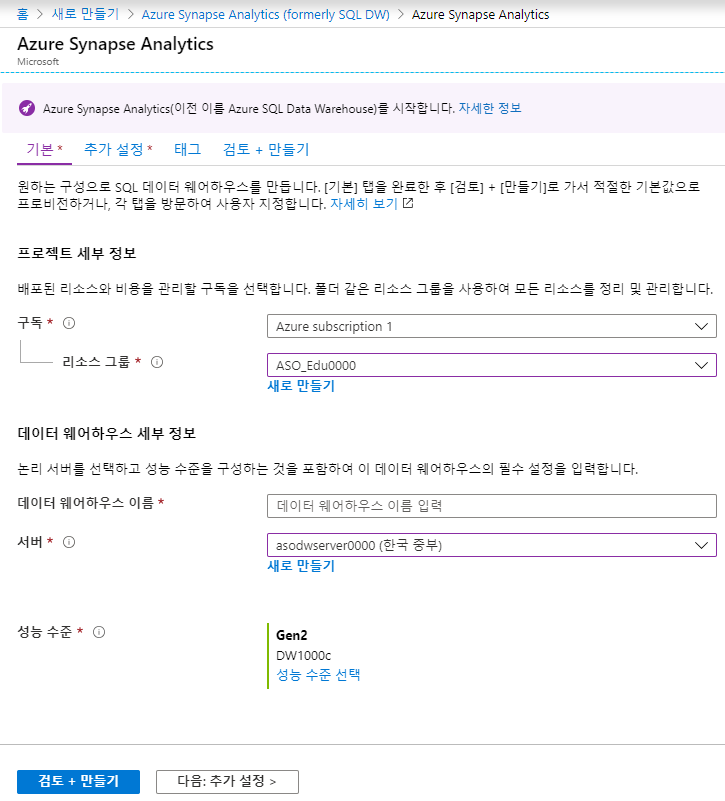
* 1. Task 2: Create an Azure SQL Data Warehouse

1. Azure 포탈로 이동하여 https://portal.azure.com/에서 로그인합니다.
2. **왼쪽 메뉴의 리소스 만들기를** 선택하여,  **DW 입력하여 검색한 후** **Azure Synapse Analytics**를 선택합니다.



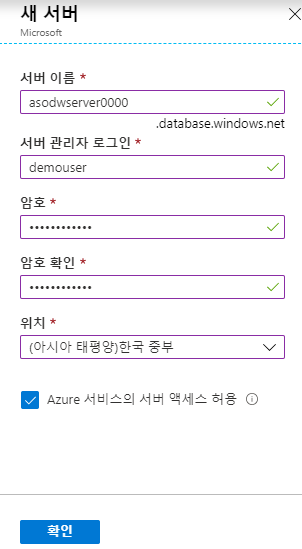
1. **Azure Synapse Analytics** 블레이드에서 **만들기**를 클릭하고 아래 내용을 입력합니다.

* 구독: **본인의 구독 선택.**
* 리소스 그룹: **이전에 생성한 리소스** 선택 -> **ASO\_EduXXXX**
* 데이터 웨어하우스 이름: **CohoDWXXXX**

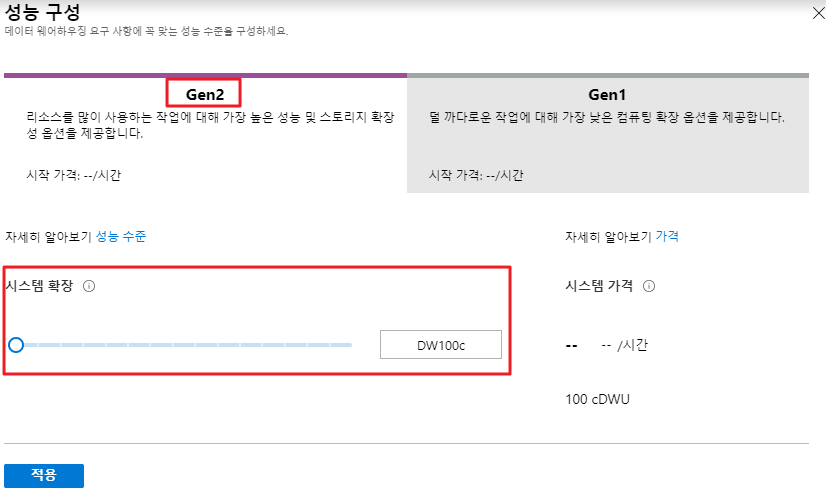


1. **서버에서 새로 만들기를 클릭하여 아래 내용을 입력 후 확인 버튼을 클릭합니다.**

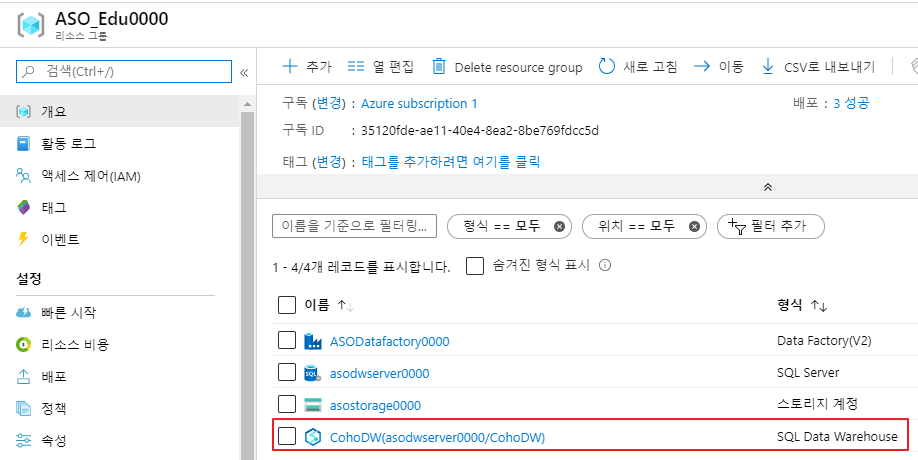
* 서버 이름: **asodwserverXXXX** (DW 계정 명)
* 서버 관리자 로그인: **demouser**
* 암호: **Demo@pass123**
* 위치: **한국 중부(Korea Central)**
* Azure 서비스의 서버 액세스 허용: **체크함**



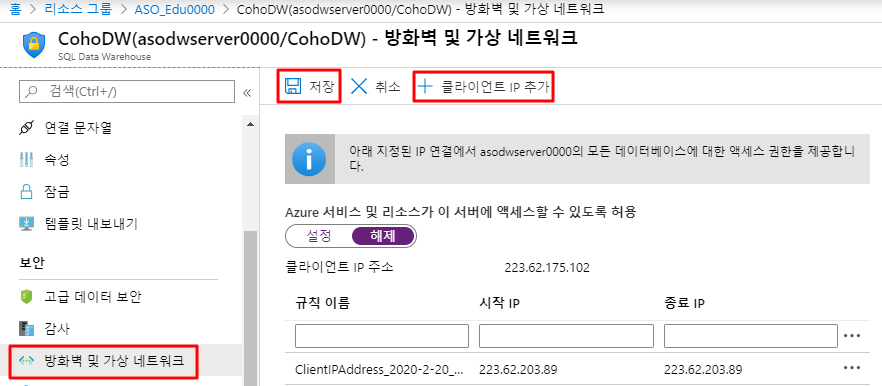
1. **성능 수준 선택**을 클릭하고 Gen2 탭을 선택한 다음 성능을 100cDWU로 설정하고 적용을 클릭합니다.



1. SQL Data Warehouse 블레이드에서 **검토+만들기**를 클릭한 후 **만들기**를 클릭합니다.
   1. Task 3: Prepare Azure SQL Data Warehouse
2. Azure 포탈로 이동하여 https://portal.azure.com/에서 로그인합니다.
3. 리소스 그룹으로 이동하여 **Data Warehouse CohoDW**를 선택합니다..

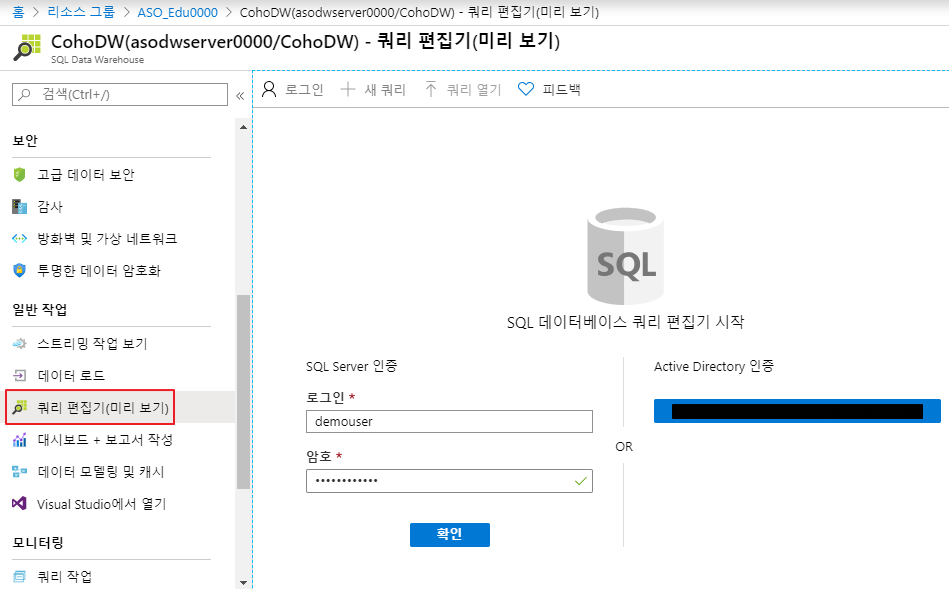


1. 왼쪽 보안 메뉴의 방화벽 및 가상 네트워크로 이동하여 **+클라이언트 IP 추가** 버튼 클릭 후 **저장** 버튼을 클릭하여 저장합니다.



1. 왼쪽 **일반 작업** 메뉴에서 **쿼리 편집기**를 클릭 후 아래의 정보를 입력하여 로그인합니다.

* 로그인: **demouser**
* 암호: **Demo@pass123**



1. 다음 쿼리를 실행하여테이블을 생성합니다.

CREATE TABLE [dbo].[DimCustomer](

       [CustomerKey] int not null, [GeographyKey] int null,

       [CustomerAlternateKey] nvarchar(30) not null, [Title] nvarchar(16) null,

       [FirstName] nvarchar(100) null, [MiddleName] nvarchar(100) null,

       [LastName] nvarchar(100) null,

       [NameStyle] bit null,

       [BirthDate] date null,

       [MaritalStatus] nchar(2) null,

       [Suffix] nvarchar(20) null,

       [Gender] nvarchar(2) null,

       [EmailAddress] nvarchar(100) null,

       [YearlyIncome] money null,

       [TotalChildren] tinyint null,

       [NumberChildrenAtHome] tinyint null,

       [EnglishEducation] nvarchar(80) null,

       [SpanishEducation] nvarchar(80) null,

       [FrenchEducation] nvarchar(80) null,

       [EnglishOccupation] nvarchar(200) null,

       [SpanishOccupation] nvarchar(200) null,

       [FrenchOccupation] nvarchar(200) null,

       [HouseOwnerFlag] nchar(2) null,

       [NumberCarsOwned] tinyint null,

       [AddressLine1] nvarchar(240) null,

       [AddressLine2] nvarchar(240) null,

       [Phone] nvarchar(40) null,

       [DateFirstPurchase] date null,

       [CommuteDistance] nvarchar(30) null

) WITH(CLUSTERED COLUMNSTORE INDEX, DISTRIBUTION = ROUND\_ROBIN);

CREATE TABLE [dbo].[DimGeography](

[GeographyKey] int not null, [City] nvarchar(60) null,

[StateProvinceCode] nvarchar(6) null,

[StateProvinceName] nvarchar(100) null,

[CountryRegionCode] nvarchar(6) null,

[EnglishCountryRegionName] nvarchar(100) null,

[SpanishCountryRegionName] nvarchar(100) null,

[FrenchCountryRegionName] nvarchar(100) null,

[PostalCode] nvarchar(30) null,

[SalesTerritoryKey] int null

) WITH(CLUSTERED COLUMNSTORE INDEX, DISTRIBUTION = ROUND\_ROBIN);

CREATE TABLE [dbo].[FactInternetSales](

       [ProductKey] int not null,

       [OrderDateKey] int not null,

       [DueDateKey] int not null,

       [ShipDateKey] int not null,

       [CustomerKey] int not null,

       [PromotionKey] int not null,

       [CurrencyKey] int not null,

       [SalesTerritoryKey] int not null,

       [SalesOrderNumber] nvarchar(40) not null,

       [SalesOrderLineNumber] tinyint not null,

       [RevisionNumber] tinyint not null,

       [OrderQuantity] smallint not null,

       [UnitPrice] money not null,

       [ExtendedAmount] money not null,

       [UnitPriceDiscountPct] float(8) not null,

       [DiscountAmount] float(8) not null,

       [ProductStandardCost] money not null,

       [TotalProductCost] money not null,

       [SalesAmount] money not null,

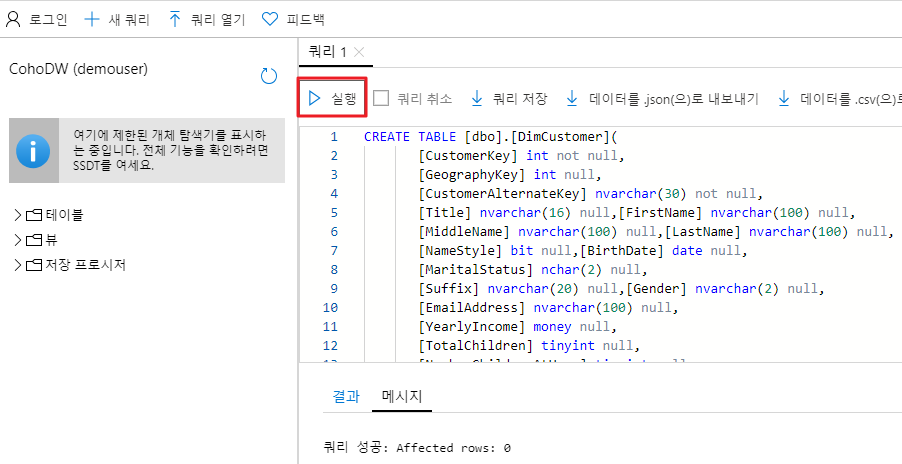
       [TaxAmt] money not null,

       [Freight] money not null,

       [CarrierTrackingNumber] nvarchar(50) null,

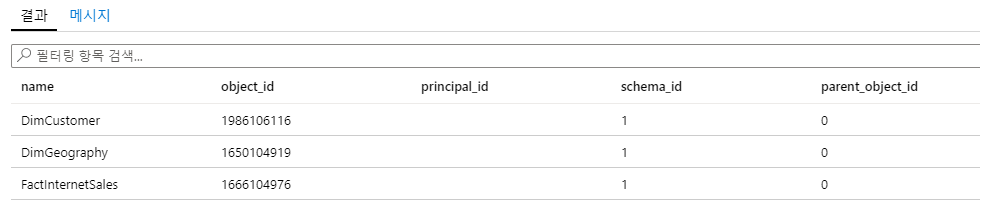
       [CustomerPONumber] nvarchar(50) null

) WITH(CLUSTERED COLUMNSTORE INDEX, DISTRIBUTION = ROUND\_ROBIN)



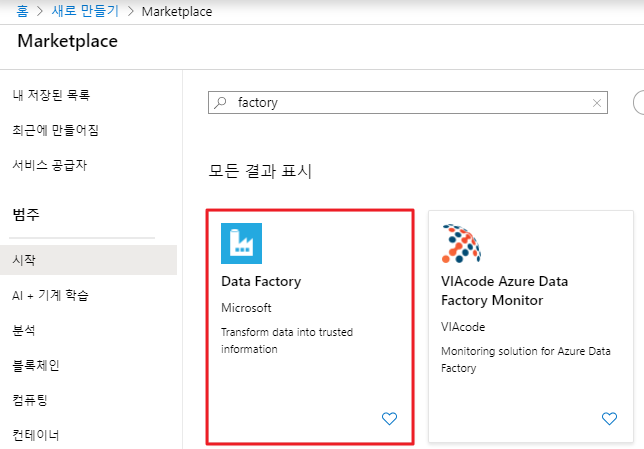
1. 다음 쿼리를 실행하여 테이블이 생성되었는지 확인합니다. 반환되는 행이 3개여야 합니다.

|  |
| --- |
| SELECT \* FROM sys.tables |



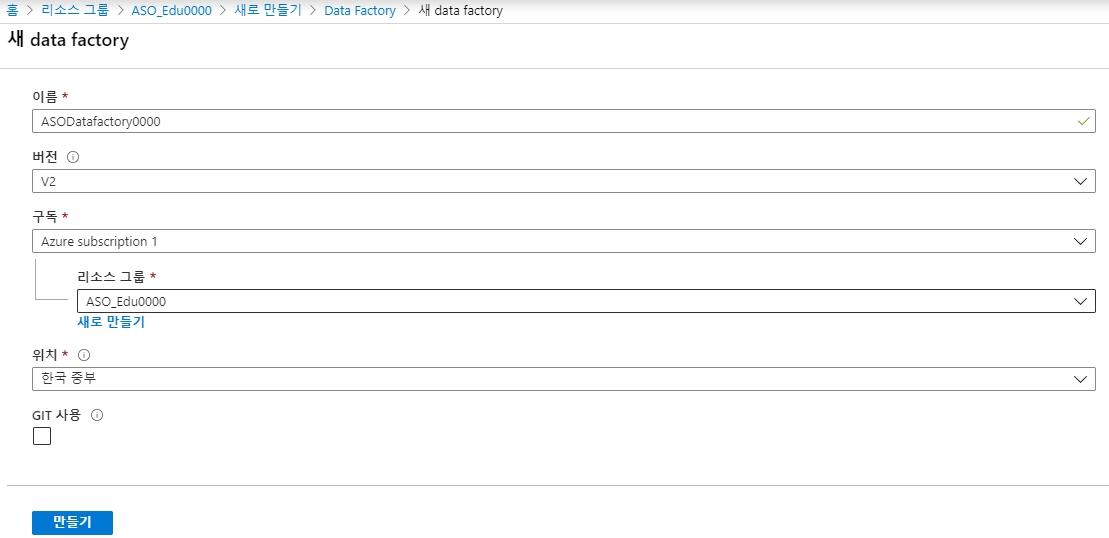
* 1. Task 4: Create an Azure Data Factory v2

1. Azure 포탈로 이동하여 https://portal.azure.com/에서 로그인합니다.
2. **리소스 만들기**를 클릭하여 **Factory**를 입력, 검색하고 **Data Factory**를 선택합니다.



1. Data Factory 블레이드에서 **만들기**를 클릭한 후 다음 정보를 입력합니다.

* 이름: **ASODatafactoryXXXX (XXXX는 사용자 별로 고유하게 입력**)
* 버전: **V2**
* 구독: **본인의 구독 선택**.
* 리소스 그룹: **ASO\_EduXXXX**
* 위치: **한국 중부(Korea Central)**
* **Git사용 미 체크**



1. **만들기** 버튼을 클릭합니다.

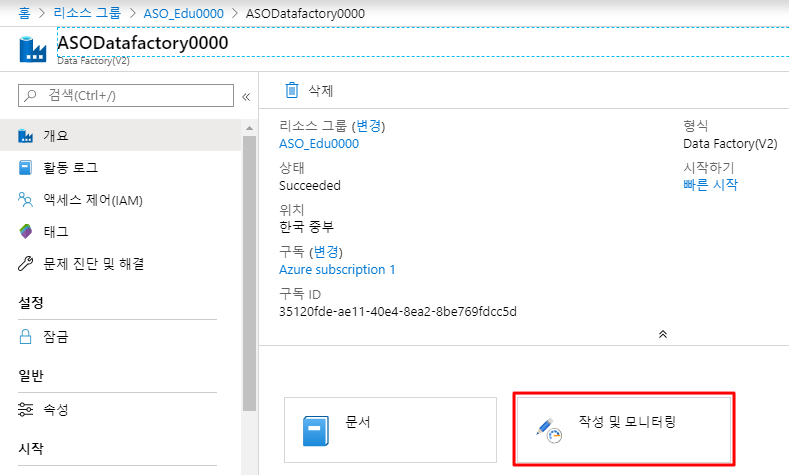
Data Factory 위치는 Data Factory의 메타데이터가 저장되고 파이프라인의 트리거가 시작되는 위치입니다. 한편, Data Factory는 다른 Azure 영역의 데이터 저장소 및 컴퓨팅 서비스에 액세스하여 데이터 저장소 간에 데이터를 이동하거나 컴퓨팅 서비스를 사용하여 데이터를 처리할 수 있습니다. 이 동작은 전 세계적으로 사용 가능한 통합 런타임을 통해 실현되어 데이터 컴플라이언스, 효율성 및 네트워크 송신 비용 절감을 보장합니다.

1. Exercise 2: Orchestrate with Azure Data Factory

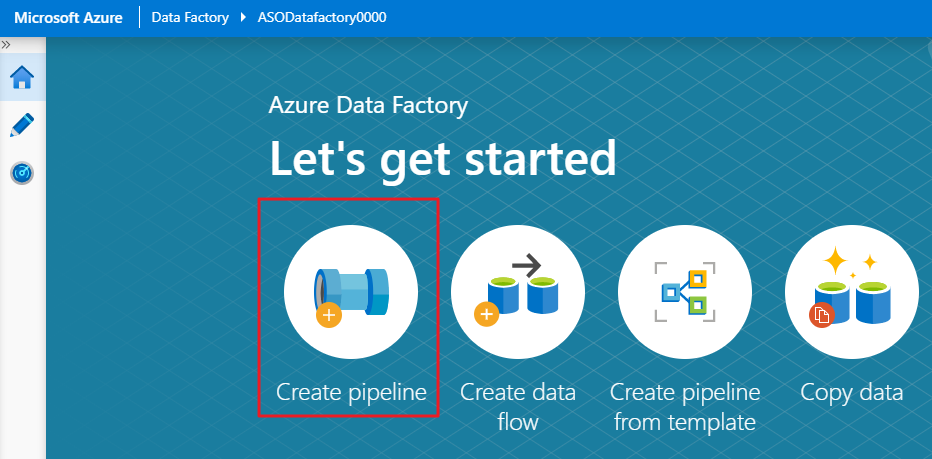
이 실습에서는 Azure Data Factory의 Pipeline을 이용하여 특정 위치의 파일의 데이터를 이관하며, Azure Databricks, Azure Machine Learning 등 Azure의 다른 서비스들과 연결하여 전체적인 데이터의 흐름을 관리합니다.

* 1. Task 1: Amazon S3 to Azure Gen2

1. 리소스 그룹으로 이동하여 **Data Factory**를 선택 후 **작성 및 모니터링**을 클릭합니다.



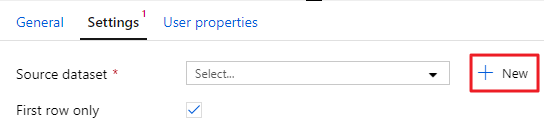
1. Data Factory 포털 홈 화면에서 **Create pipline**을 선택합니다.



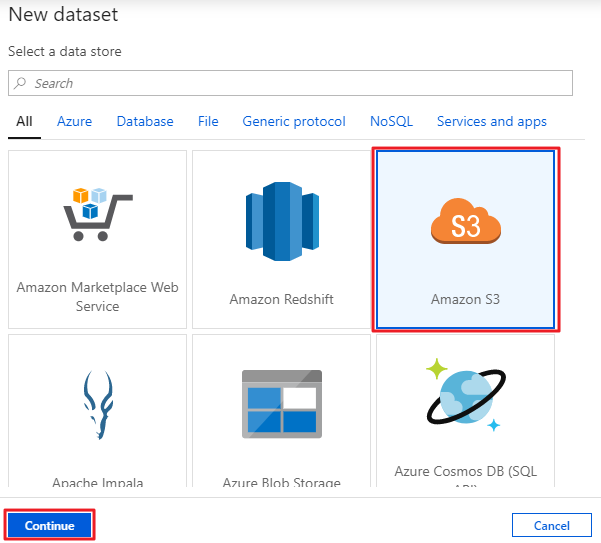
1. **Activities > General > Lookup**을 드래그해서 옮긴 후 하단 **General** 탭에서 다음을 입력합니다.

* Name: **Check Amazon S3**

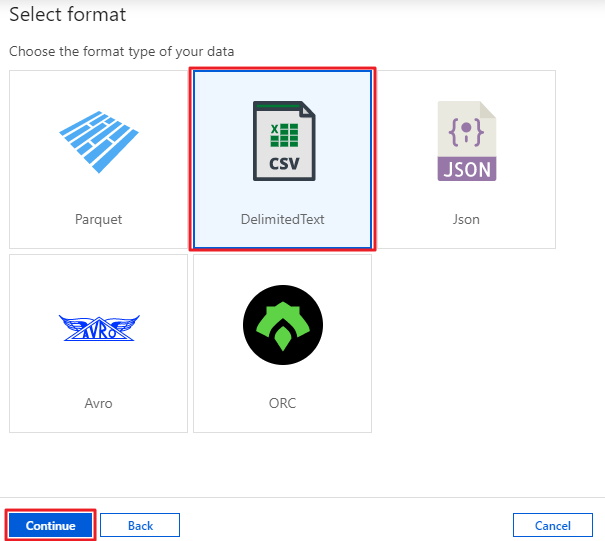
1. **Setting**탭으로 이동하여 Source dataset의 **+New**를 클릭합니다.



1. **New dataset**의 **All** 목록에서 **Amazon S3** 를 선택합니다.

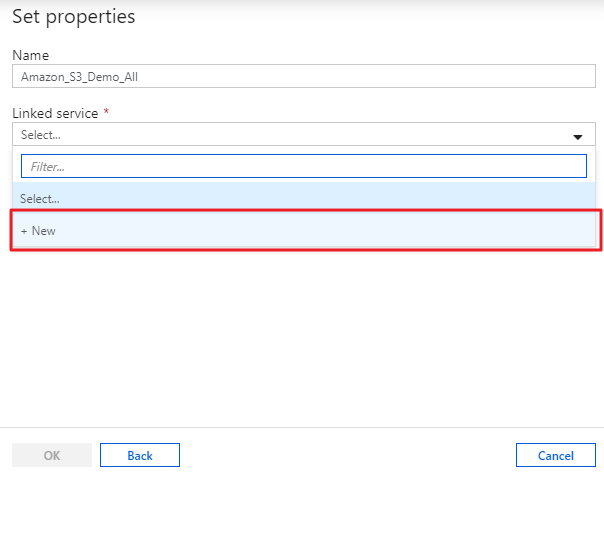


1. **format은 DelimitedText**를선택합니다.



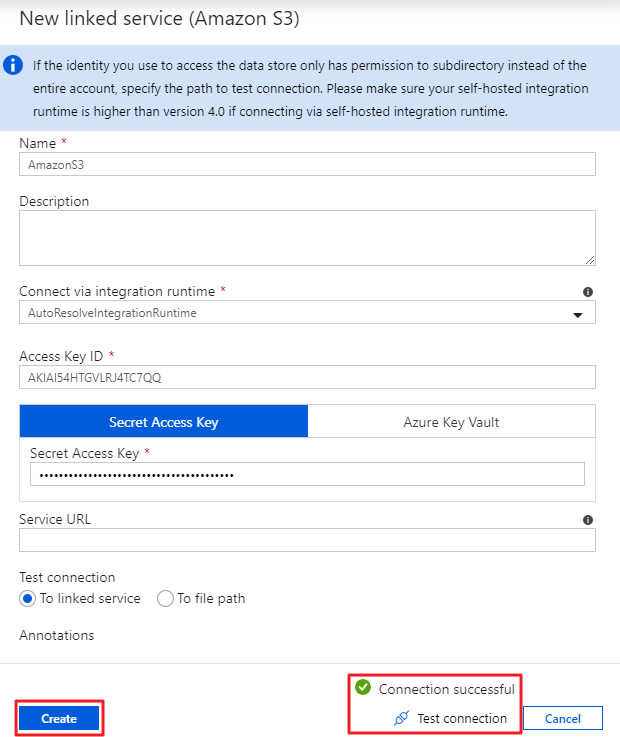
1. Name 항목 입력 후 **Linked service** 에서 **+New**를선택합니다.

* Name : **Amazon\_S3\_Demo\_All**



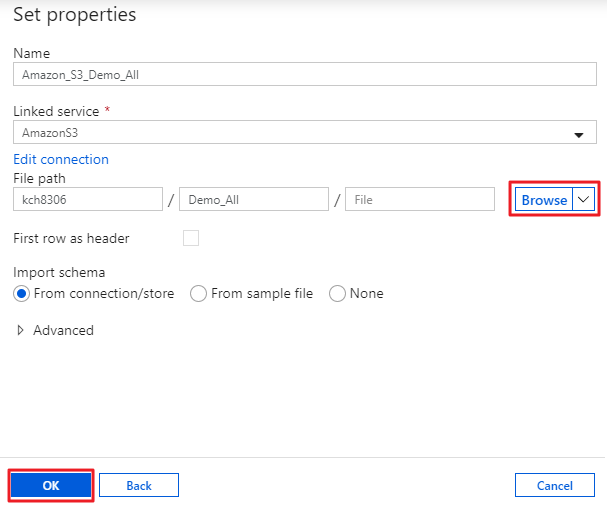
1. **New linked service (Amazon S3)** 블레이드가 열리면, 아래와 같이 정보를 입력하고 **Test Connection** 버튼을 클릭하여 Connection successful 메시지가 출력되면 **Create** 버튼을 클릭합니다.

* Name : **AmazonS3**
* Access Key ID : **AKIAI54HTGVLRJ4TC7QQ**
* Secret Access Key : **OOr4sejC3lCrmT7Nuv4RaOezH7UDatbh4sl+ztgx**

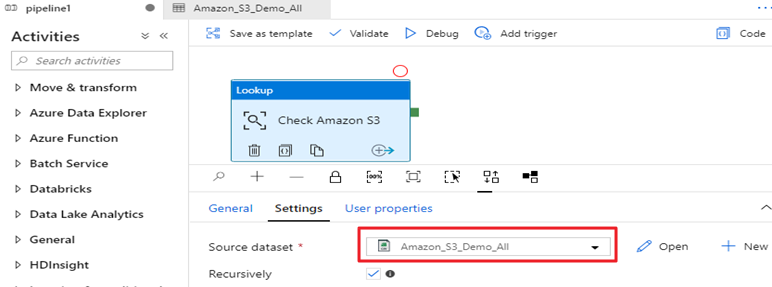


1. **Set Properties** 블레이드가 열리면 **File path** 항목의 **Browse** 버튼을 클릭하여 파일 경로를 **kch8306 > Demo\_All**로 지정 후 **OK** 버튼을 클릭하여 완료합니다.

* Name : **AmazonS3**
* File path : **kch8306 > Demo\_All**



1. **Lookup** 액티비티의 **Setting**탭의 **Source Dateset**이 세팅된 것을 확인합니다.

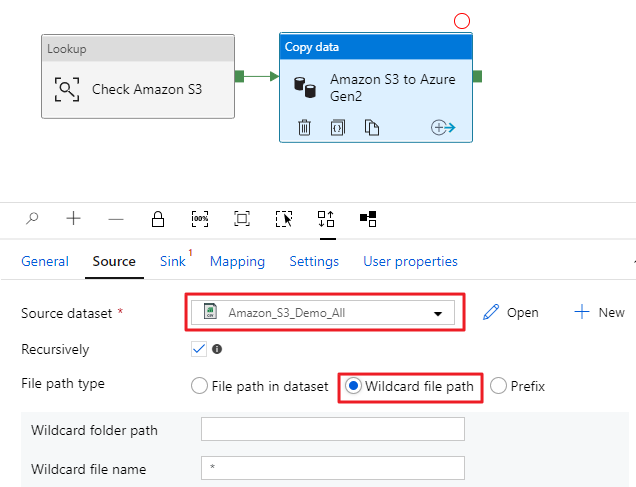


1. **Activities > Move & transform > Copy data**를 드래그해서 옮긴 후 하단 **General** 탭에서 다음을 입력합니다.

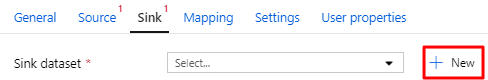
* Name: **Amazon S3 to Azure Gen2**

1. **Lookup** 액티비티와 **Copy data** 액티비티를 연결시킨 후 **Source**탭으로 이동하여 다음을 입력합니다.

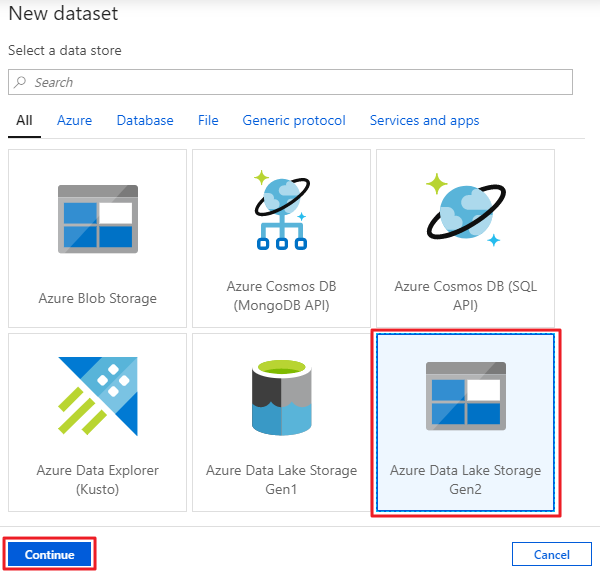
* Source dataset : **Amazon\_S3\_Demo\_All**
* File path Type : **Wildcard file path**



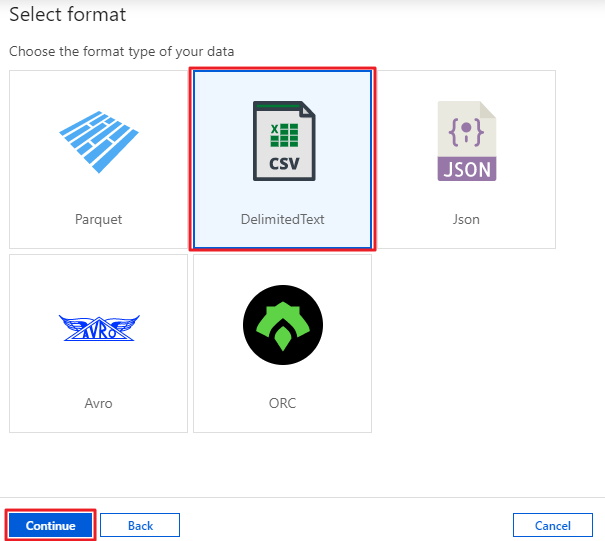
1. Sink탭으로 이동하여 Sink dataset의 **+New**를 클릭합니다.



1. New Datasource 블레이드가 나타나면 Azure Data Lake Storage Gen2 를 선택 후 Continue 버튼을 클릭합니다.

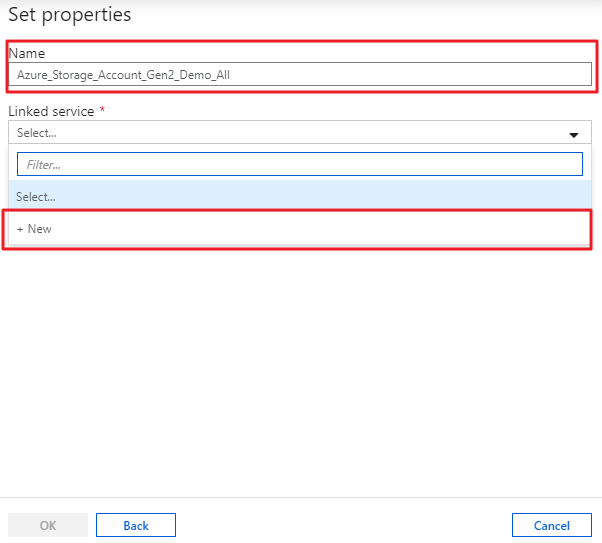


1. format은 **DelimitedText**를선택합니다.



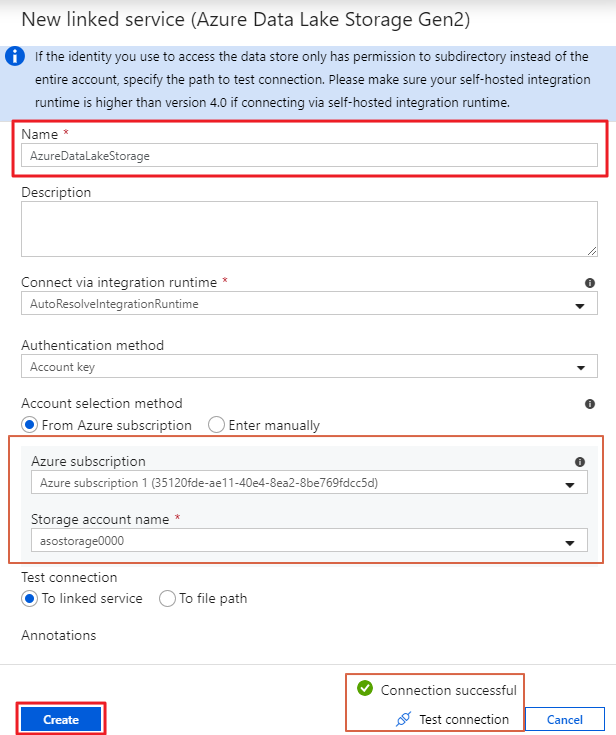
1. **Set Properties** 블레이드가 열리면 Name 항목을 입력 후 Linked service 의 **+New** 를 선택합니다.

* Name: **Azure\_Storage\_Account\_Gen2\_Demo\_All**



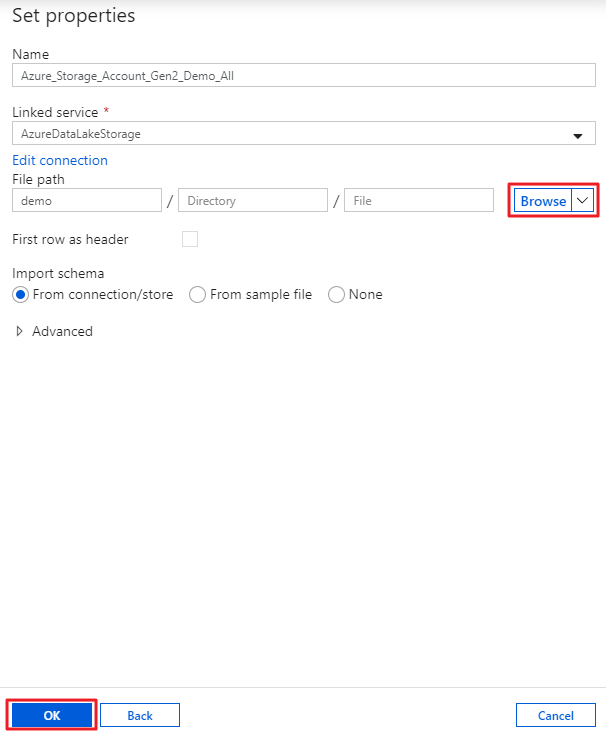
1. **New linked service (Azure Data Lake Storage Gen2)** 블레이드가 열리면, 아래와 같이 정보를 입력하고 **Test Connection** 버튼을 클릭하여 Connection successful 메시지가 출력되면 **Create** 버튼을 클릭합니다.

* Name : **AzureDataLakeStorage**
* Azure subscription: **본인의 구독**
* Storage account name: **asostorageXXXX**

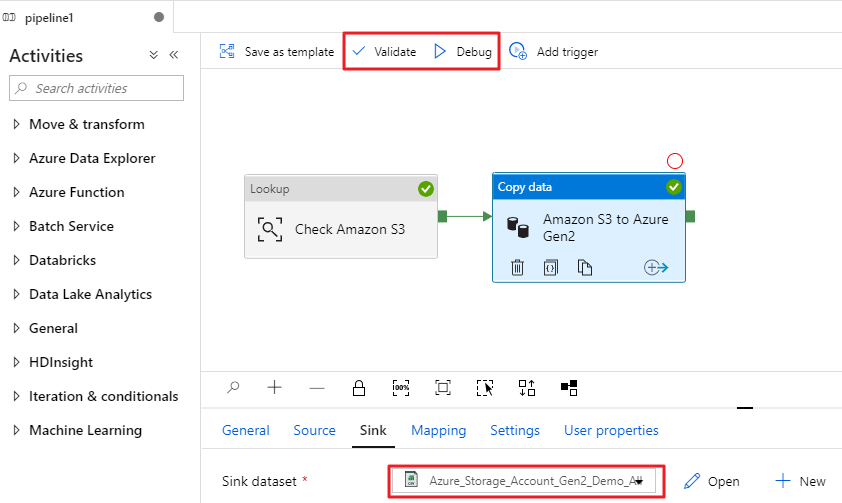


1. **Set Properties** 블레이드가 열리면 아래의 정보를 입력 후 **OK** 버튼을 클릭하여 완료합니다.

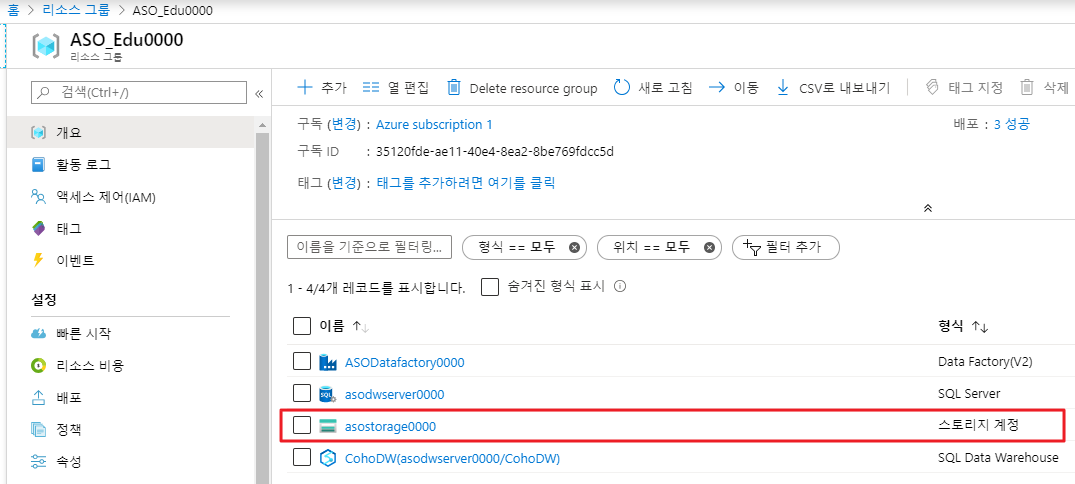
* Name : **Azure\_Storage\_Account\_Gen2\_Demo\_All**
* Linked service : **AzureDataLakeStorage**
* File path : **demo /** (browse 버튼 클릭하여 경로 지정)



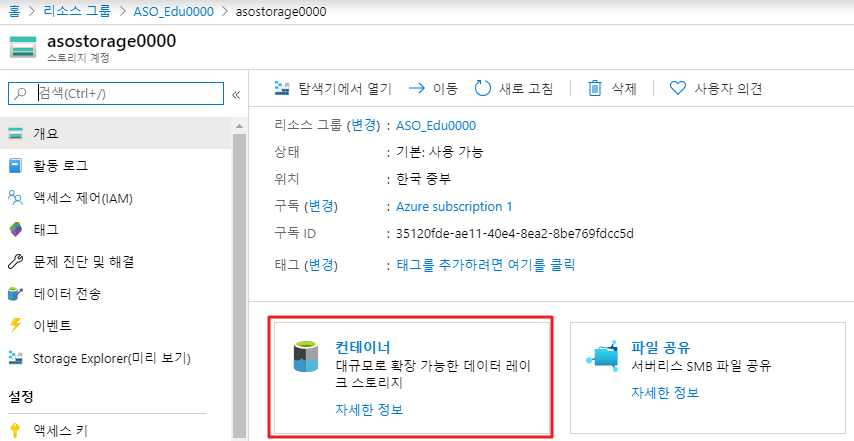
1. **Copy data** 액티비티의 **Sink**탭의 **Source Dataset**이 세팅된 것을 확인 후 상단의 **validate** 버튼 클릭 후 **Debug** 버튼을 클릭하여 실행합니다.



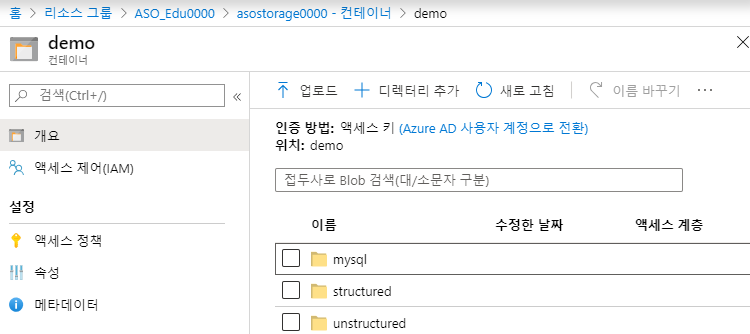
1. 실행된 결과를 확인하기 위해 Azure Portal 의 리소스그룹으로 이동하여 스토리지 계정을 선택합니다.



1. **컨테이너** 버튼을 클릭합니다.

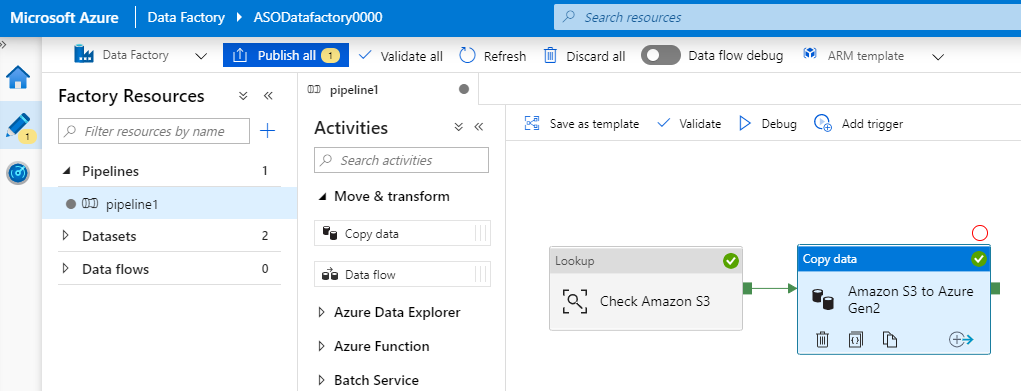


1. 컨테이너 목록에서 **demo**를 선택하여 아래와 같이 폴더가 생성 되었는지 확인합니다.



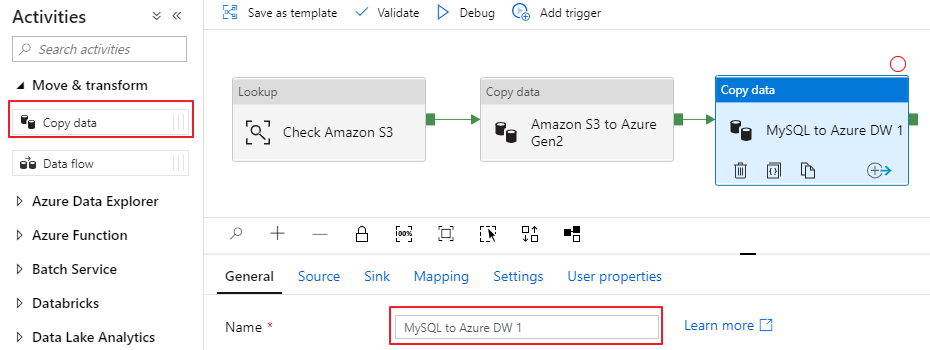
* 1. Task 2: MySQL to Azure Synapse Analytics

1. Data Factory 포털 화면으로 이동합니다.

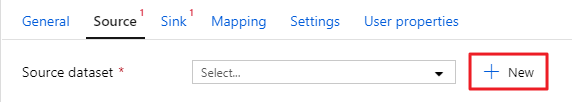


1. **Activities > Move & transform > Copy data**를 드래그해서 옮긴 후 하단 **General** 탭에서 다음을 입력합니다.

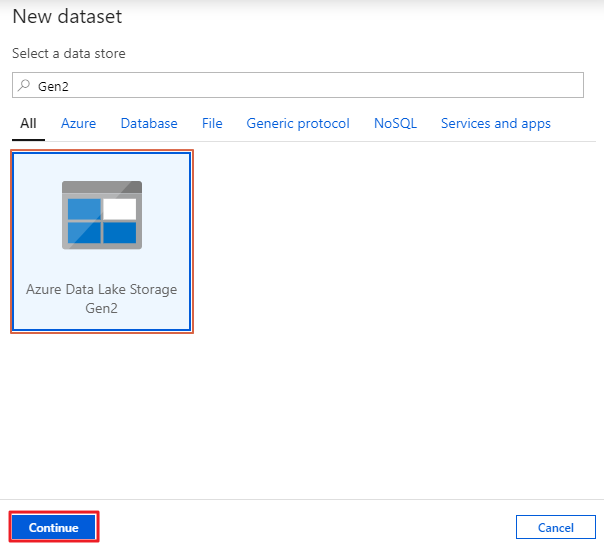
* Name: **MySQL to Azure DW 1**



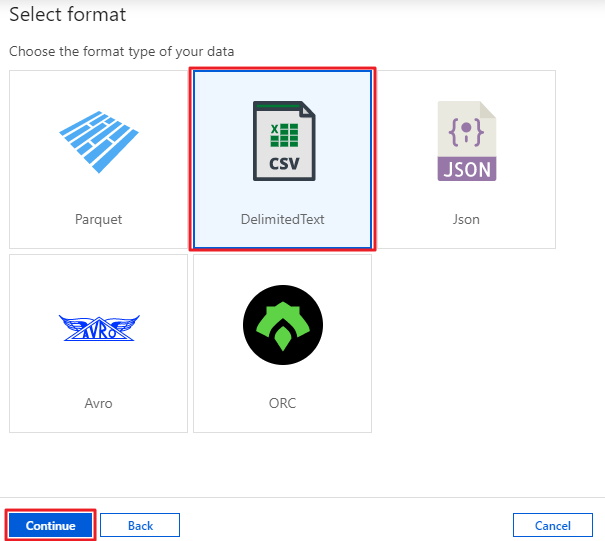
1. Source 탭으로 이동하여 Source dataset의 **+New** 를 클릭합니다.



1. New dataset 블레이드에서 Gen2를 검색하여 **Azure Data Lake Storage Gen2**를 선택 후 **Continue**를 클릭합니다.

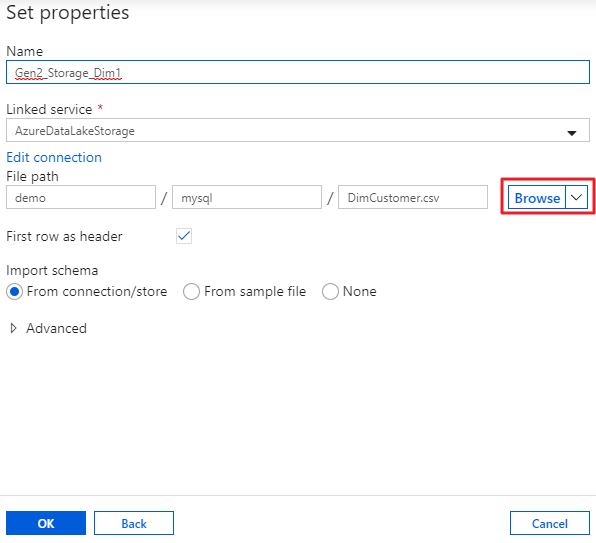


1. format**은 DelimitedText**를선택합니다

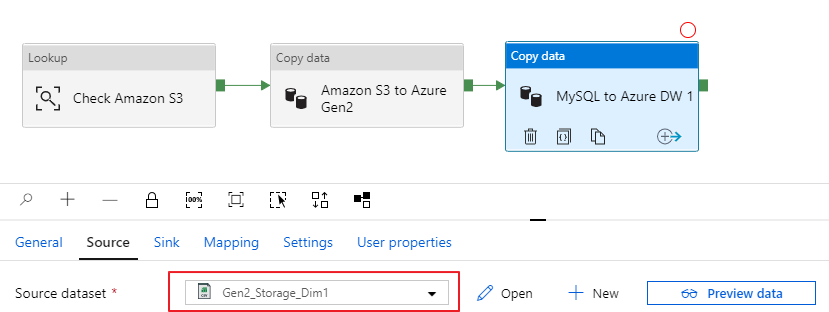


1. Set Properties블레이드가 열리면 아래의 정보를 입력 후 **OK**버튼을 클릭합니다.

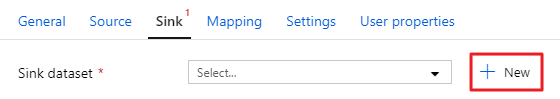
* Name: **Gen2\_Storage\_Dim1**
* Linked service : **AzureDataLakeStorage**
* File path **: demo / mysql / DimCustomer.csv** (browse 버튼 클릭하여 경로 지정)
* First row as header : **Check**



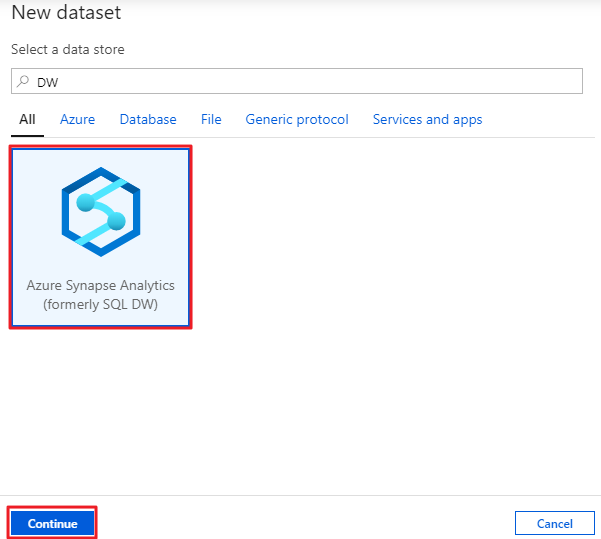
1. Source탭의 **Source Dateset**이 세팅된 것을 확인합니다.



1. Sink탭으로 이동하여 Sink dataset의 **+New**를 클릭합니다.

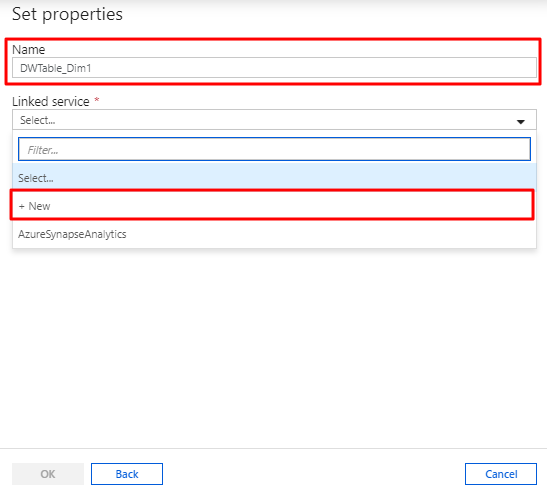


1. New dataset 블레이드에서 **DW**를 검색하여 **Azure Synapse Analytics**를 선택 후 **Continue**를 클릭합니다.



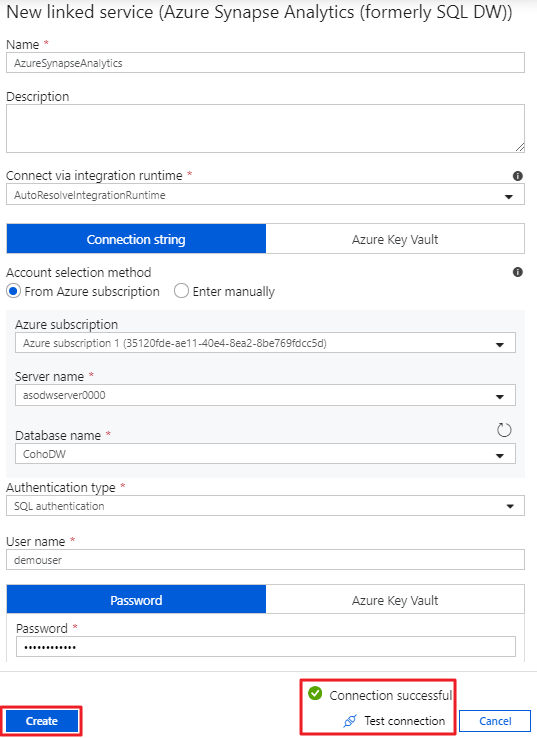
1. **Set Properties** 블레이드가 열리면 Name 항목을 입력 후 Linked service 항목을 클릭한 후 **+New** 를 선택합니다.

* Name: **DWTable\_Dim1**



1. New linked service (Azure Synapse Analytics (formerly SQL DW))블레이드가 열리면, 아래와 같이 정보를 입력하고 **Test Connection** 버튼을 클릭하여 Connection successful 메시지가 출력되면 **Create** 버튼을 클릭합니다.

* Name : **AzureSynapseAnalytics**
* Azure subscription : **본인의 구독**
* Server name : **asodwserverXXXX** (DW 계정 명)
* Database name : **CohoDW**
* User name : **demouser**
* Password : **Demo@pass123**

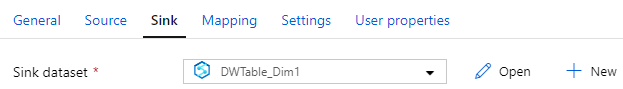


1. Set properties 화면이 열리면 아래와 같이 입력 후 OK 버튼을 클릭합니다.

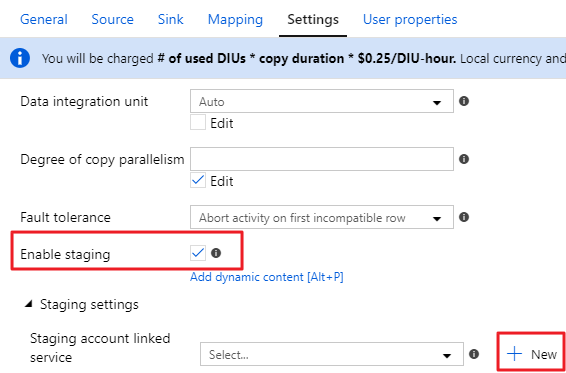
* Name: **DWTable\_Dim1**
* Linked service : **AzureSynapseAnalytics**
* Table Name : **dbo.DimCustomer**



1. **Copy data** 액티비티의 **Sink**탭의 **Source Dateset**이 세팅된 것을 확인합니다.

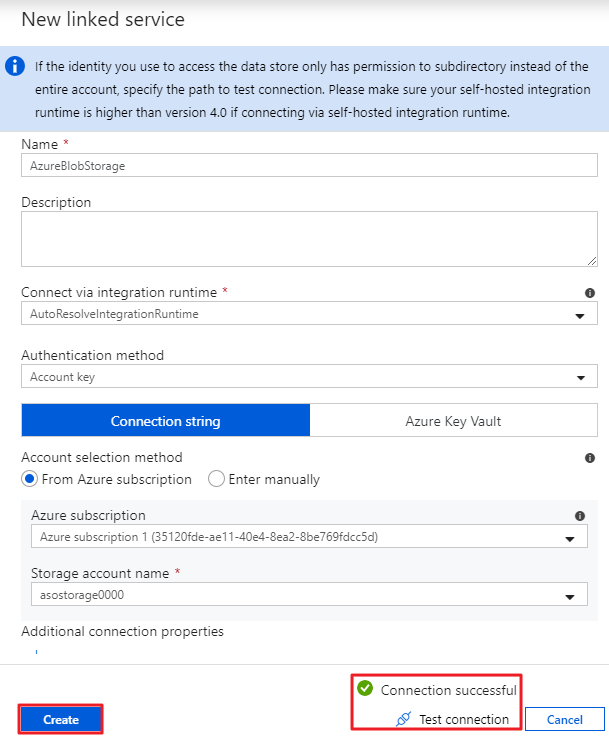


1. Settings 탭으로 이동하여 Enable staging 항목에 체크 후 Staging account linked service 항목의 **+New**를 클릭합니다.

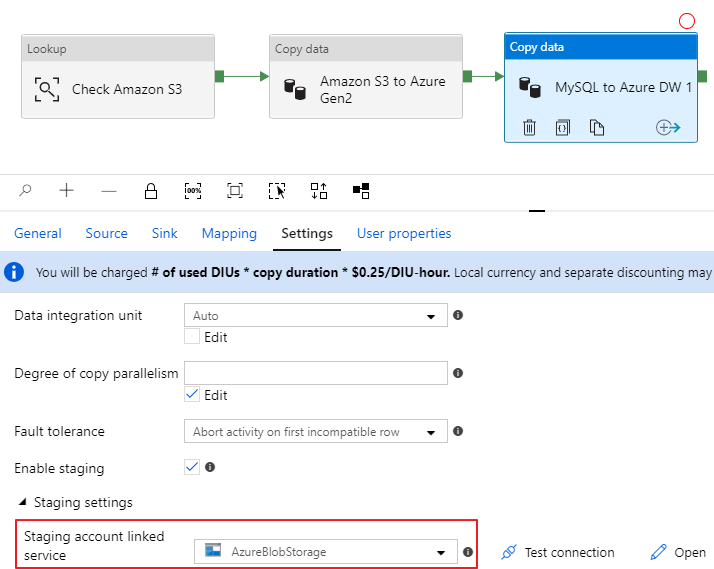


1. 다음 내용을 입력 후 **Test connection**을 클릭하여 연결을 확인한 후 **Create** 버튼을 클릭합니다.

* Name: **AzureBlobStorage(임의입력)**
* Account selection method: **From Azure subscription**
* Azure subscription: **본인의 구독**
* Storage account name: **asodwserverXXXX** (DW 계정 명)

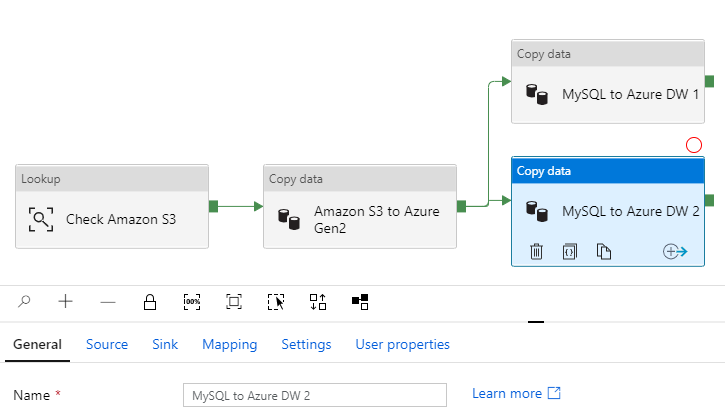


1. Settings 탭으로 이동하여 Enable staging 항목에 체크 후 Staging account linked service 항목을 **AzureBlobStorage**로 선택합니다.

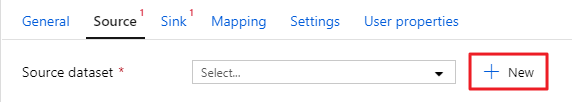


1. **Activities > Move & transform > Copy data**를 드래그해서 옮긴 후 하단 **General** 탭에서 다음을 입력합니다.

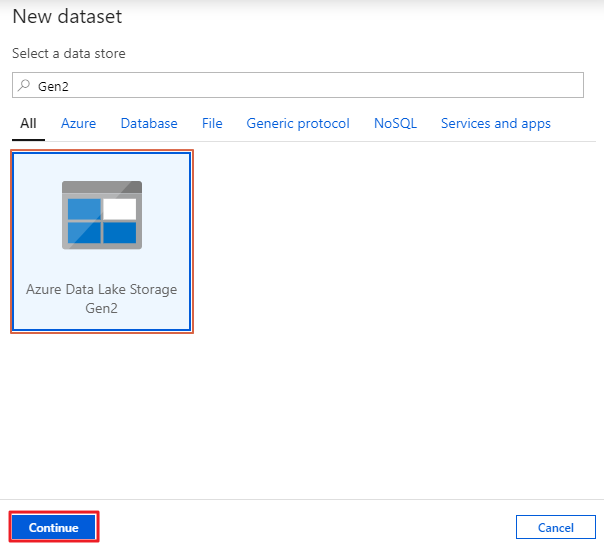
* Name: **MySQL to Azure DW 2**



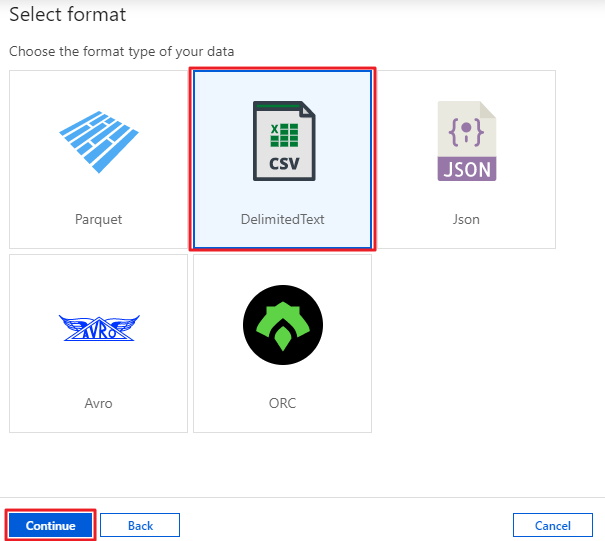
1. Source 탭으로 이동하여 Source dataset의 **+New** 를 클릭합니다.



1. New dataset 블레이드에서 Gen2를 검색하여 **Azure Data Lake Storage Gen2**를 선택 후 **Continue**를 클릭합니다.

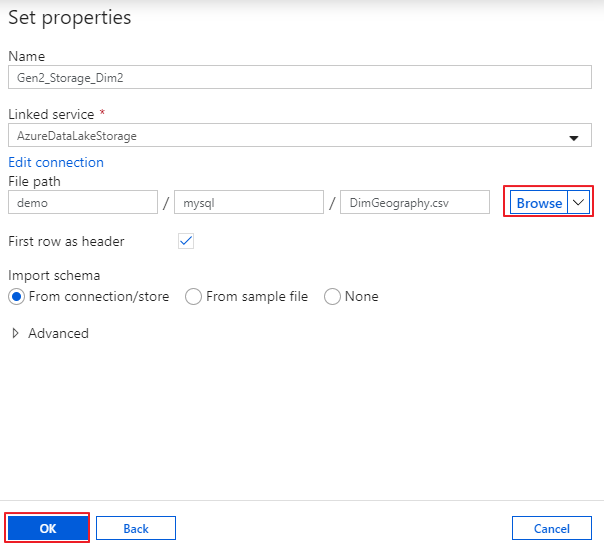


1. format**은 DelimitedText**를선택합니다

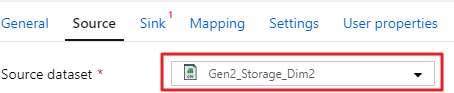


1. Set Properties블레이드가 열리면 아래의 정보를 입력 후 **OK**버튼을 클릭합니다.

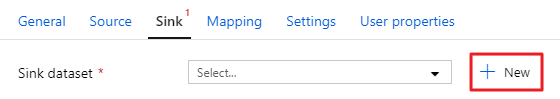
* Name: **Gen2\_Storage\_Dim2**
* Linked service : **AzureDataLakeStorage**
* File path **: demo / mysql / DimGeography.csv**
* First row as header : **Check**



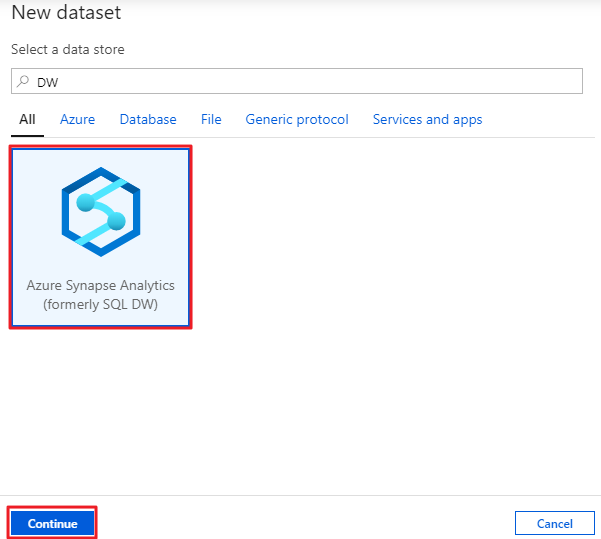
1. Source탭의 **Source Dateset**이 세팅된 것을 확인합니다.



1. Sink탭으로 이동하여 Sink dataset의 **+New**를 클릭합니다.

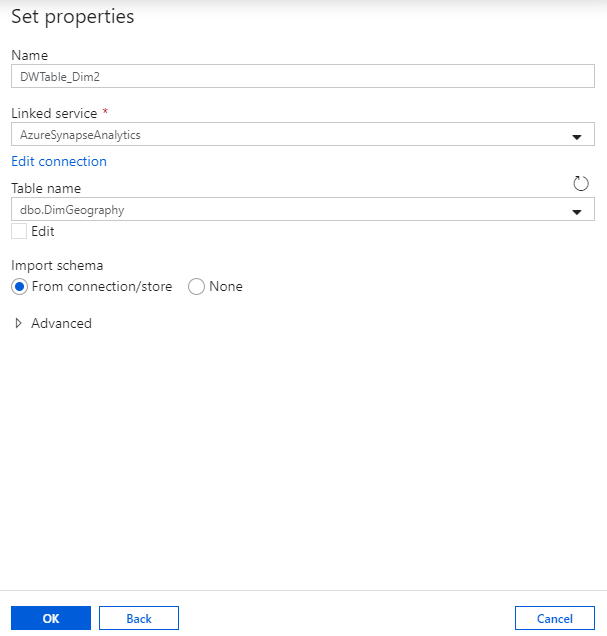


1. New dataset 블레이드에서 **DW**를 검색하여 **Azure Synapse Analytics**를 선택 후 **Continue**를 클릭합니다.

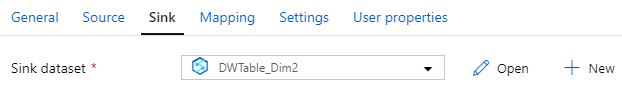


1. Set properties 화면이 열리면 아래와 같이 입력 후 OK 버튼을 클릭합니다.

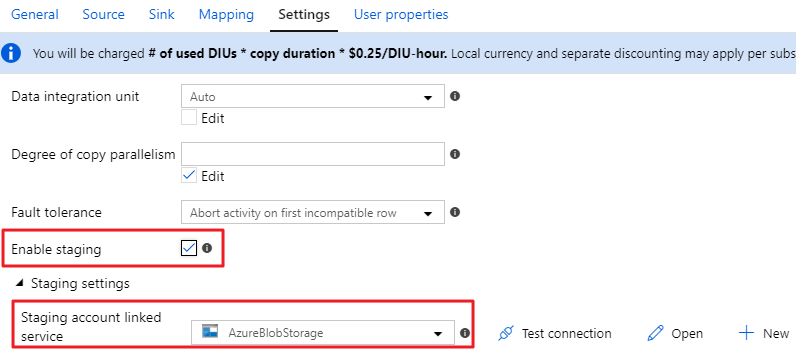
* Name: **DWTable\_Dim2**
* Linked service : **AzureSynapseAnalytics**
* Table Name : **dbo.DimGeography**



1. **Copy data** 액티비티의 **Sink**탭의 **Source Dateset**이 세팅된 것을 확인합니다.

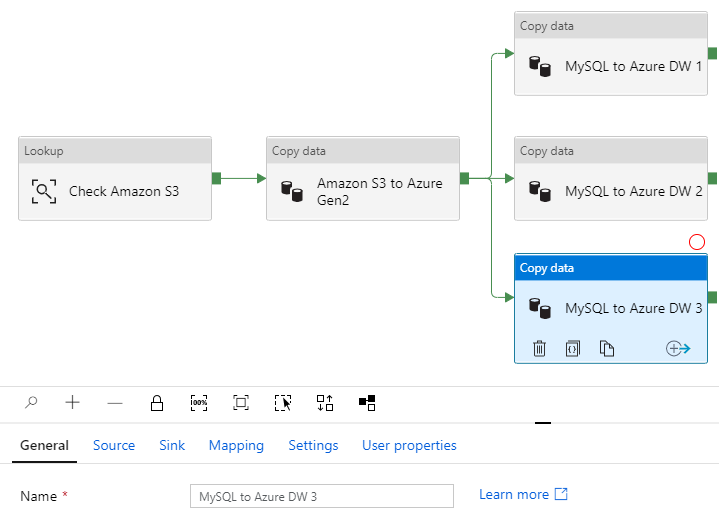


1. Settings 탭으로 이동하여 Enable staging 항목에 체크 후 Staging account linked service 항목은 **AzureBlobStorage**를 선택합니다.

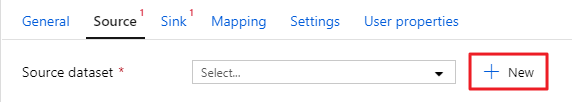


1. **Activities > Move & transform > Copy data**를 드래그해서 옮긴 후 하단 **General** 탭에서 다음을 입력합니다.

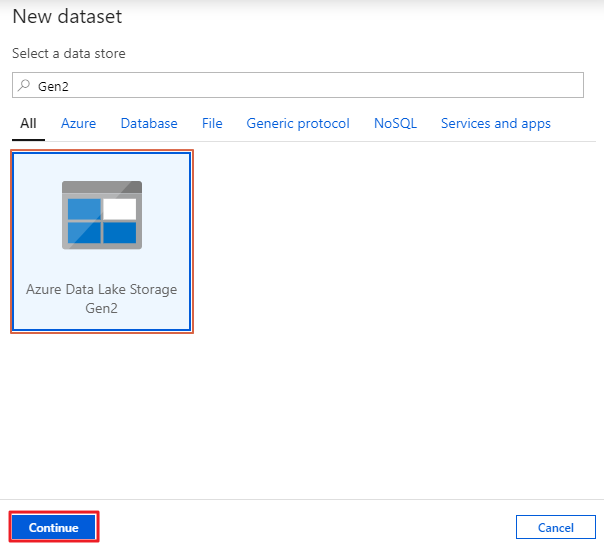
* Name: **MySQL to Azure DW 3**



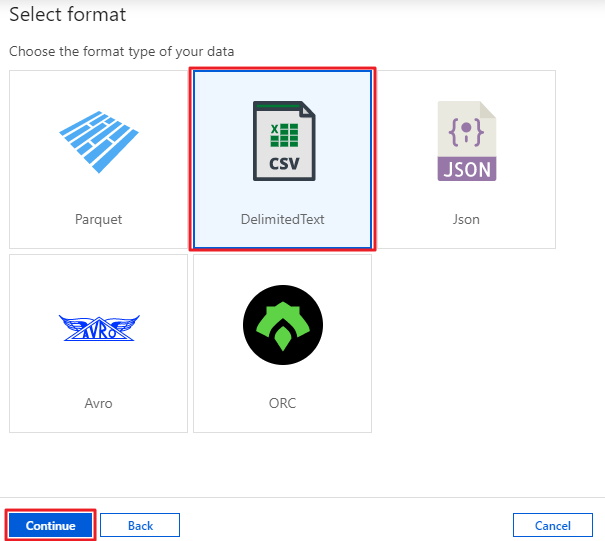
1. Source 탭으로 이동하여 Source dataset의 **+New** 를 클릭합니다.



1. New dataset 블레이드에서 Gen2를 검색하여 **Azure Data Lake Storage Gen2**를 선택 후 **Continue**를 클릭합니다.

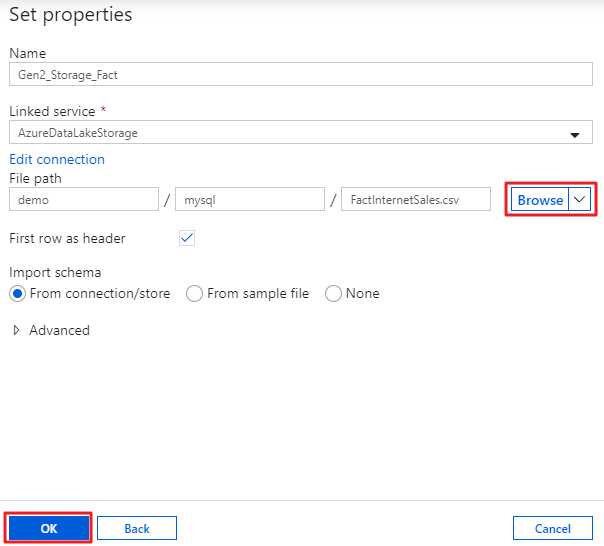


1. format**은 DelimitedText**를선택합니다

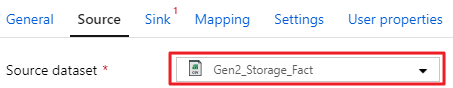


1. Set Properties블레이드가 열리면 아래의 정보를 입력 후 **OK**버튼을 클릭합니다.

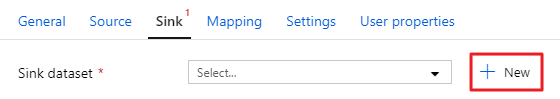
* Name: **Gen2\_Storage\_Fact**
* Linked service : **AzureDataLakeStorage**
* File path **: demo / mysql / FactInternetSales.csv** (Browse 버튼 클릭하여 경로지정)
* First row as header : **Check**



1. Source탭의 **Source Dateset**이 세팅된 것을 확인합니다.

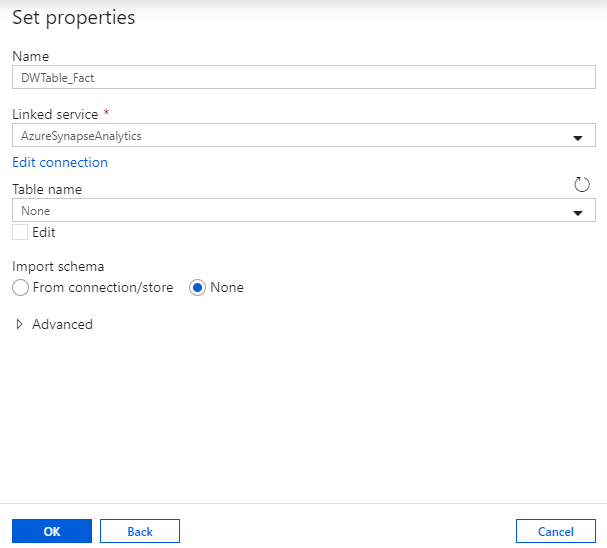


1. Sink탭으로 이동하여 Sink dataset의 **+New**를 클릭합니다.

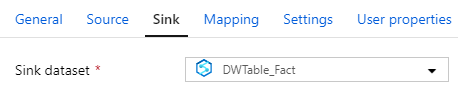


1. Set properties 화면이 열리면 아래와 같이 입력 후 OK 버튼을 클릭합니다.

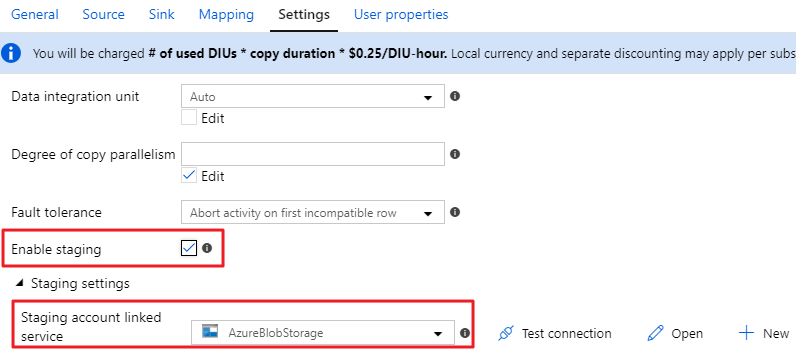
* Name: **DWTable\_Fact**
* Linked service : **AzureSynapseAnalytics**
* Table Name : **dbo.FactInternetsales**



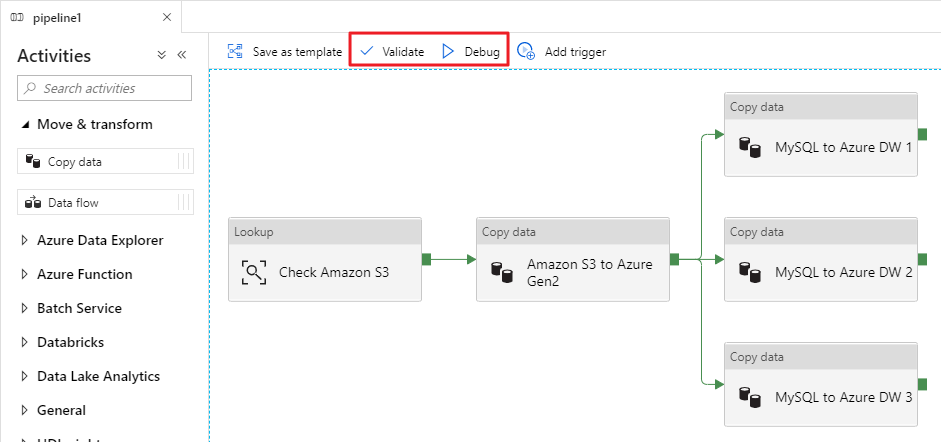
1. **Copy data** 액티비티의 **Sink**탭의 **Source Dateset**이 세팅된 것을 확인합니다.



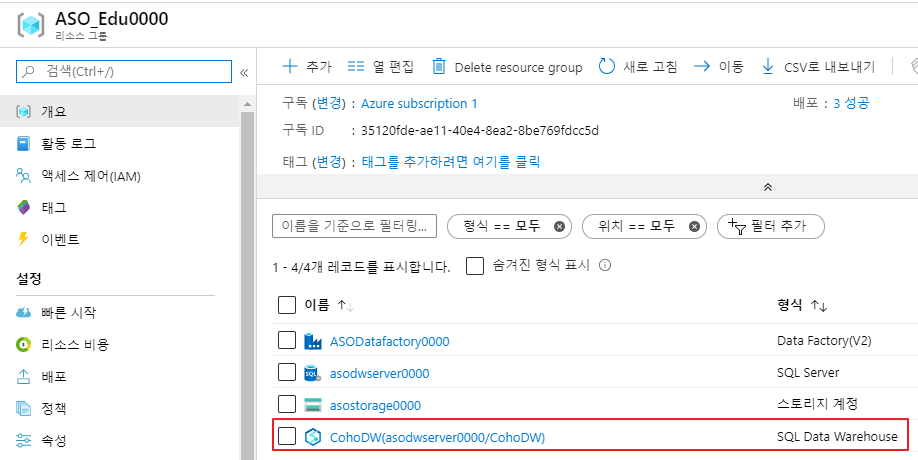
1. Settings 탭으로 이동하여 Enable staging 항목에 체크 후 Staging account linked service 항목은 **AzureBlobStorage**를 선택합니다.



1. 상단의 **validate** 버튼 클릭 후 **debug** 버튼을 클릭하여 실행합니다.

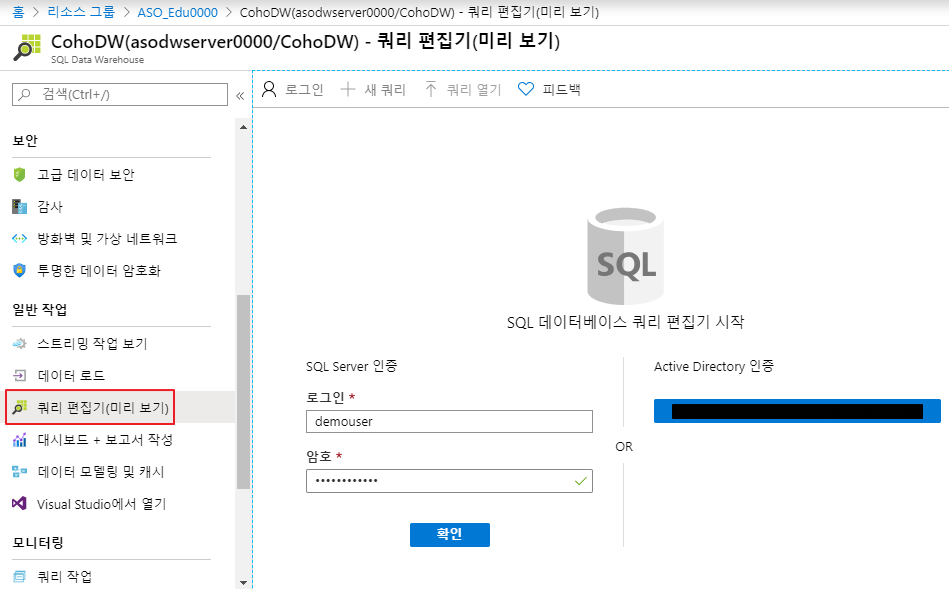


1. 실행 결과를 확인하기 위해 Azure Portal 의 리소스 그룹으로 이동하여 **CohoDW**를 선택합니다..



1. 왼쪽 메뉴에서 **쿼리 편집기**를 클릭 후 아래의 정보를 입력하여 로그인합니다.

* 로그인: **demouser**
* 암호: **Demo@pass123**

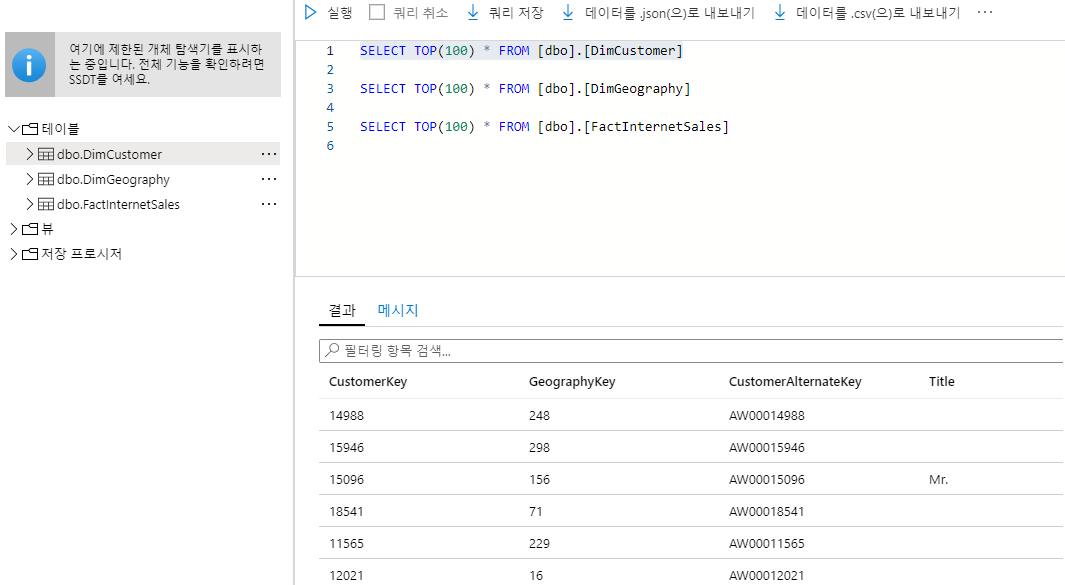


1. 다음 쿼리를 1라인씩 블록 지정 하여 실행하여테이블 별 데이터가 정상적으로 조회되는지 확인합니다.

SELECT TOP(100) \* FROM [dbo].[DimCustomer]

SELECT TOP(100) \* FROM [dbo].[DimGeography]

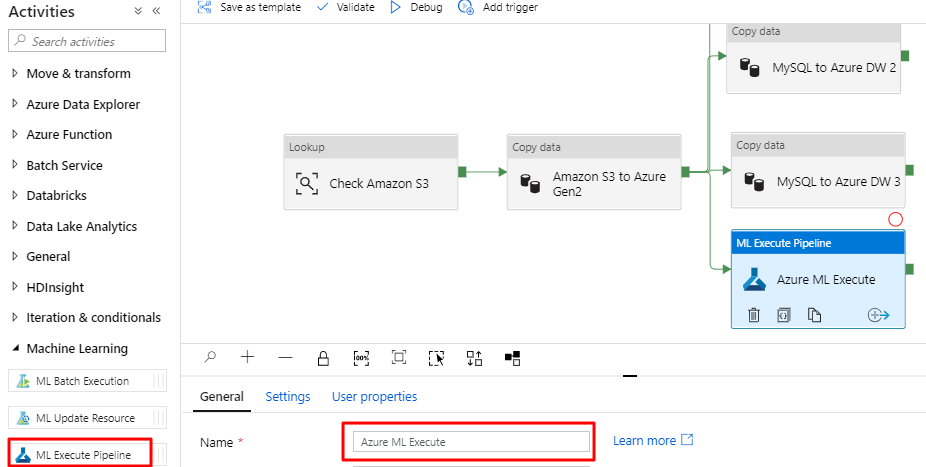
SELECT TOP(100) \* FROM [dbo].[FactInternetSales]



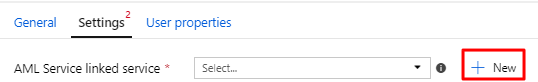
* 1. Task 3: Azure ML Execute

1. Data Factory 포털 화면으로 이동합니다.
2. **Activities > Machine Learning> ML Execute Pipeline**을 드래그해서 옮긴 후 하단 **General** 탭에서 다음을 입력합니다.

* Name: **Azure ML Execute**

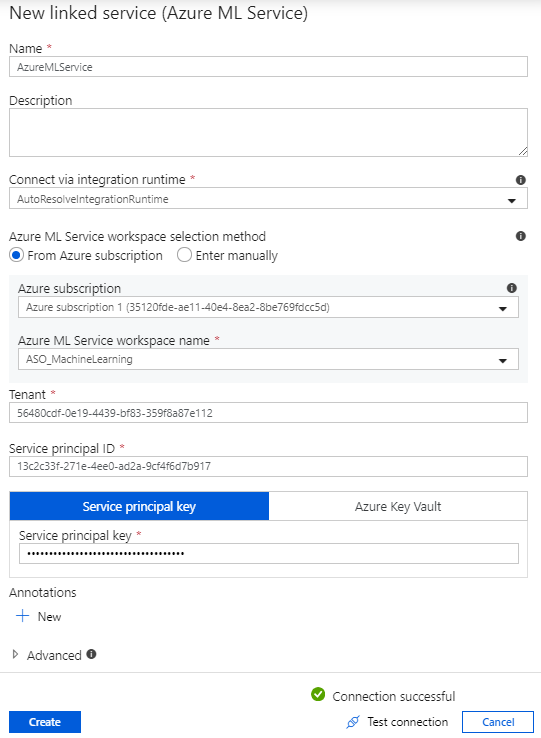


1. **Setting**탭으로 이동하여 AML Service linked service의 **+New**를 클릭합니다.



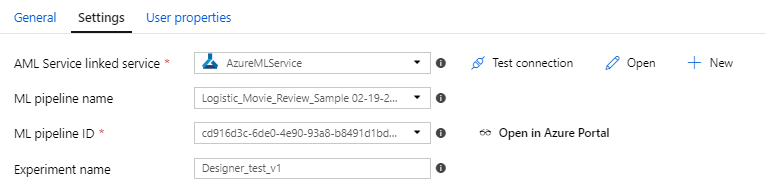
1. New linked service (Azure Synapse Analytics (Azure ML Service))블레이드가 열리면, 아래와 같이 정보를 입력하고 **Test Connection** 버튼을 클릭하여 Connection successful 메시지가 출력되면 **Create** 버튼을 클릭합니다.

* Name : **AzureMLService**
* Azure ML Service workspace selection method: **From Azure subscription**
* Azure subscription : **본인의 구독**
* Azure ML Service workspace name: **ASO\_MachineLearning**
* Tenant : **56480cdf-0e19-4439-bf83-359f8a87e112**
* Service principal ID: **13c2c33f-271e-4ee0-ad2a-9cf4f6d7b917**
* Service principal key : **40932bbf-4739-4c21-95a1-aa08f1a3e5bc**



1. AML Service linked service 항목이 세팅된 것을 확인 후 아래의 정보를 입력합니다.

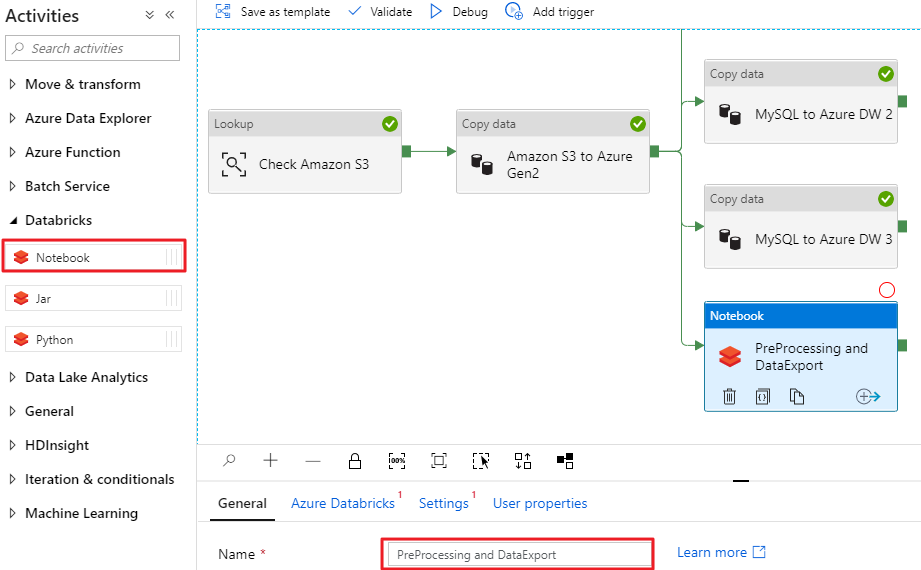
* ML pipeline name: **Logistic\_Movie\_Review\_Sample**
* ML pipeline ID: **cd916d3c-6de0-4e90-93a8-b8491d1bd978..**
* Experiment name : **Designer\_test\_v1**



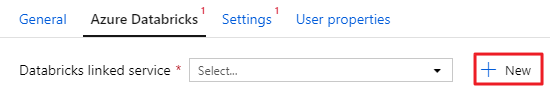
* 1. Task 4: PreProcessing and DataExport

1. **Activities > Databricks > Notebook**을 드래그해서 옮긴 후 하단 **General** 탭에서 다음을 입력합니다.

* Name: **PreProcessing and DataExport**

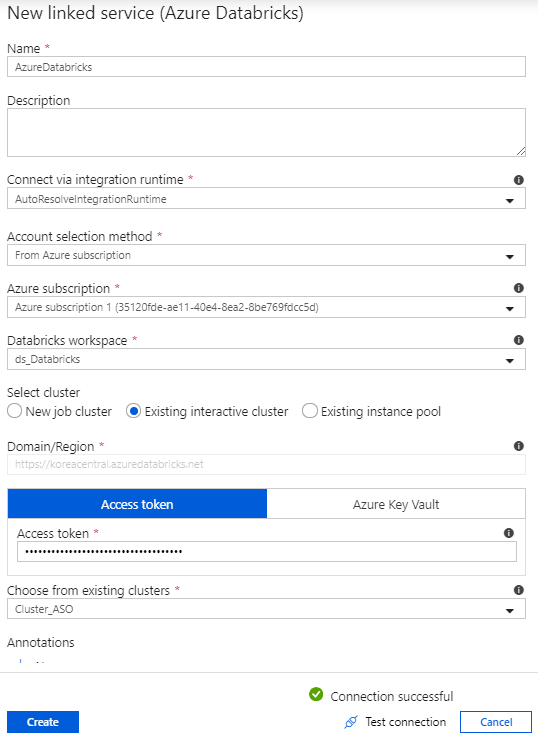


1. **Azure Databricks**탭으로 이동하여 Databricks linked service의 **+New**를 클릭합니다.



1. New linked service (Azure Synapse Analytics (Azure Databricks))블레이드가 열리면, 아래와 같이 정보를 입력하고 **Test Connection** 버튼을 클릭하여 Connection successful 메시지가 출력되면 **Create** 버튼을 클릭합니다.

* Name : **AzureDatabricks**
* Account selection method: **From Azure subscription**
* Azure subscription : **본인의 구독**
* Databricks workspace: **ds\_Databricks**
* Access token : **dapid1fb7b77a6d0e7f9fb2ccf337367a9a5**
* Select cluster : **Existing interactive cluster**
* Existing cluster ID: **Cluster\_ASO**

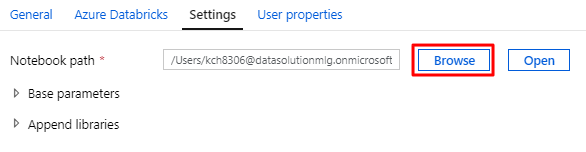


1. Databricks linked service 항목이 세팅된 것을 확인합니다.



1. Settings 탭으로 이동하여 Notebook path 항목의 Browse 버튼을 클릭하여 다음과 같이 경로를 지정합니다.

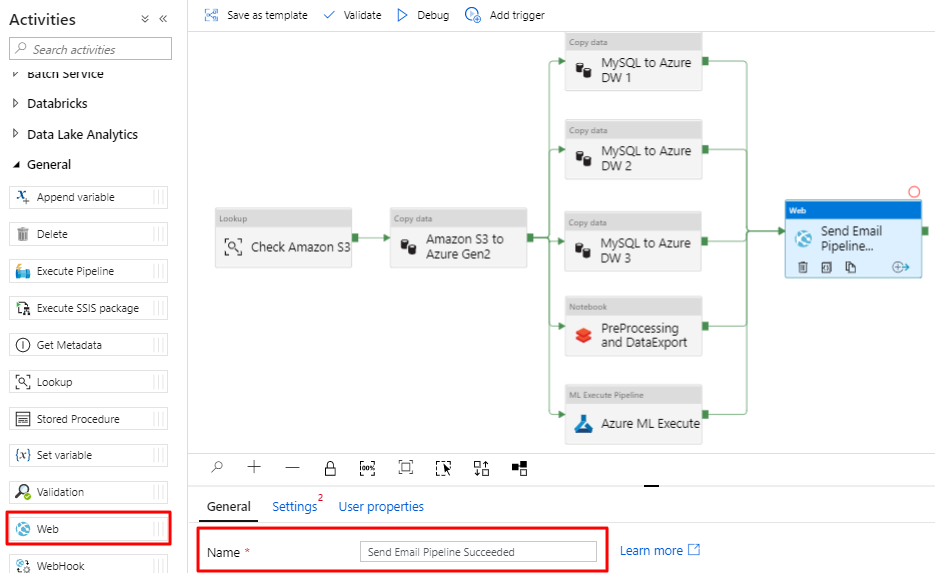
* Notebook path: **/Users/kch8306@datasolutionmlg.onmicrosoft.com/PreProcessing\_ASO**



* 1. Task 5: Send Email

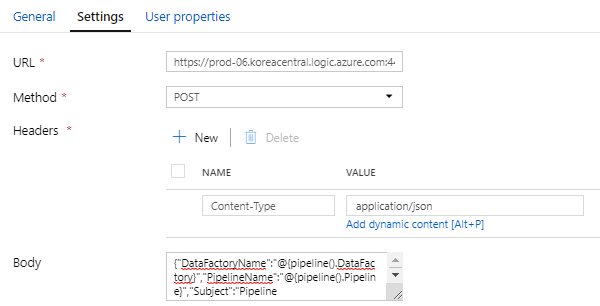
1. **Activities > General> Web**을 드래그해서 옮긴 후 하단 **General** 탭에서 다음을 입력합니다.

* Name: **Send Email Pipeline Succeeded**

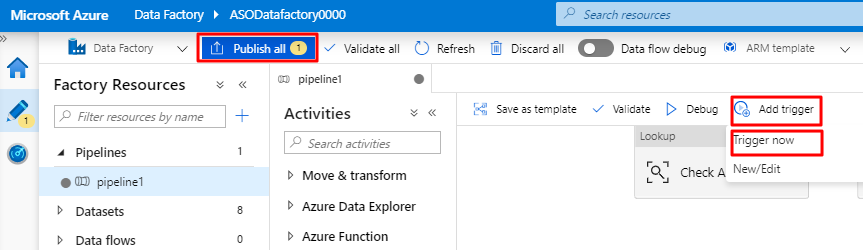


1. **Settings** 탭으로 이동하여 아래의 정보를 입력합니다.

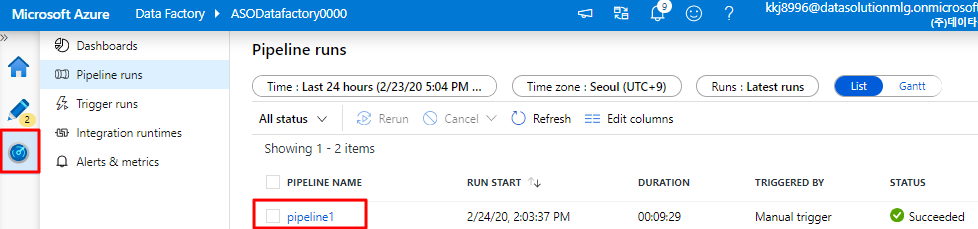
* URL:**https://prod-06.koreacentral.logic.azure.com:443/workflows/21f3052675d24be0aa4298a9b0490ce8/triggers/manual/paths/invoke?api-version=2016-10-01&sp=%2Ftriggers%2Fmanual%2Frun&sv=1.0&sig=BfXuXTp05p20wnvDFzkUgi3xbf3XGSB-1qapyC\_EDFA**
* Method: **POST**
* Headers: **+New 클릭 후 Content-Type application/json 추가**
* Body: **{"DataFactoryName":"@{pipeline().DataFactory}","PipelineName":"@{pipeline().Pipeline}","Subject":"Pipeline Succeeded","Message":"Succeeded!","EmailTo":"본인의 메일계정"}**



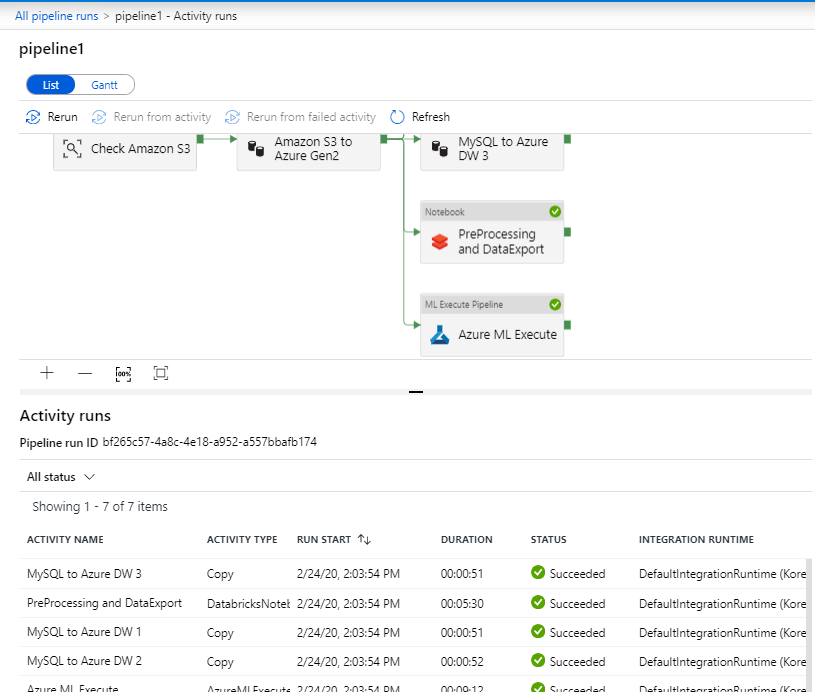
1. 최종 실행을 위해 왼쪽 상단의 **Publish All** 버튼 클릭 후 **Add triger** 의 Trigger now를 클릭합니다.



1. 왼쪽 메뉴의 Moniter 아이콘을 클릭하여 진행 상태를 확인합니다.



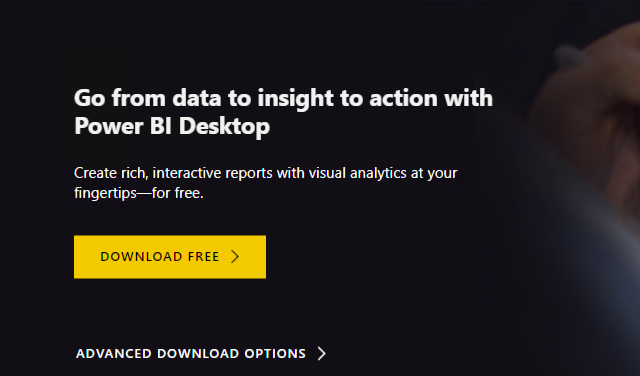
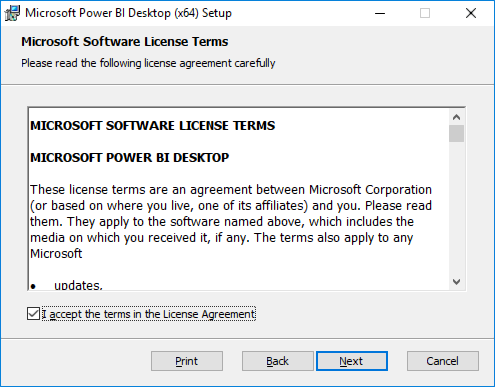
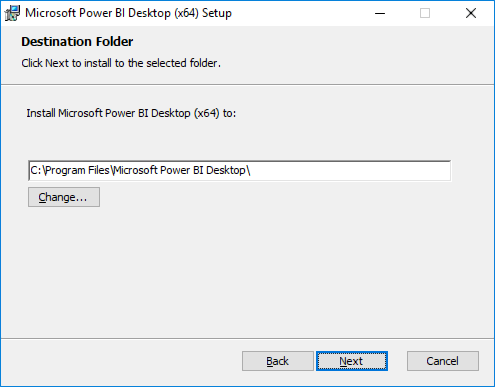
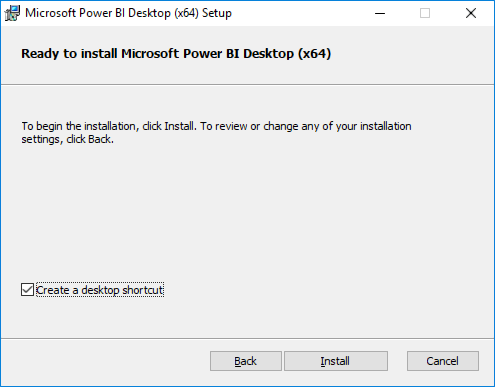
1. PIPELINE NAME을 클릭하여 상세화면으로 이동합니다.

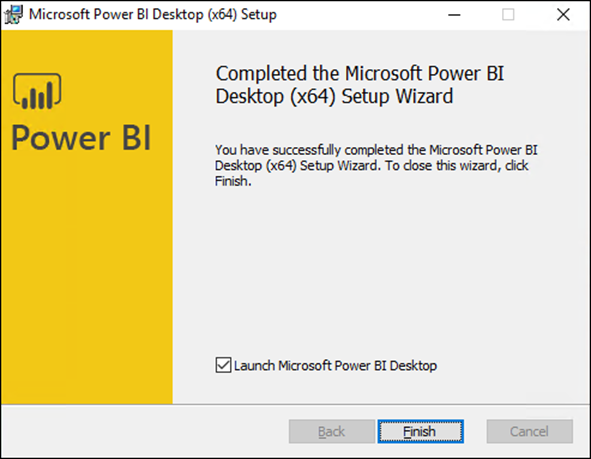


1. Exercise 3: Visualize data with Power BI Desktop

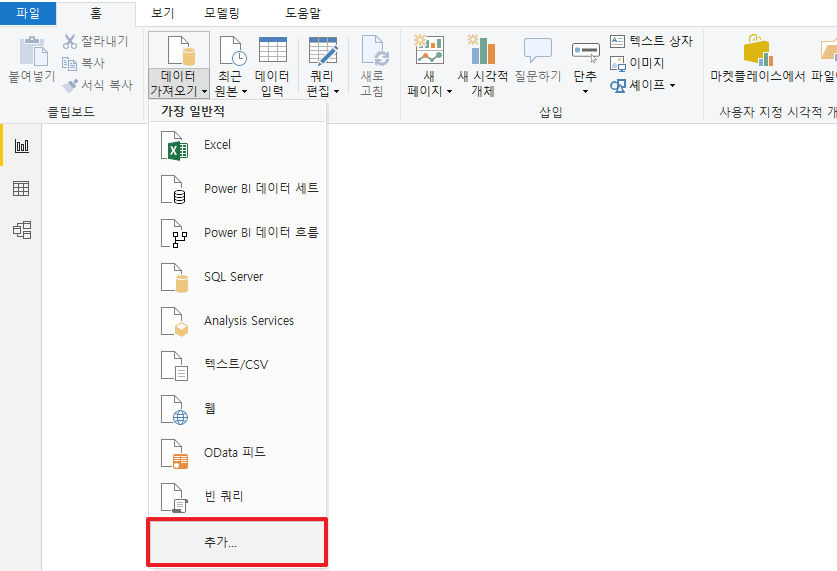
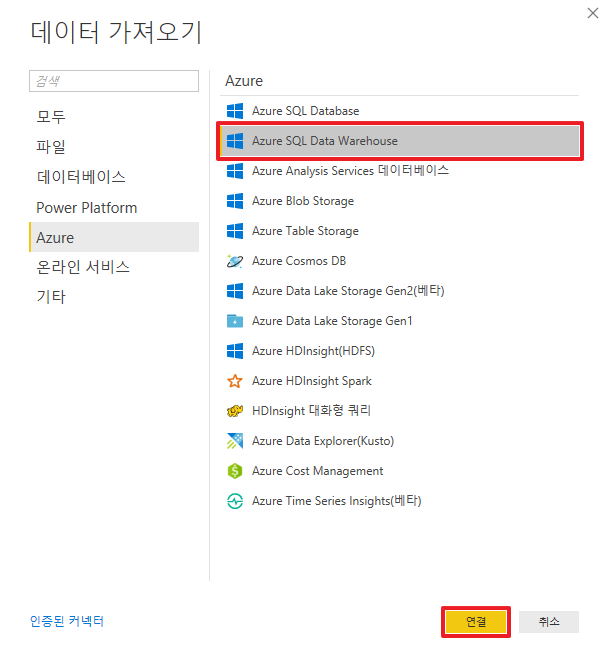
이 실습에서는 Data Warehouse의 데이터를 이용하여 Power BI Desktop과의 통합을 설정합니다.

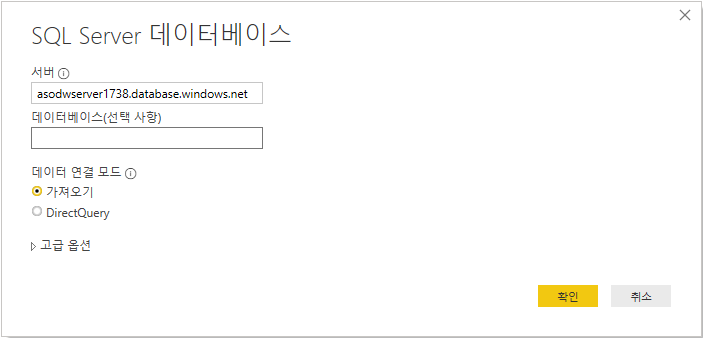
* 1. Task 1: Install Power BI Desktop

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.

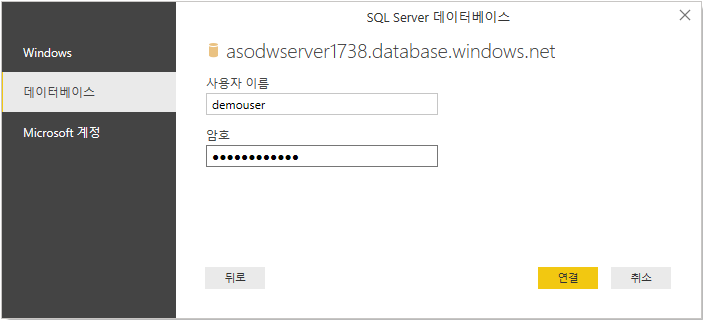


* 1. Task 2: Query data with Power BI Desktop

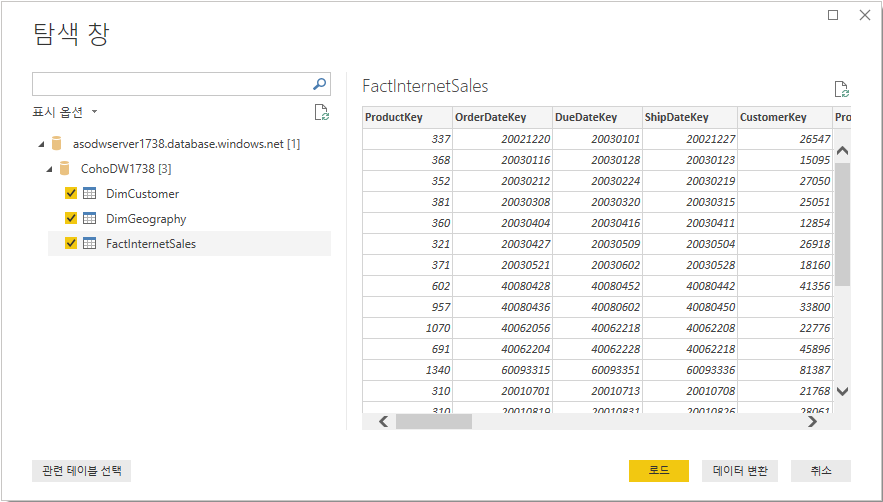
1. Power BI에서 **데이터 가져오기** 버튼을 클릭한 후 **추가...**을 선택합니다.  
   
2. 데이터 가져오기 창에서 왼쪽 목록에서 Azure를 선택합니다. 그런 다음 Azure SQL Data Warehouse를 선택하고 연결을 클릭합니다.  
   
3. SQL Server 데이터베이스 화면에서 Azure Data Warehouse 서버의 이름을 입력, 데이터 연결 모드는 가져오기를 선택한 다음 **확인**을 클릭합니다.

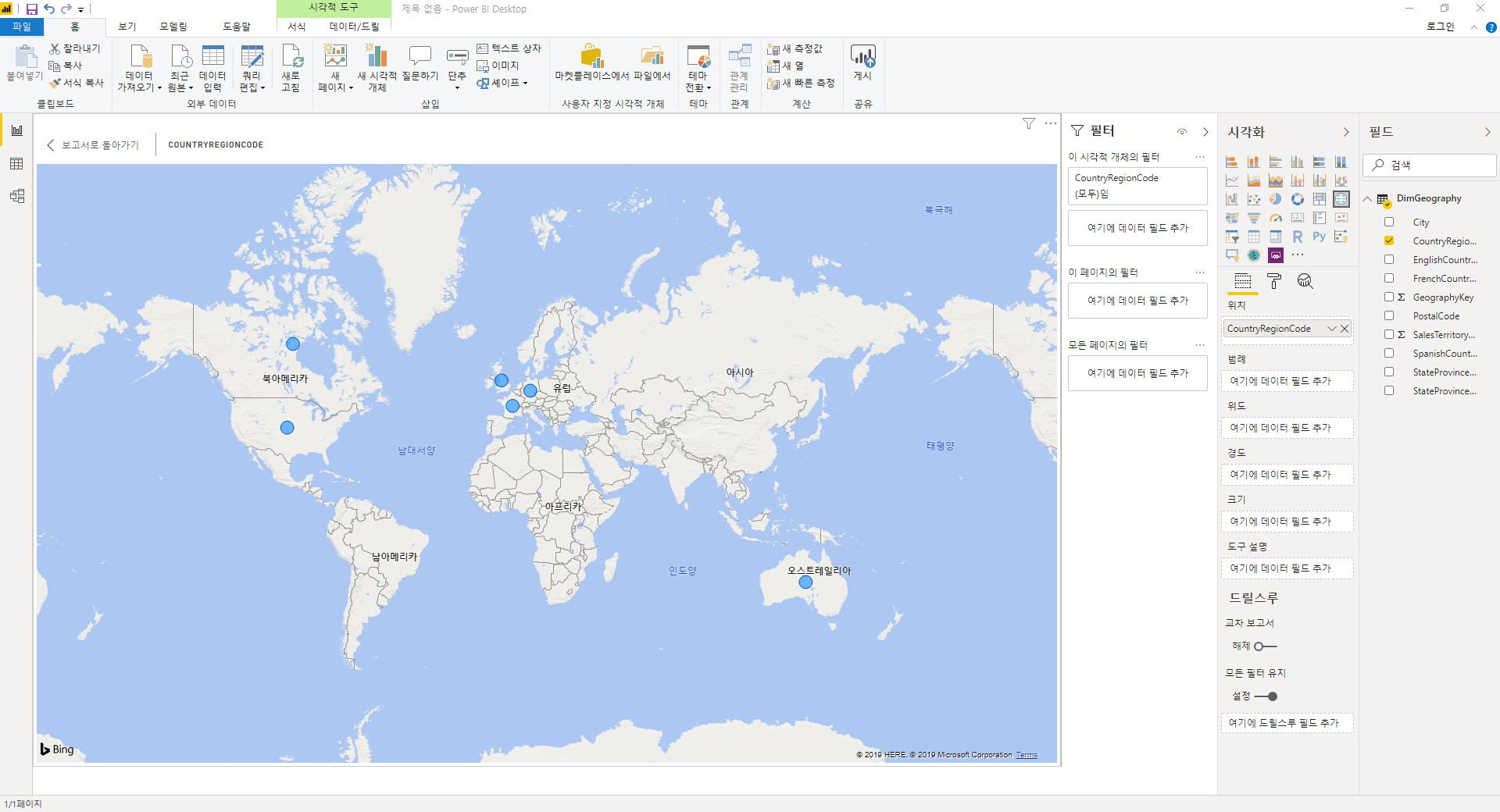
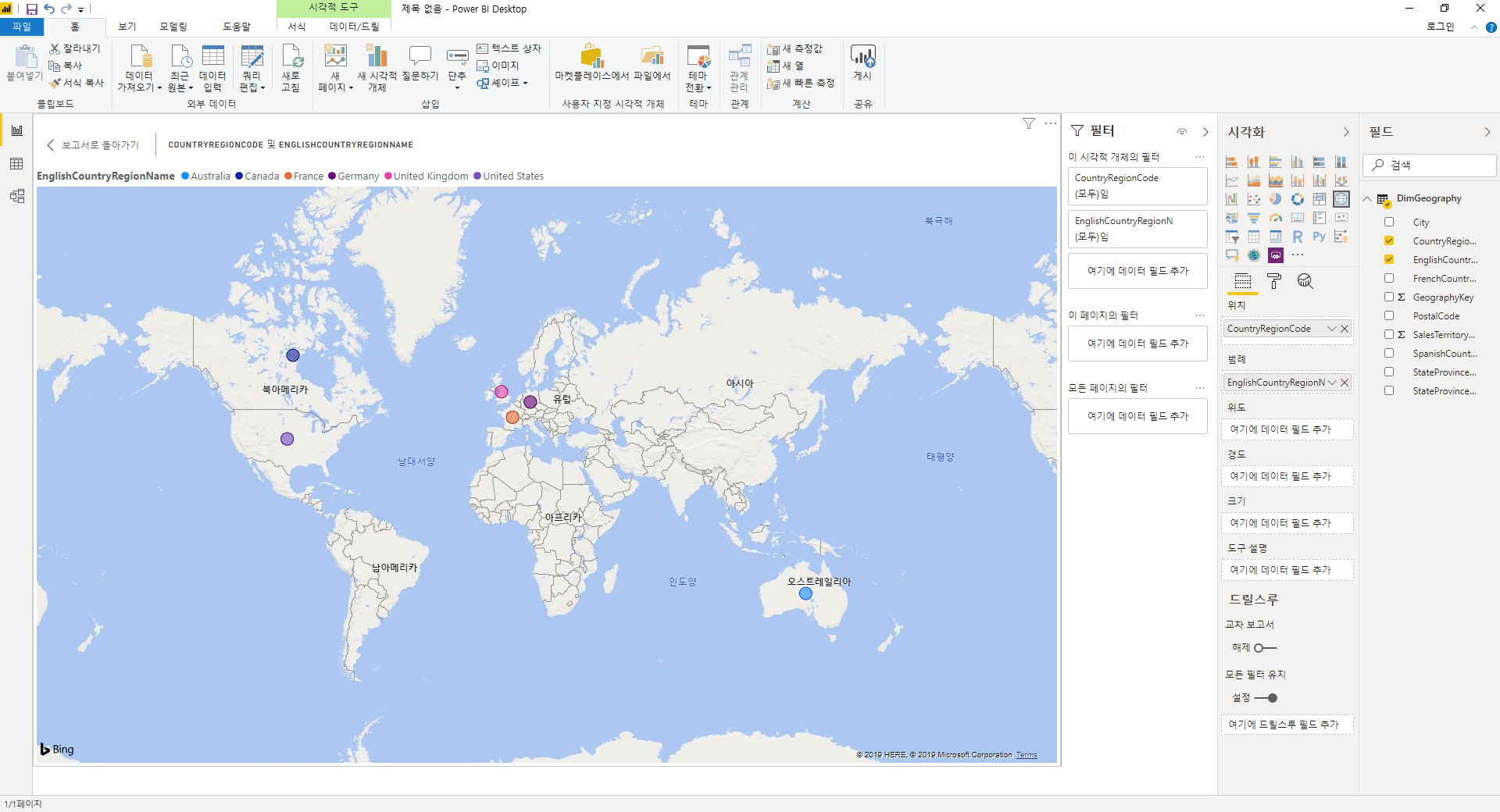
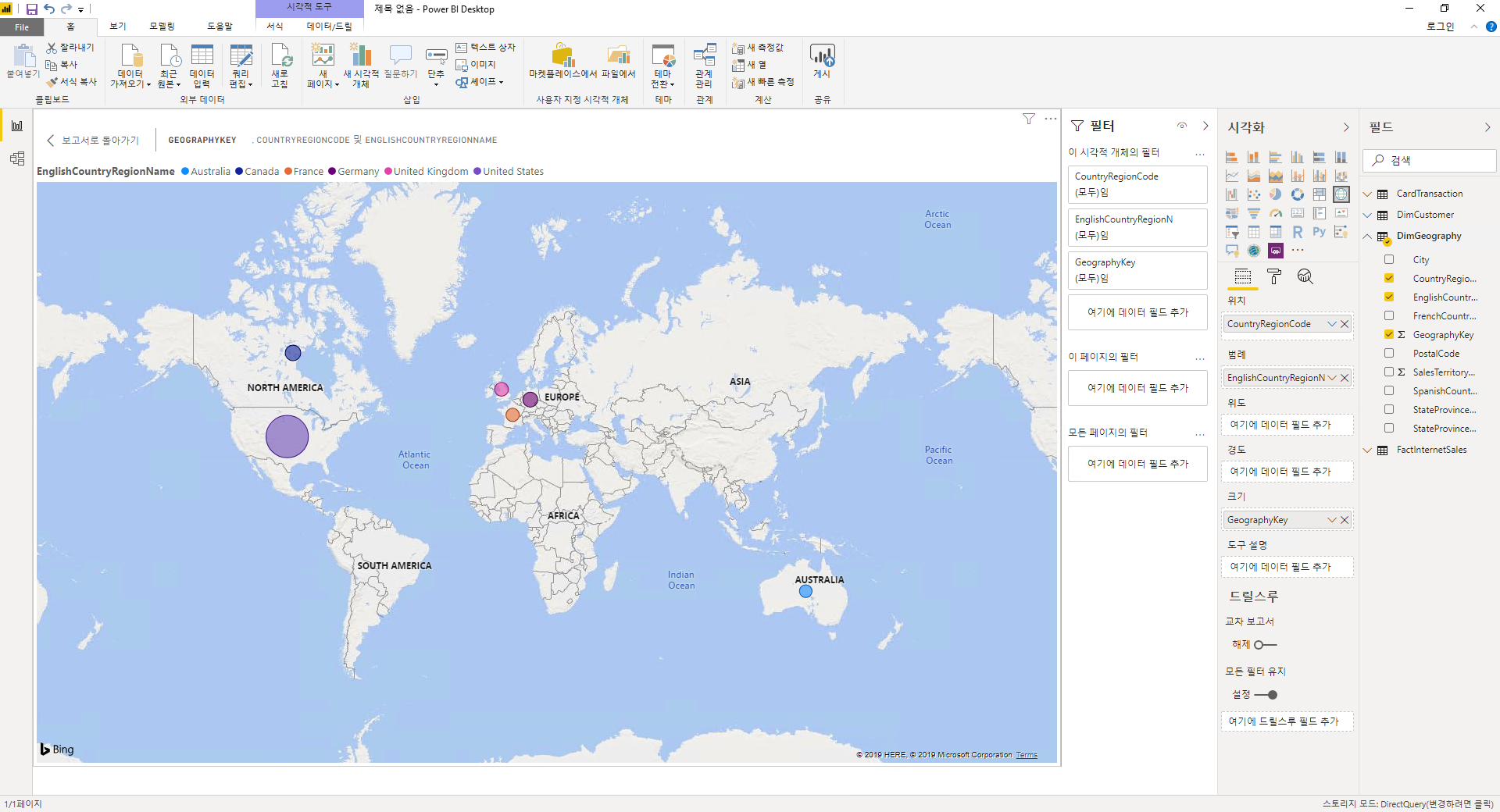


1. 데이터베이스를 선택한 후 Data Warehouse의 사용자 이름과 암호를 입력한 후 **연결**을 클릭합니다.



1. 모든 테이블을 선택한 후 **로드** 버튼을 클릭합니다.



1. 오른쪽에 있는 필드 블레이드에서 DimGeography 테이블을 확장하고 CountryRegionCode 옆에 있는 상자를 선택합니다. Power BI는 지리 데이터를 인식하고 Map 시각화가 자동으로 시작됩니다.  
   
2. 이 원이 실제로 무엇을 의미하는지에 대해 좀 더 구체적인 세부 사항을 보고 싶습니다. 국가/지역을 식별하는 범례를 추가해 봅시다. Legend 아래에서 EnglishCountryRegionName 필드를 드래그합니다.  
   
3. 원의 크기를 변경시켜 보겠습니다. FactInternetSales 테이블을 확장하고 수치를 가진 필드인 SalesAmount를 드래그해서 Map에 추가해봅니다. 

실습이 완료되었습니다. 이번 실습을 통해 온-프레미스 Data Warehouse를 Azure SQL Data Warehouse로 마이그레이션 하는 과정에 대해 알아보았습니다

* Data Warehouse 마이그레이션 계획, 사내 구축 환경의 데이터베이스에서 클라우드로의 효율적인 마이그레이션
* Azure SQL Data Warehouse의 데이터 배포, ETL 작업을 Azure Data Factory에 연결

금일 교육에 사용된 파일들은 <https://github.com/azure-datasolution/ASO> 를 통해 다운로드 받으실 수 있습니다.