



Azure Service

Hands on Lab Guide

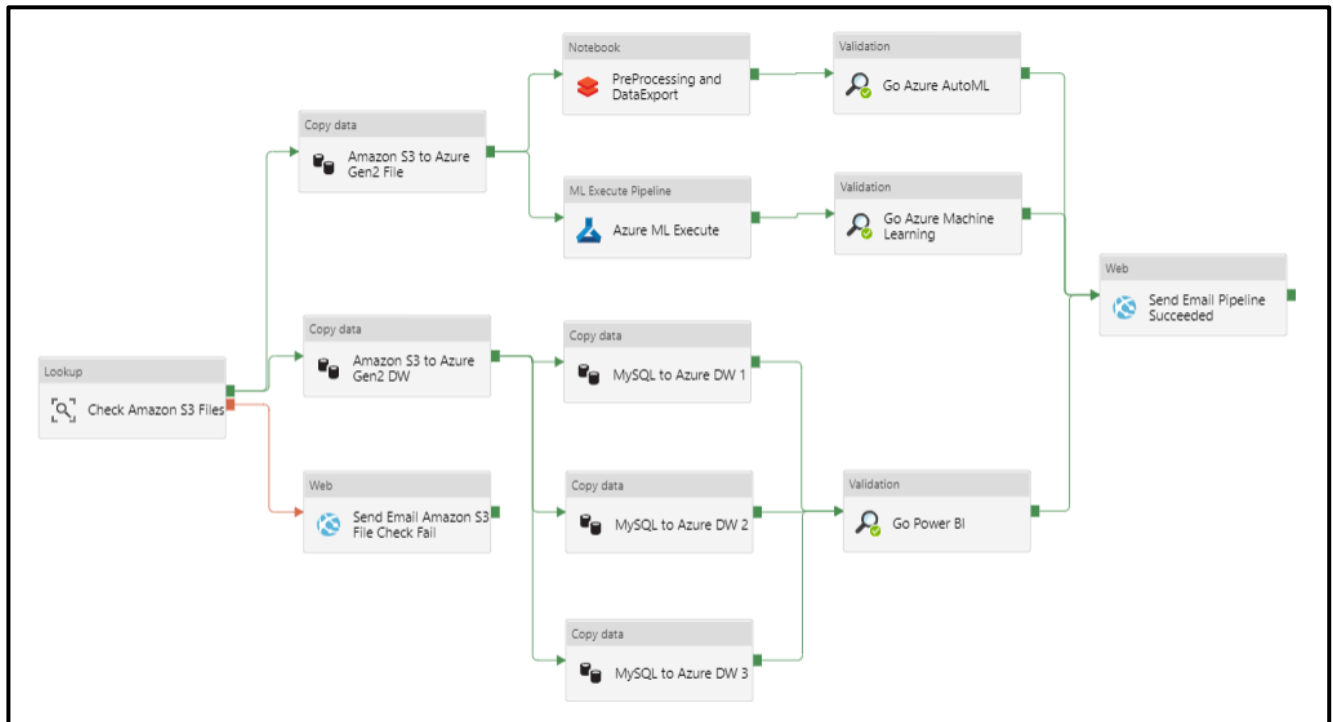
Contents

1. 사전 준비.....	2
1.1. 시나리오 개요.....	2
1.2. Azure 서비스 및 관련 제품.....	2
1.3. 필수 조건.....	2
2. 학습 목표.....	3
2.1. 솔루션 아키텍처.....	3
2.2. 대상 환경 정보.....	3
3. Exercise 1: Configure Azure Services.....	4
3.1. Task 1: Create an Azure Storage Account.....	4
3.2. Task 2: Create an Azure SQL Data Warehouse.....	8
3.3. Task 3: Prepare Azure SQL Data Warehouse.....	10
3.4. Task 4: Create an Azure Data Factory v2.....	14
4. Exercise 2: Orchestrate with Azure Data Factory.....	15
4.1. Task 1: Amazon S3 to Azure Gen2.....	15
4.2. Task 2: MySQL to Azure Synapse Analytics.....	25
4.3. Task 3: Azure ML Execute.....	43
4.4. Task 4: PreProcessing and DataExport.....	45
4.5. Task 5: Send Email.....	47
5. Exercise 3: Visualize data with Power BI Desktop.....	49
5.1. Task 1: Install Power BI Desktop.....	49
5.2. Task 2: Query data with Power BI Desktop.....	52

1. 사전 준비

1.1. 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.



1.2. Azure 서비스 및 관련 제품

- Azure Data Factory
- Azure Storage account
- Azure SQL Data Warehouse
- Azure Databricks Service
- Machine Learning
- Power BI

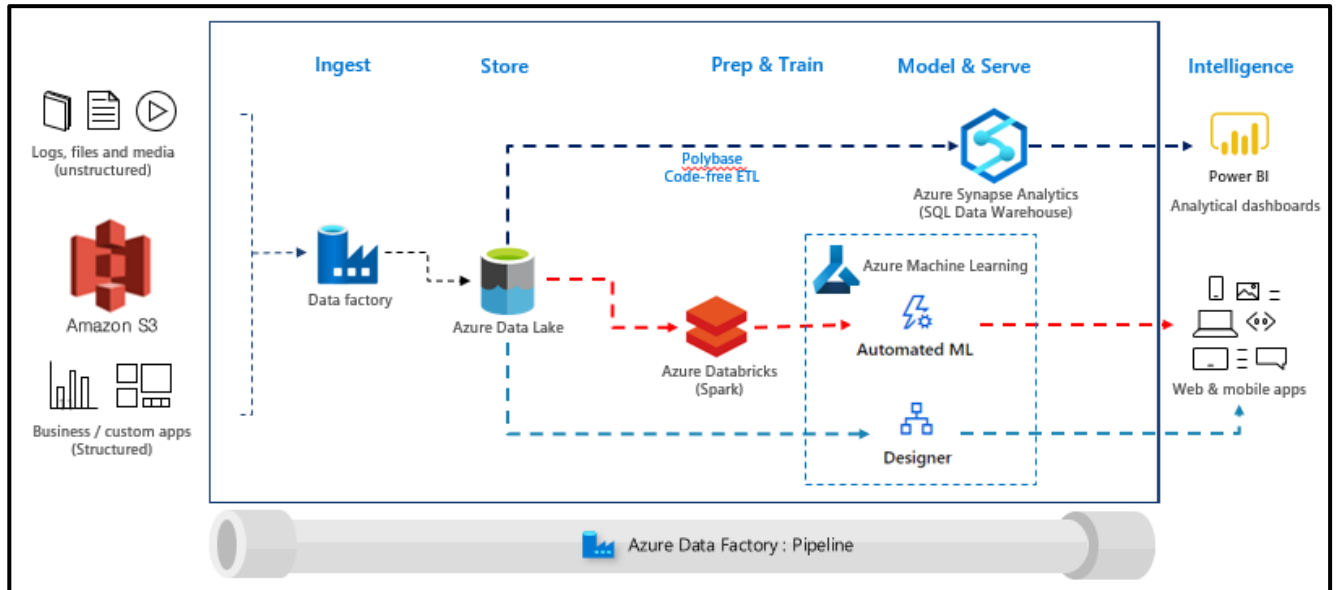
1.3. 필수 조건

- Azure Subscription

2. 학습 목표

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

2.1. 솔루션 아키텍처



2.2. 대상 환경 정보

이 연습에서는 이 실습을 위한 소스 환경을 배포합니다.

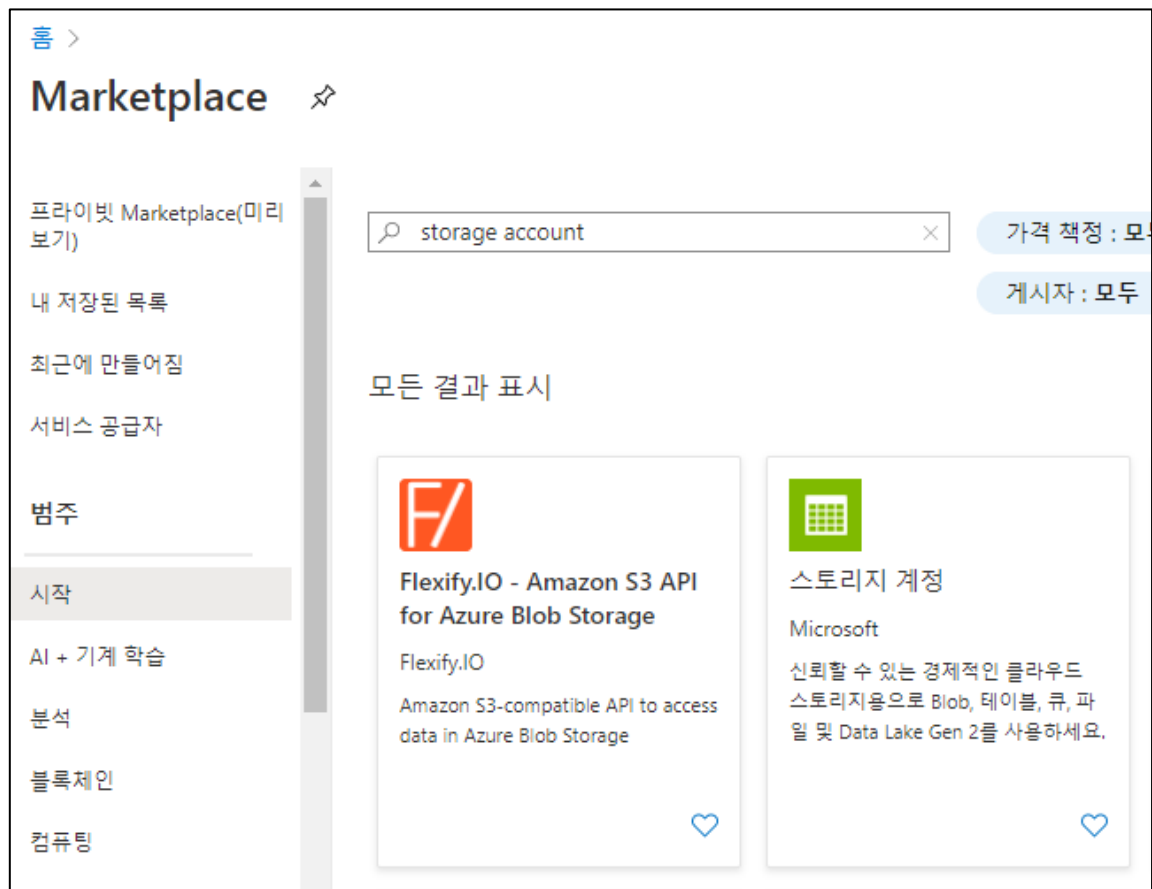
1. Azure Portal (<https://portal.azure.com>)에서 이 랩에 사용할 구독으로 로그인 했는지 확인합니다.

3. 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.

3.1. Task 1: Create an Azure Storage Account

1. Azure 포털로 이동하여 <https://portal.azure.com/>에서 로그인합니다.
2. 홈 화면 상단의 리소스 만들기를 선택하여, **storage account**를 검색하고 스토리지 계정을 선택합니다.



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

- ✓ 구독: 본인의 구독 선택
- ✓ 리소스 그룹: 새로 만들기 클릭하여 생성 -> ASO_EduXXXX
- ✓ 스토리지 계정 이름: asostorageXXXX
- ✓ 위치: (아시아 태평양)한국 중부
- ✓ 성능: 표준
- ✓ 계정 종류: StorageV2 (범용 v2)
- ✓ Replication: LRS(로컬 중복 스토리지)
- ✓ 액세스 계층: 핫

홈 > 새로 만들기 > Marketplace > Storage 계정 - Blob, File, Table, Queue > 스토리지 계정 만들기

스토리지 계정 만들기

[기본 사항](#) [네트워킹](#) [고급](#) [태그](#) [검토 + 만들기](#)

Azure Storage는 가용성, 보안, 내구성, 확장성 및 중복성이 뛰어난 클라우드 스토리지를 제공하는 Microsoft 관리 서비스입니다. Azure Storage는 Azure Blob(개체), Azure Data Lake Storage Gen2, Azure Files, Azure 큐 및 Azure 테이블을 포함합니다. 스토리지 계정의 비용은 사용량 및 아래에서 선택한 옵션에 따라 다릅니다. [Azure Storage 계정에 대한 자세한 정보](#)

프로젝트 정보

배포된 리소스와 비용을 관리할 구독을 선택합니다. 둘더 같은 리소스 그룹을 사용하여 모든 리소스를 정리 및 관리합니다.

구독 * Azure subscription 1

리소스 그룹 * ASO_Edu0000

[새로 만들기](#)

인스턴스 정보

기본 배포 모델은 최신 Azure 기능을 지원하는 Resource Manager입니다. 대신 클래식 배포 모델을 사용하여 배포하도록 선택할 수 있습니다. [클래식 배포 모델 선택](#)

스토리지 계정 이름 * ① asostorage0000 ✓

위치 * (아시아 태평양)한국 중부

성능 ① ☒ 표준 ☐ 프리미엄

계정 종류 ① StorageV2(범용 v2)

복제 ① LRS(로컬 중복 스토리지)

액세스 계층(기본값) ① ☐ 블 ☒ 핫

[검토 + 만들기](#) < 이전 다음: 네트워킹 >

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

홈 > 새로 만들기 > Marketplace > Storage 계정 - Blob, File, Table, Queue > 스토리지 계정 만들기

스토리지 계정 만들기

기본 사항 네트워킹 **고급** 태그 검토 + 만들기

보안
 보안 전송 필요 ① ☐ 사용 안 함 ☒ 사용

Azure Files
 대용량 파일 공유 ① ☒ 사용 안 함 ☐ 사용

데이터 보호
 Blob 일시 삭제 ① ☒ 사용 안 함 ☐ 사용
 ⓘ 데이터 보호 및 계층 구조 네임스페이스를 동시에 사용하도록 설정할 수 없습니다.

Data Lake Storage Gen2
 계층 구조 네임스페이스 ① ☐ 사용 안 함 ☒ 사용

NFS v3 ① ☒ 사용 안 함 ☐ 사용
 ⓘ 등록은 현재 구독별로 NFS v3 기능을 활용하는 데 필요합니다. [NFS v3 등록](#)

검토 + 만들기 < 이전 다음: 태그 >

5. 검토+만들기를 클릭하고 구성 선택 사항을 확인한 후 만들기를 선택합니다.
6. 새 스토리지 계정이 완성되면 리소스로 이동하여 컨테이너를 클릭합니다.

홈 > Microsoft.StorageAccount-20200220105915 - 개요 > asostorage0000

asostorage0000

스토리지 계정

검색(Ctrl+/) << 🔍 탐색기에서 열기 → 이동 ↺ 새로 고침 🗑️ 삭제 ❤️ 사용자 의견

개요

- 활동 로그
- 액세스 제어(IAM)
- 태그
- 문제 진단 및 해결
- 데이터 전송
- 이벤트
- Storage Explorer(미리 보기)
- 설정
- 액세스 키

리소스 그룹 (변경) : ASO_Edu0000

상태 : 기본: 사용 가능

위치 : 한국 중부

구독 (변경) : Azure subscription 1

구독 ID : 35120fde-ae11-40e4-8ea2-8be769fdcc5d

태그 (변경) : 태그를 추가하려면 여기를 클릭

컨테이너
대규모로 확장 가능한 데이터 레이크 스토리지
[자세한 정보](#)

파일 공유
서버리스 SMB 파일 공유
[자세한 정보](#)

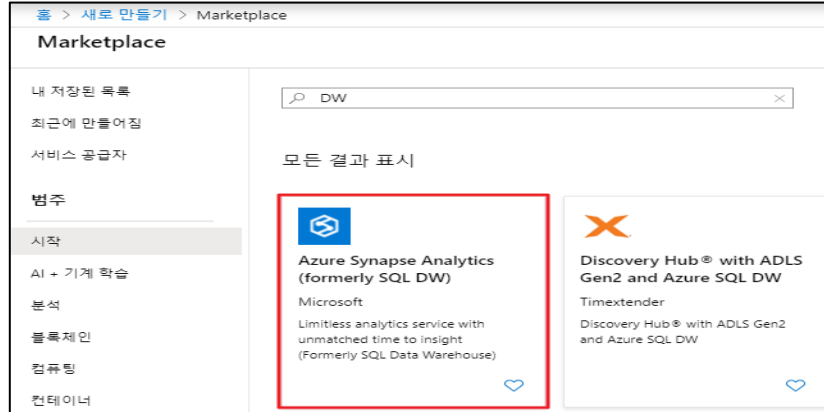
7. +컨테이너를 클릭한 후 **demo**를 입력하고 **확인** 버튼을 클릭합니다.

8. 위와 같은 방식으로 +컨테이너를 클릭 하여 **structured**를 입력하고 **확인** 버튼을 클릭합니다.

이름	마지막으로 수정한 날...	공용 액세스 ...	임대 단계
<input type="checkbox"/> demo	2020. 2. 20. 오전 11:18...	프라이빗	사용 가능 ...
<input type="checkbox"/> structured	2020. 2. 20. 오후 1:41:18	프라이빗	사용 가능 ...

3.2. Task 2: Create an Azure SQL Data Warehouse

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



3. **Azure Synapse Analytics** 블레이드에서 만들기를 클릭하고 아래 내용을 입력합니다.
 - ✓ 구독: 본인의 구독 선택.
 - ✓ 리소스 그룹: 이전에 생성한 리소스 선택 -> **ASO_EduXXXX**
 - ✓ SQL 풀 이름: **CohoDWXXXX**

Azure Synapse Analytics
Microsoft

Azure Synapse Analytics(이전 이름 Azure SQL Data Warehouse)를 시작합니다. [자세한 정보](#)

배포된 리소스와 비용을 관리할 구독을 선택합니다. 폴더 같은 리소스 그룹을 사용하여 모든 리소스를 정리 및 관리합니다.

구독 *

Azure subscription 1

리소스 그룹 *

ASO_Edu0000

[새로 만들기](#)

SQL 풀 정보

논리 서버 선택, 성능 수준 구성 등을 포함하여 이 SQL 풀의 필수 설정을 입력합니다.

SQL 풀 이름 *

CohoDW0000

서버 *

(새로 만들기) asodwserver0000 (한국 중부)

[새로 만들기](#)

성능 수준 *

Gen2

DW100c

[성능 수준 선택](#)

검토 + 만들기

다음: 추가 설정 >

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

✓ 서버 이름: **asodwserverXXXX** (DW 계정 명)

✓ 서버 관리자 로그인: **demouser**

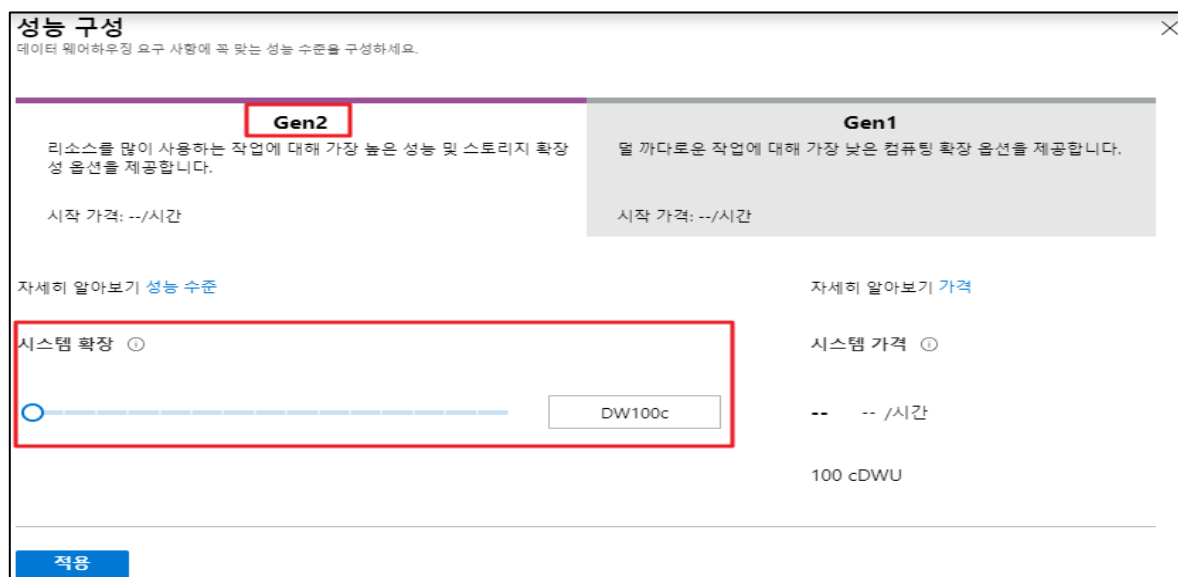
✓ 암호: **Demo@pass123**

✓ 위치: **(아시아 태평양)한국 중부**

✓ Azure 서비스의 서버 액세스 허용: **체크함**



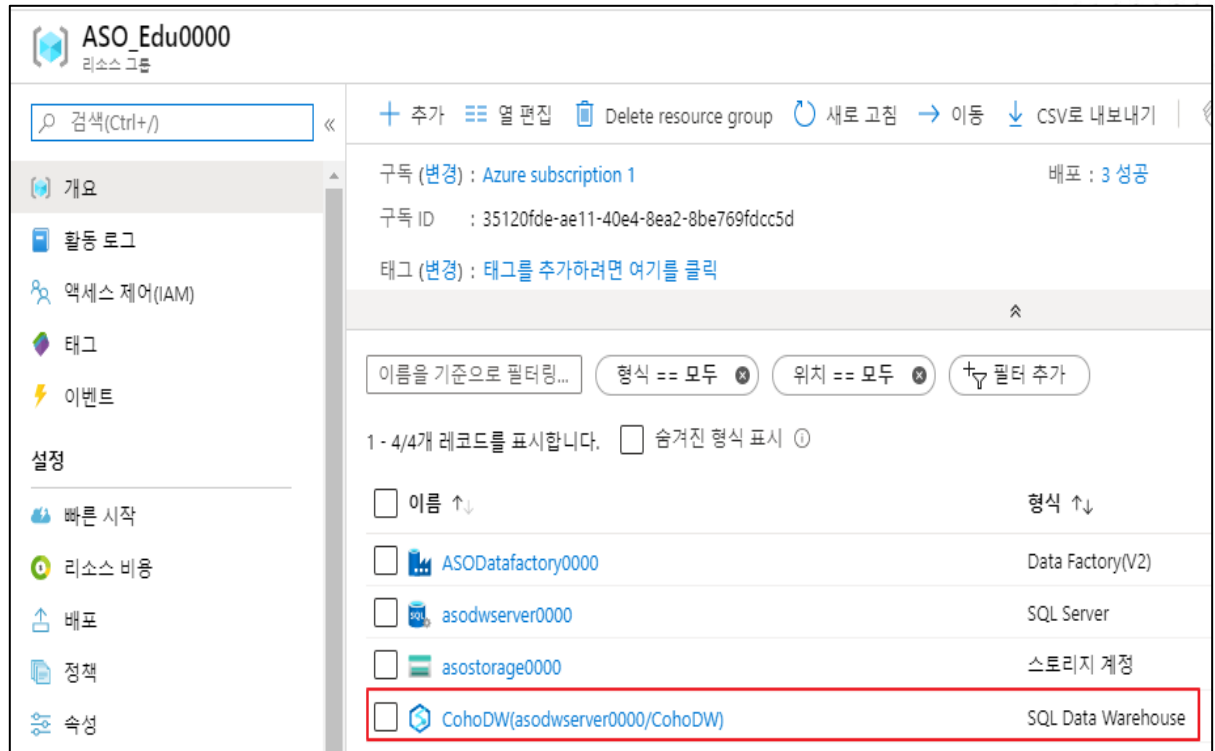
5. 성능 수준 선택을 클릭하고 Gen2 탭을 선택한 다음 성능을 DW100c로 설정하고 적용을 클릭합니다.



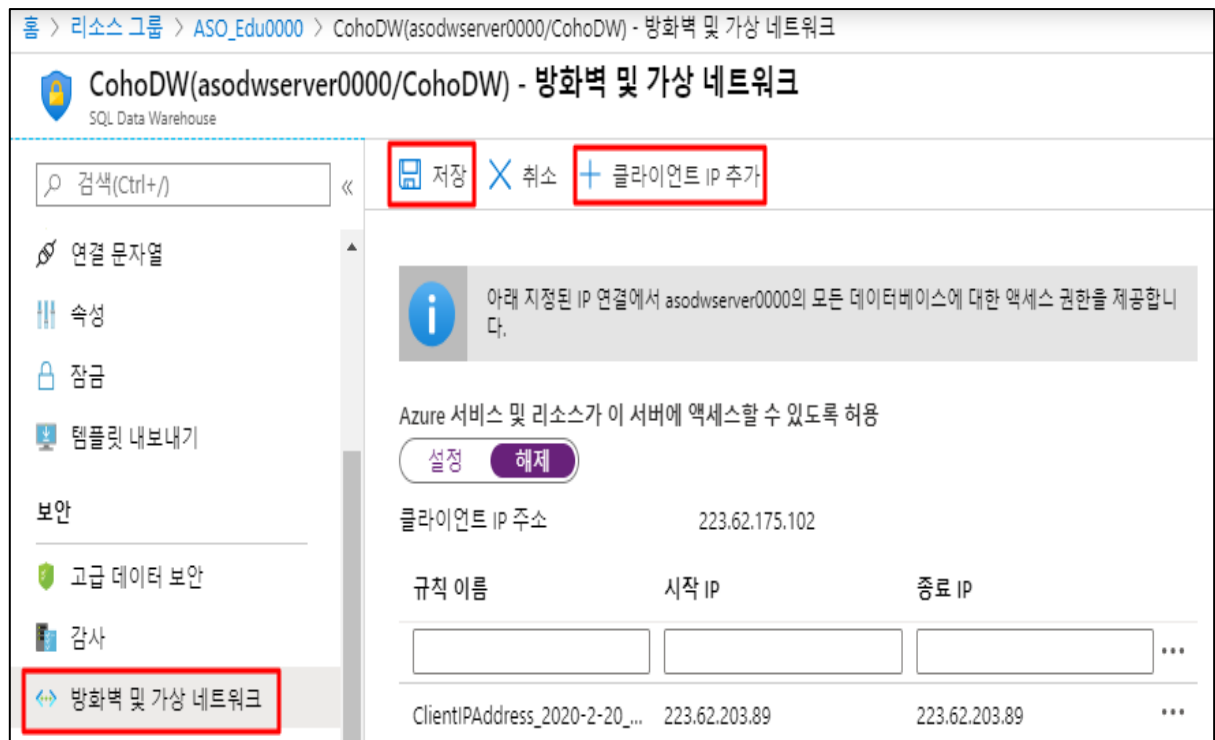
6. SQL Data Warehouse 블레이드에서 검토+만들기를 클릭한 후 만들기를 클릭합니다.

3.3. Task 3: Prepare Azure SQL Data Warehouse

1. Azure 포털로 이동하여 <https://portal.azure.com/>에서 로그인합니다.
2. 리소스 그룹으로 이동하여 **Data Warehouse (CohoDW)**를 선택합니다..



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



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

✓ 로그인: **demouser**

✓ 암호: **Demo@pass123**

홈 > 리소스 그룹 > ASO_Edu0000 > CohoDW(asodwserver0000/CohoDW) - 쿼리 편집기(미리 보기)

CohoDW(asodwserver0000/CohoDW) - 쿼리 편집기(미리 보기)
SQL Data Warehouse

검색(Ctrl+F) << 로그인 + 새 쿼리 ↑ 쿼리 열기 피드백

보안

- 고급 데이터 보안
- 감사
- 방화벽 및 가상 네트워크
- 투명한 데이터 암호화

일반 작업

- 스트리밍 작업 보기
- 데이터 로드
- 쿼리 편집기(미리 보기)**
- 대시보드 + 보고서 작성
- 데이터 모델링 및 캐시
- Visual Studio에서 열기

모니터링

- 쿼리 작업

SQL

SQL 데이터베이스 쿼리 편집기 시작

SQL Server 인증

로그인 *

demouser

암호 *

확인

Active Directory 인증

OR

5. 다음 쿼리를 실행하여 테이블을 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);

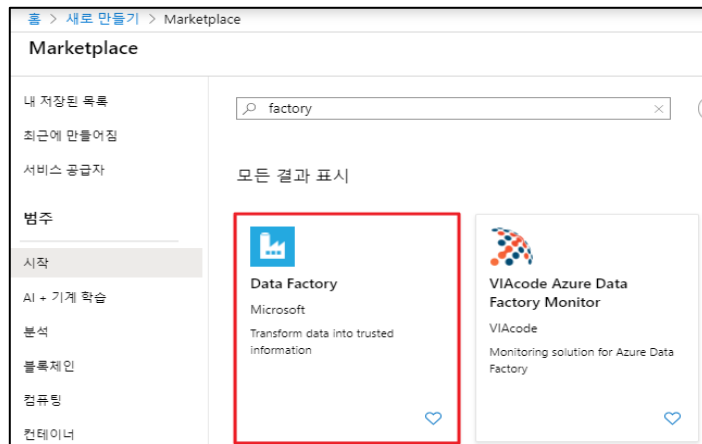
CREATE TABLE [dbo].[DimGeography](
    [GeographyKey] int not null, [City] nvarchar(60) null,
    [StateProvinceCode] nvarchar(6) null,
    [StateProvinceName] nvarchar(100) null,
    [CountryRegionCode] nvarchar(6) null,
    [EnglishCountryRegionName] nvarchar(100) null,
    [SpanishCountryRegionName] 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)

```


3.4. Task 4: Create an Azure Data Factory v2

1. Azure 포털로 이동하여 <https://portal.azure.com/>에서 로그인합니다.
2. 리소스 만들기를 클릭하여 **Factory**를 입력, 검색하고 **Data Factory**를 선택합니다.



3. Data Factory 블레이드에서 만들기를 클릭한 후 다음 정보를 입력합니다.

- ✓ 구독: 본인의 구독 선택
- ✓ 리소스 그룹: **ASO_EduXXXX**
- ✓ 지역: **한국 중부**
- ✓ 이름: **ASODatafactoryXXXX** (XXXX는 사용자 별로 고유하게 입력)
- ✓ 버전: **V2**

4. 다음: **Git configuration** 버튼 클릭 후 **Configure Git later**에 체크하고 **검토+만들기** 버튼을 클릭합니다.

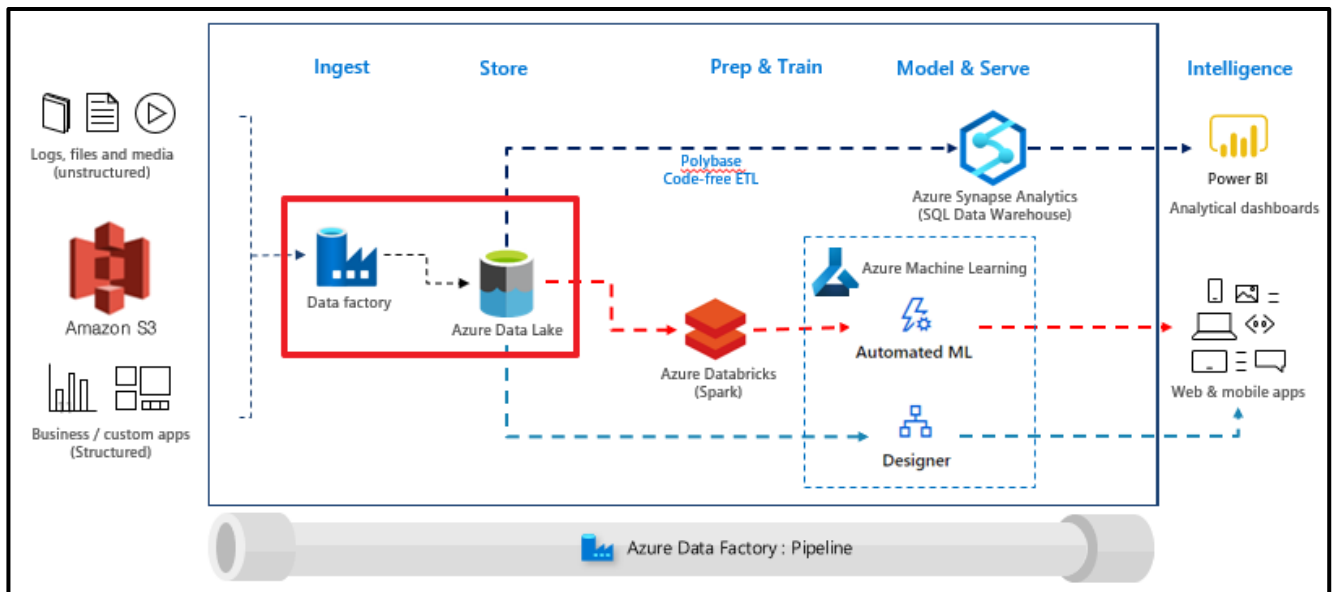
Data Factory 위치는 Data Factory의 메타데이터가 저장되고 파이프라인의 트리거가 시작되는 위치입니다. 한편, Data Factory는 다른 Azure 영역의 데이터 저장소 및 컴퓨팅 서비스에 액세스하여 데이터 저장소 간에 데이터를 이동하거나 컴퓨팅 서비스를 사용하여 데이터를 처리할 수 있습니다. 이 동작은 전 세계적으로 사용 가능한 통합 런타임을 통해 실현되어 데이터 컴플라이언스, 효율성 및 네트워크 송신 비용 절감을 보장합니다.

4. Exercise 2: Orchestrate with Azure Data Factory

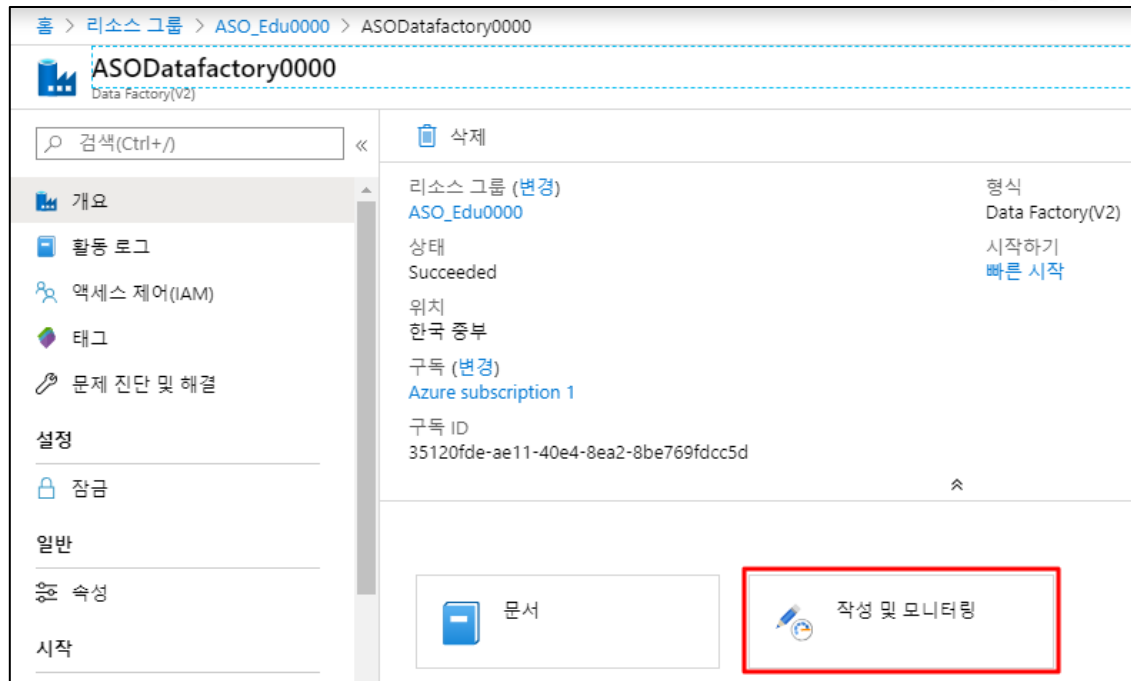
이 실습에서는 Azure Data Factory의 Pipeline을 이용하여 특정 위치의 파일의 데이터를 이관하며, Azure Databricks, Azure Machine Learning 등 Azure의 다른 서비스들과 연결하여 전체적인 데이터의 흐름을 관리합니다.

4.1. Task 1: Amazon S3 to Azure Gen2

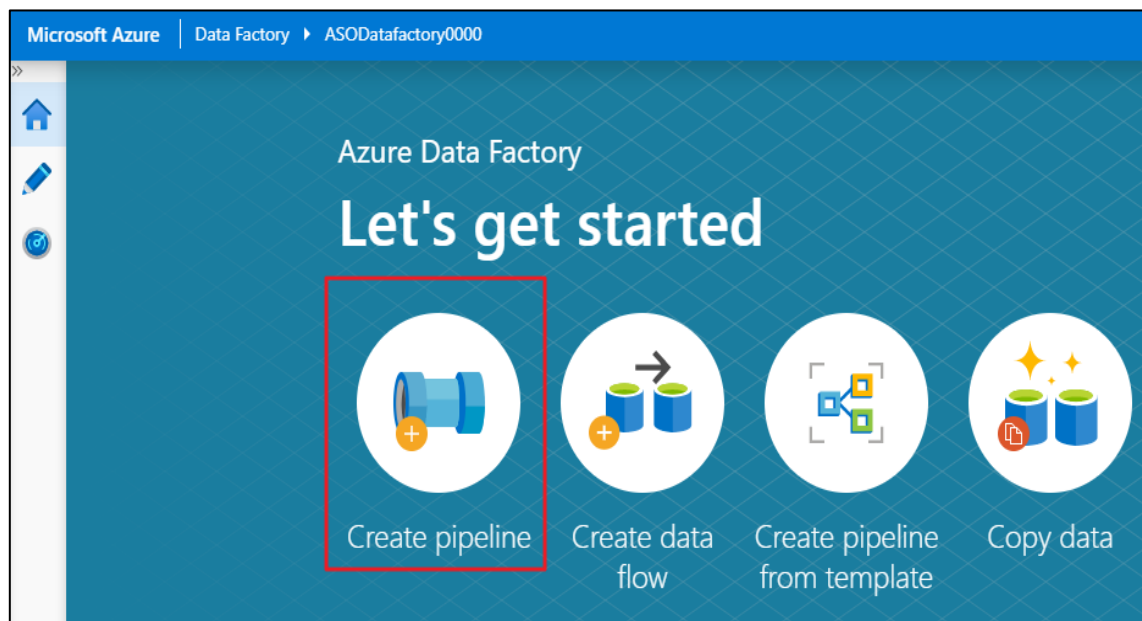
이 작업은 Amazon S3에 파일이 존재하는지 확인한 후 해당파일을 Azure 서비스에서 사용하기 위해 Azure Data Lake Store Gen2로 복사합니다.



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

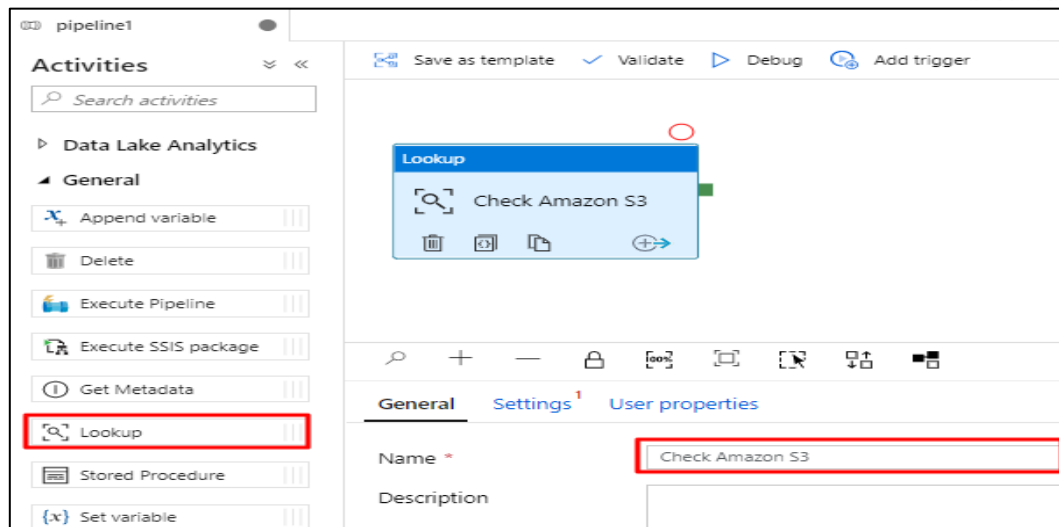


2. Data Factory 포털 홈 화면에서 **Create pipeline**을 선택합니다.

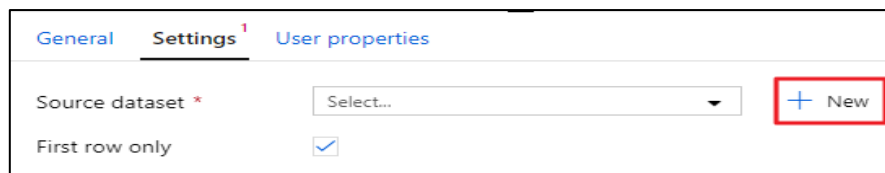


3. **Activities > General > Lookup**을 드래그해서 옮긴 후 하단 **General** 탭에서 다음을 입력합니다.

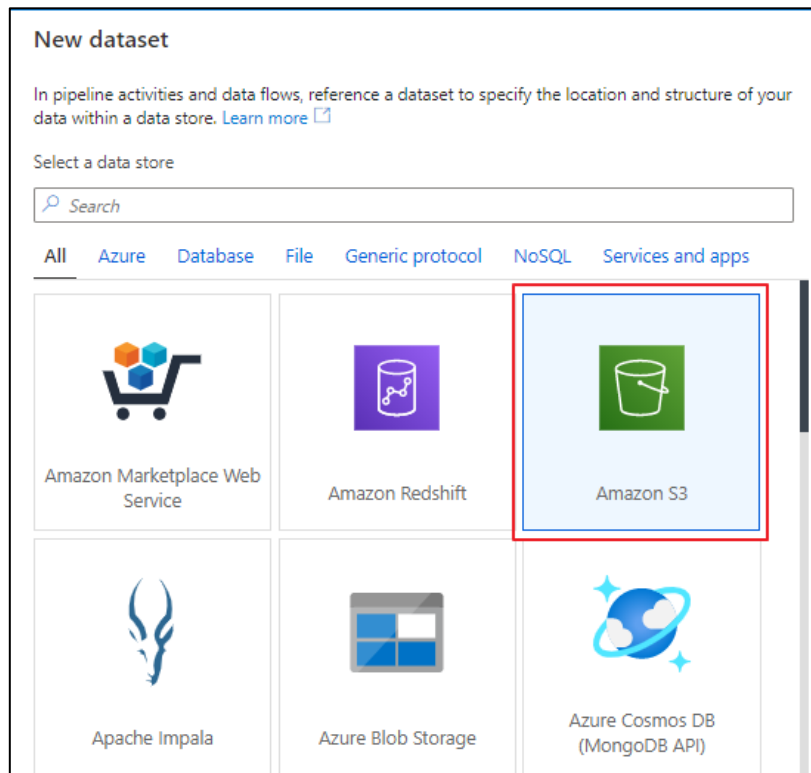
✓ Name: **Check Amazon S3**



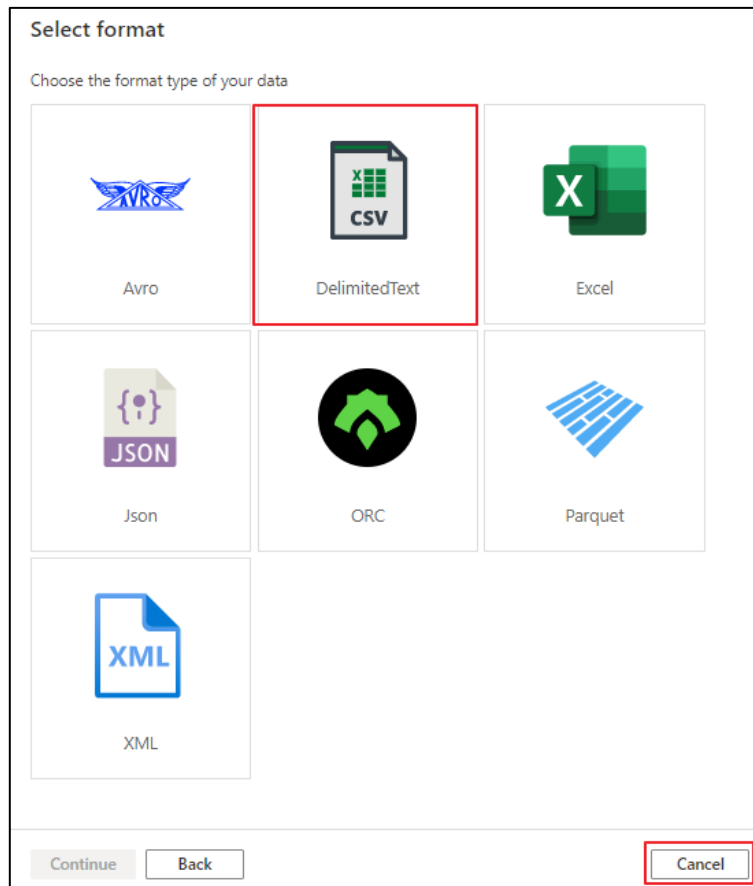
4. **Setting** 탭으로 이동하여 Source dataset의 **+New**를 클릭합니다.



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

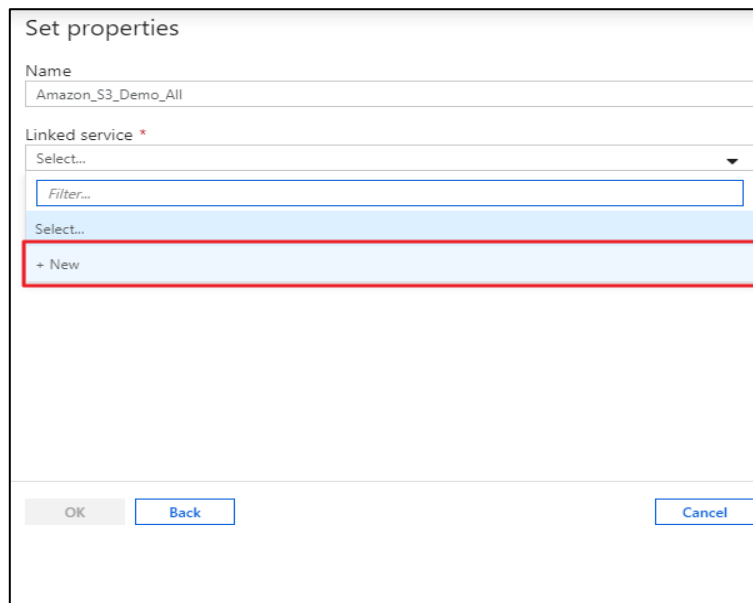


6. **format**은 **DelimitedText**를 선택합니다.



7. Name 항목 입력 후 **Linked service** 에서 **+New**를 선택합니다.

✓ Name : **Amazon_S3_Demo_All**



8. **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 : **AKIAI54HTGVLJRJ4TC7QQ**
- ✓ Secret Access Key : **OOr4sejC3lCrmT7Nuv4RaOezH7UDatbh4sl+ztgx**

New linked service (Amazon S3)

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

Name *
AmazonS3

Description

Connect via integration runtime *
AutoResolveIntegrationRuntime

Access Key ID *
AKIAI54HTGVLJRJ4TC7QQ

Secret Access Key *

Service URL

Test connection
☒ To linked service ☐ To file path

Annotations

Create **Test connection** **Cancel**

9. **Set Properties** 블레이드가 열리면 **File path** 항목의 **Browse** 버튼을 클릭하여 파일 경로를 **asodemo > Demo_All**로 지정 후 **OK** 버튼을 클릭하여 완료합니다.

- ✓ Name : **Amazon_S3_Demo_All**
- ✓ Linked service: **AmazonS3**
- ✓ File path : **asodemo > Demo_All**

Set properties

Name
Amazon_S3_Demo_All

Linked service *
AmazonS3

[Edit connection](#)

File path
asodemo / Demo_All / File **Browse**

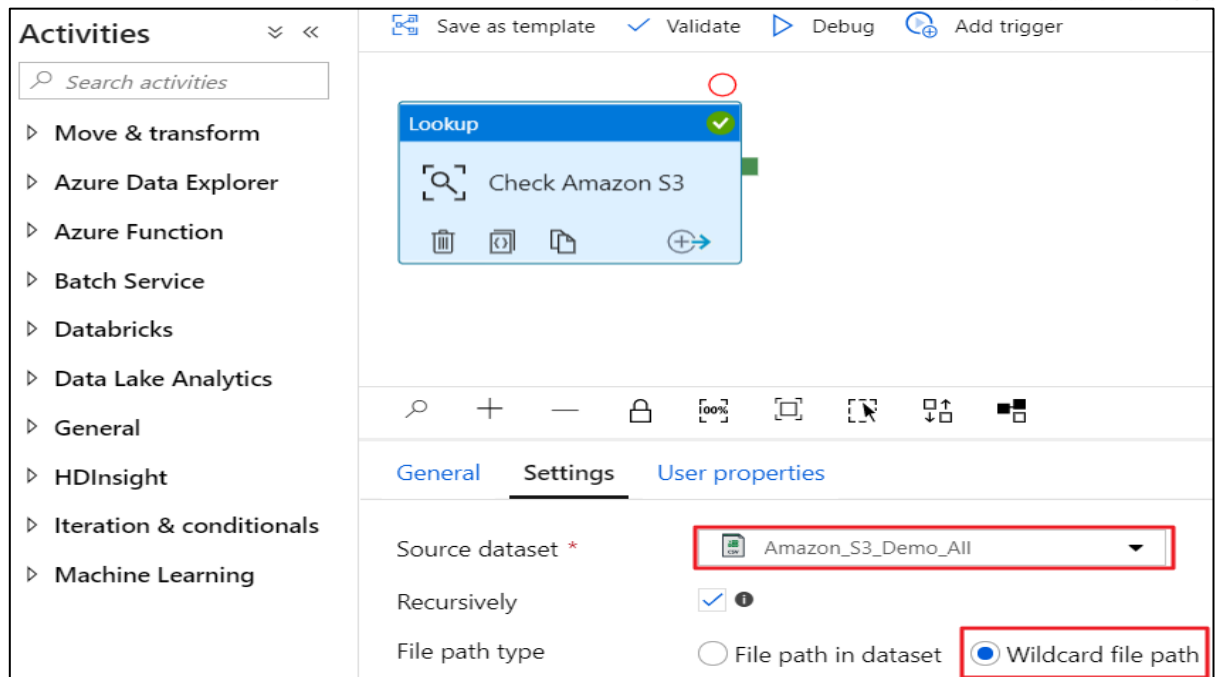
First row as header ☐

Import schema
☒ From connection/store ☐ From sample file ☐ None

Advanced

OK **Back** **Cancel**

10. **Lookup** 액티비티의 **Setting**탭의 **Source Dataset**이 세팅된 것을 확인 후 **File path Type** 항목의 **Wildcard file path**를 선택합니다.



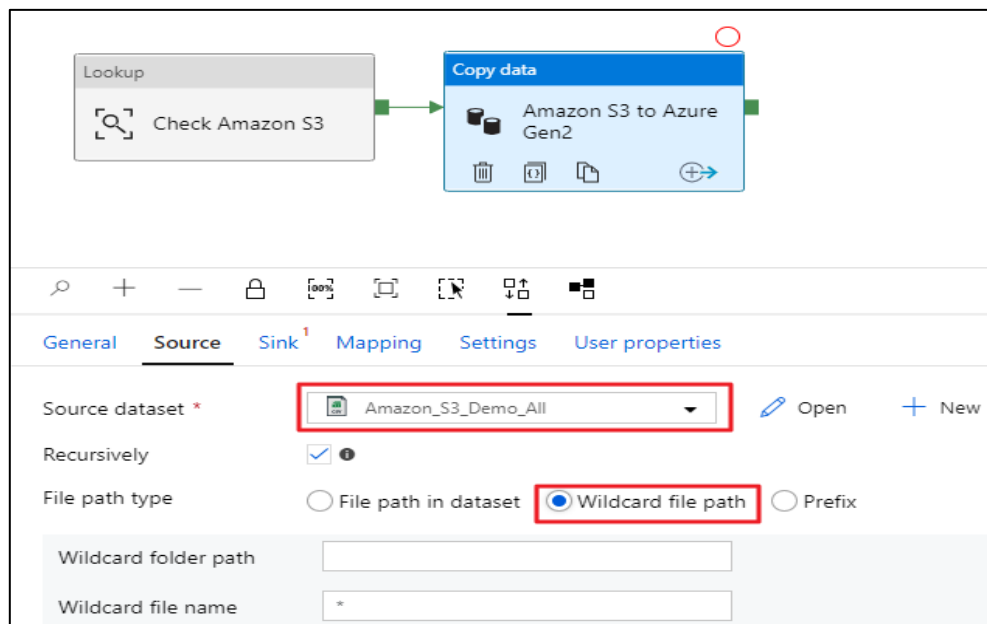
11. **Activities** > **Move & transform** > **Copy data**를 드래그해서 옮긴 후 하단 **General** 탭에서 다음을 입력합니다.

✓ Name: **Amazon S3 to Azure Gen2**

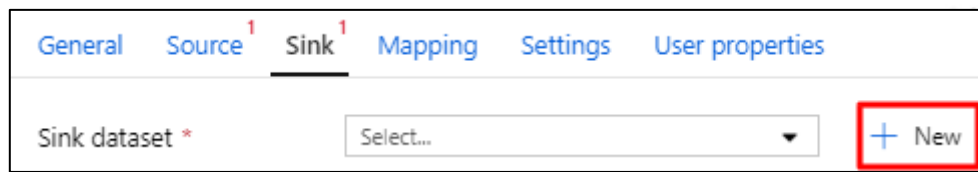
12. **Lookup** 액티비티와 **Copy data** 액티비티를 연결시킨 후 **Source**탭으로 이동하여 다음을 입력합니다.

✓ Source dataset : **Amazon_S3_Demo_All**

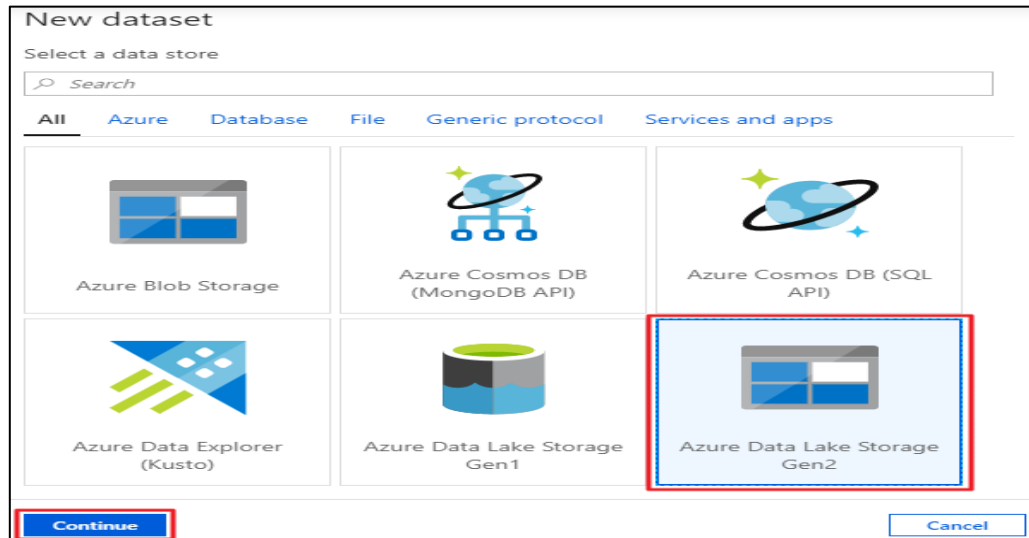
✓ File path Type : **Wildcard file path**



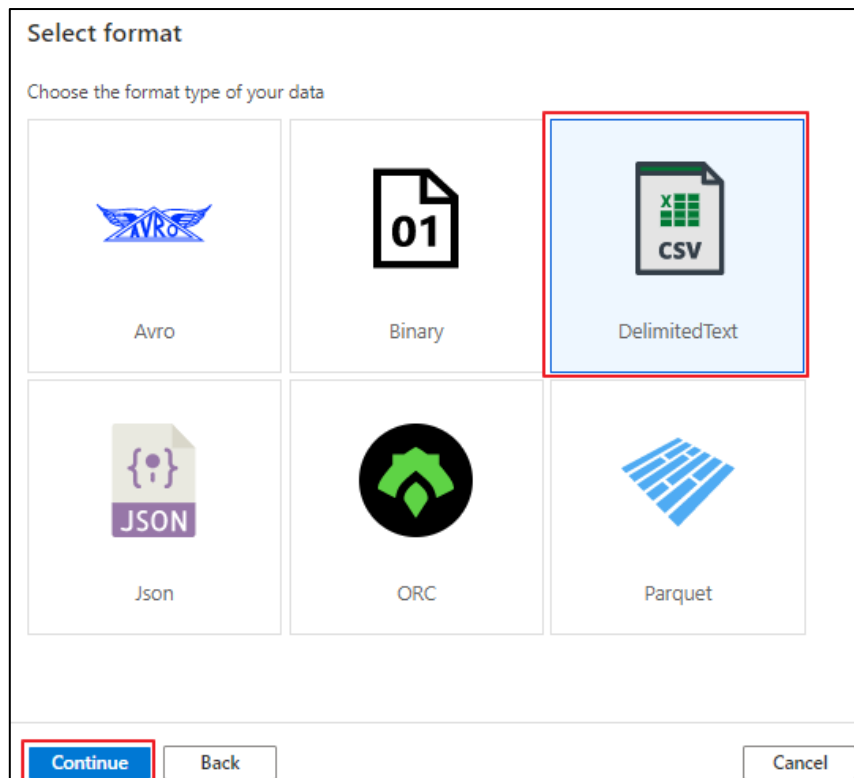
13. **Sink** 탭으로 이동하여 Sink dataset의 **+New**를 클릭합니다.



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

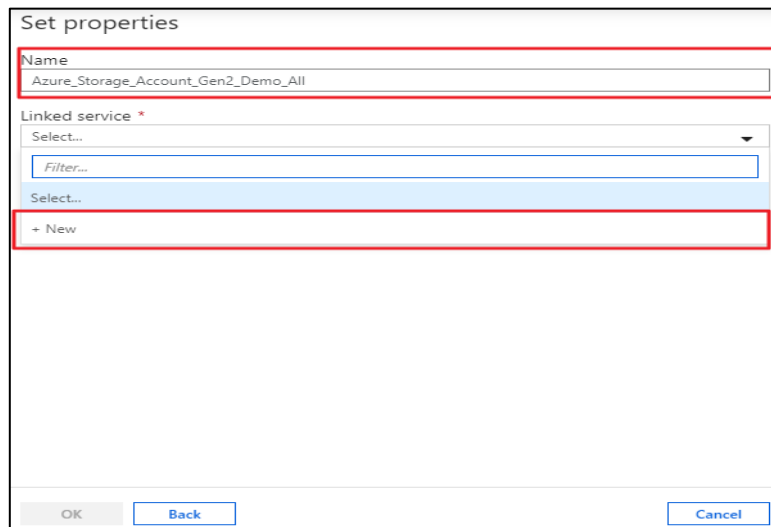


15. format은 **DelimitedText**를 선택합니다.



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

✓ Name: **Azure_Storage_Account_Gen2_Demo_All**

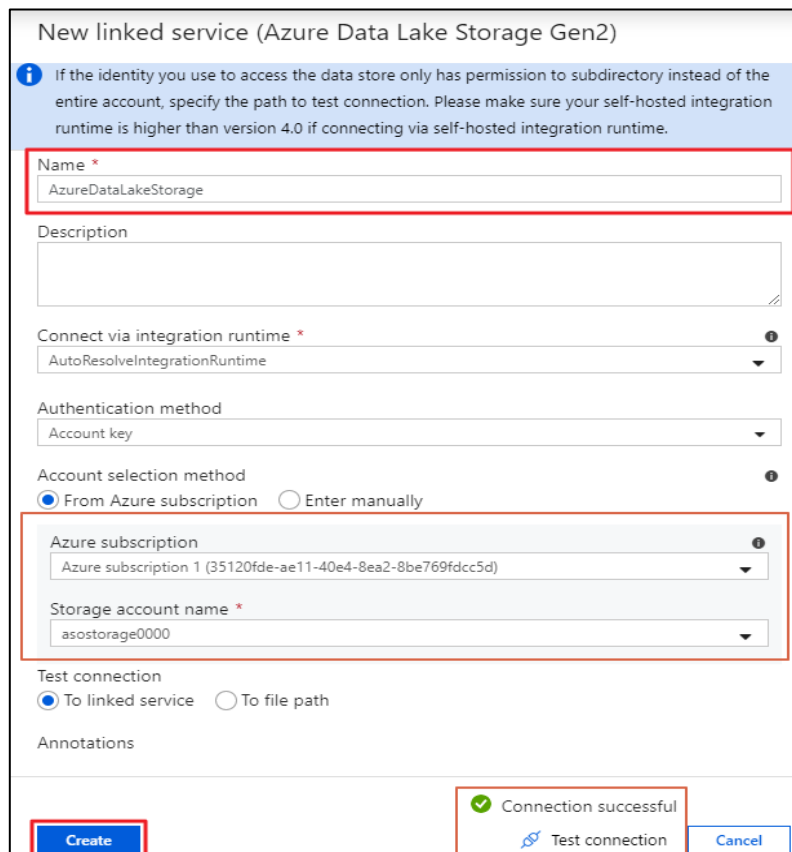


17. **New linked service (Azure Data Lake Storage Gen2)** 블레이드가 열리면, 아래와 같이 정보를 입력하고 **Test Connection** 버튼을 클릭하여 Connection successful 메시지가 출력되면 **Create** 버튼을 클릭합니다.

✓ Name : **AzureDataLakeStorage**

✓ Azure subscription: 본인의 구독

✓ Storage account name: **asostorageXXXX**



18. **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)

Set properties

Name
Azure_Storage_Account_Gen2_Demo_All

Linked service *
AzureDataLakeStorage

Edit connection

File path
demo / Directory / File **Browse**

First row as header ☐

Import schema
☒ From connection/store ☐ From sample file ☐ None

Advanced

OK Back Cancel

19. Copy data 액티비티의 **Sink** 탭의 Sink Dataset이 세팅된 것을 확인 후 상단의 **validate** 버튼 클릭 후 **Debug** 버튼을 클릭하여 실행합니다.

pipeline1

Activities

Save as template **Validate** Debug Add trigger

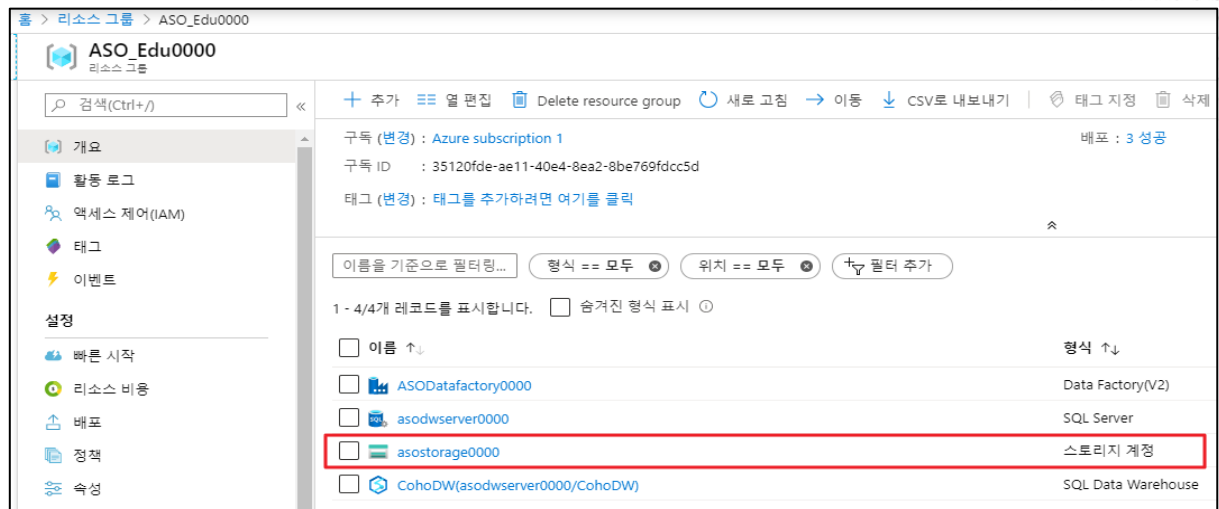
Lookup Check Amazon S3

Copy data Amazon S3 to Azure Gen2

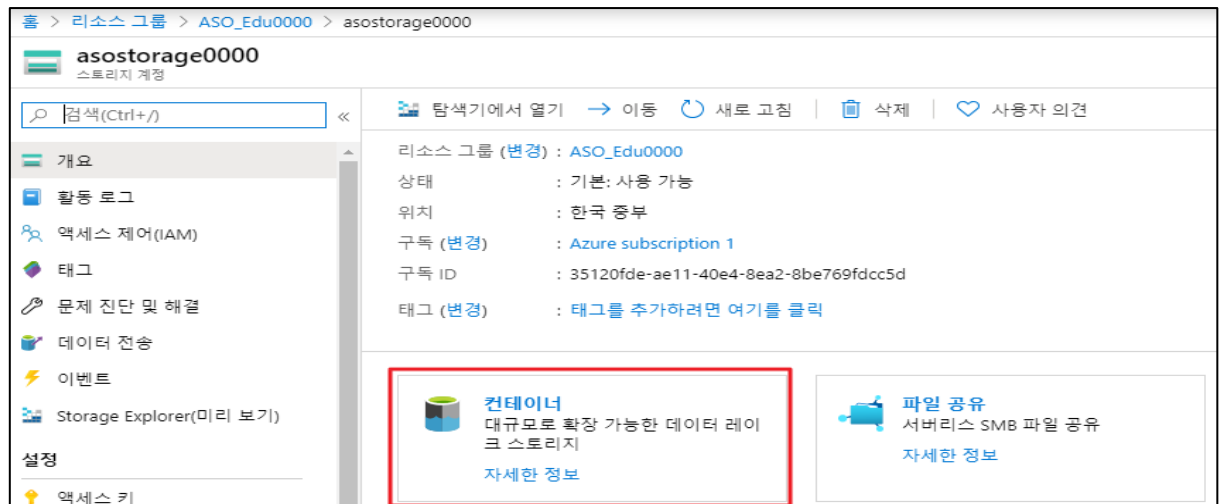
General Source **Sink** Mapping Settings User properties

Sink dataset * **Azure_Storage_Account_Gen2_Demo_All** Open New

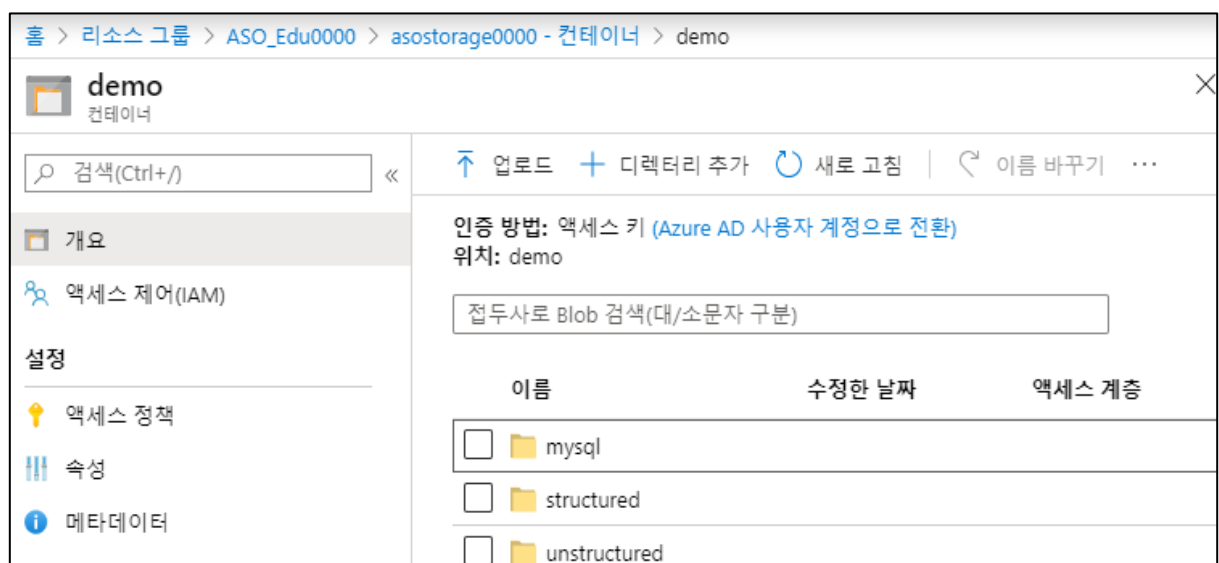
20. 실행된 결과를 확인하기 위해 Azure Portal 의 리소스그룹으로 이동하여 스토리지 계정을 선택합니다.



21. 컨테이너 버튼을 클릭합니다.

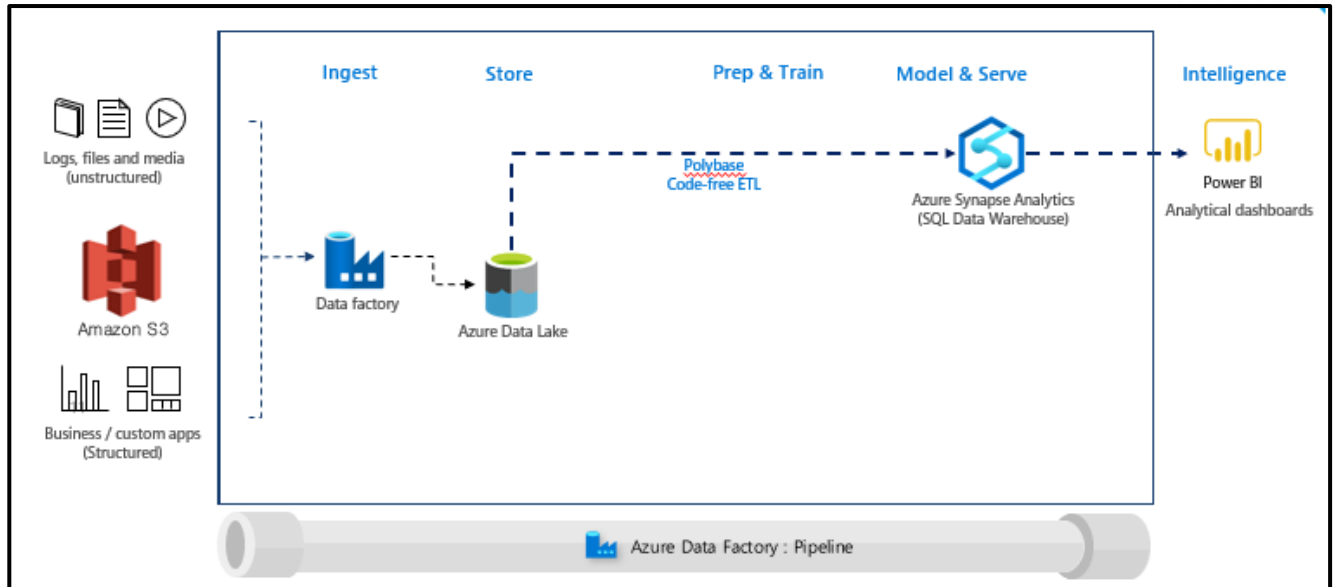


22. 컨테이너 목록에서 **demo**를 선택하여 아래와 같이 폴더가 생성 되었는지 확인합니다.



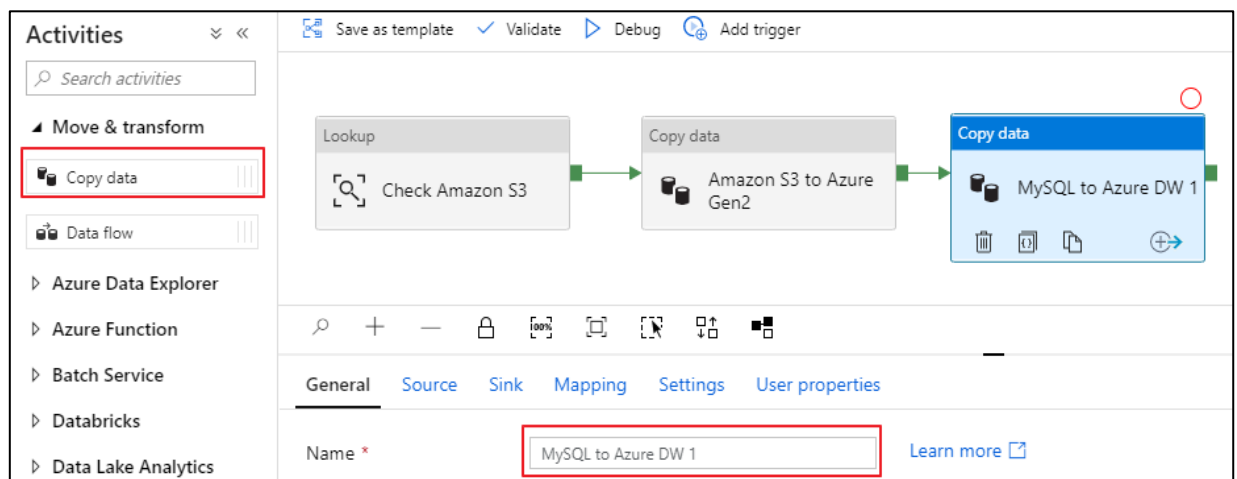
4.2. Task 2: MySQL to Azure Synapse Analytics

이 작업은 Task 1에서 복사한 파일을 이용하여 Azure Data Warehouse에 데이터를 이관합니다.

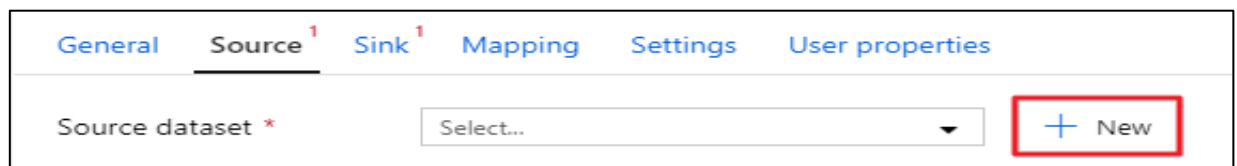


1. **Activities > Move & transform > Copy data**를 드래그해서 옮긴 후 하단 **General** 탭에서 다음을 입력합니다.

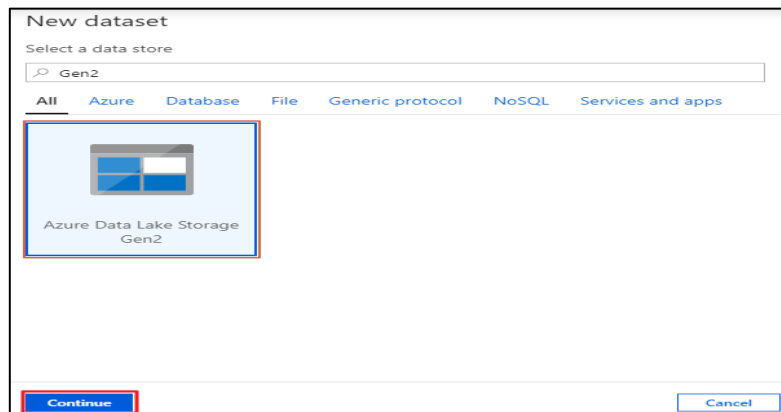
✓ Name: **MySQL to Azure DW 1**



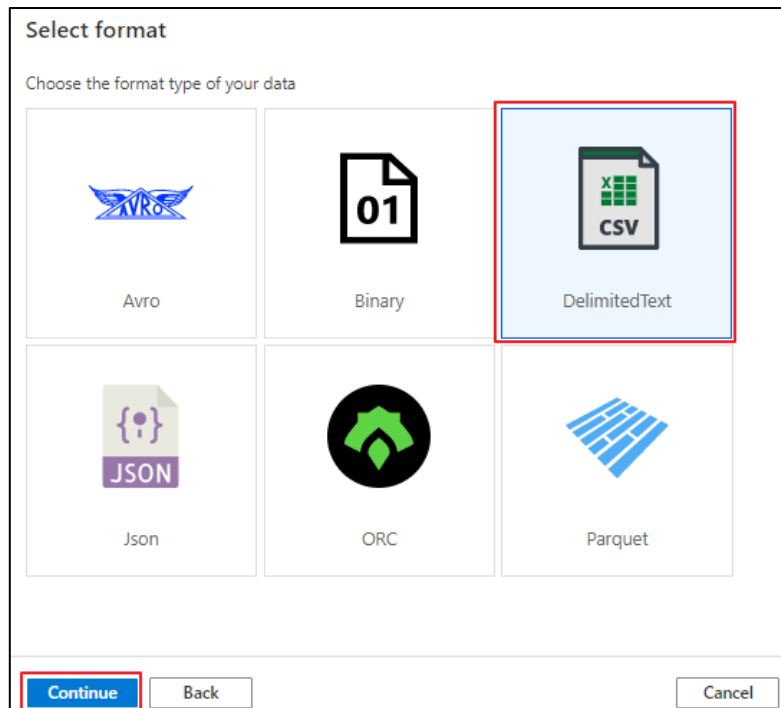
2. **Source** 탭으로 이동하여 **Source dataset**의 **+New** 를 클릭합니다.



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



4. format은 **DelimitedText**를 선택합니다



5. Set Properties 블레이드가 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**

Set properties

Name

Linked service *

[Edit connection](#)

File path
 / / Browse ▾

First row as header ☒

Import schema
☒ From connection/store ☐ From sample file ☐ None

▸ Advanced

OK Back Cancel

6. **Source** 탭의 Source Dataset이 세팅된 것을 확인합니다.

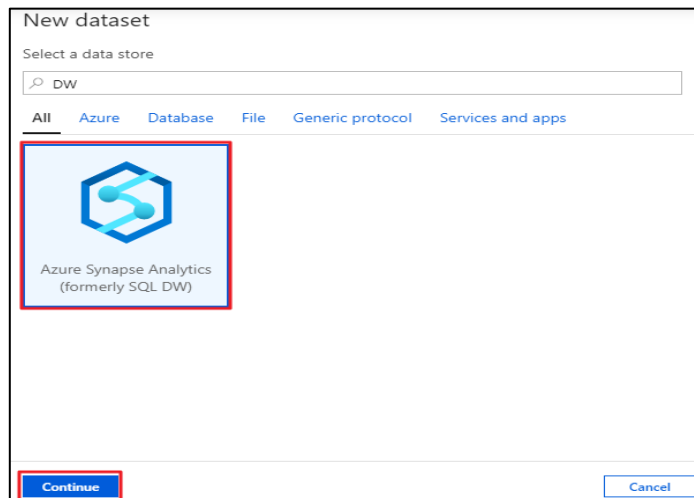
The screenshot shows a Data Factory pipeline with three tasks: 'Lookup' (Check Amazon S3), 'Copy data' (Amazon S3 to Azure Gen2), and 'Copy data' (MySQL to Azure DW 1). The 'Source' tab is selected, and the 'Source dataset' dropdown is set to 'Gen2_Storage_Dim1', which is highlighted with a red box. The 'Sink' tab is also visible, and the 'Sink dataset' dropdown is set to 'Select...'. The '+ New' button is highlighted with a red box.

7. **Sink** 탭으로 이동하여 Sink dataset의 +New를 클릭합니다.

[General](#) [Source](#) [Sink¹](#) [Mapping](#) [Settings](#) [User properties](#)

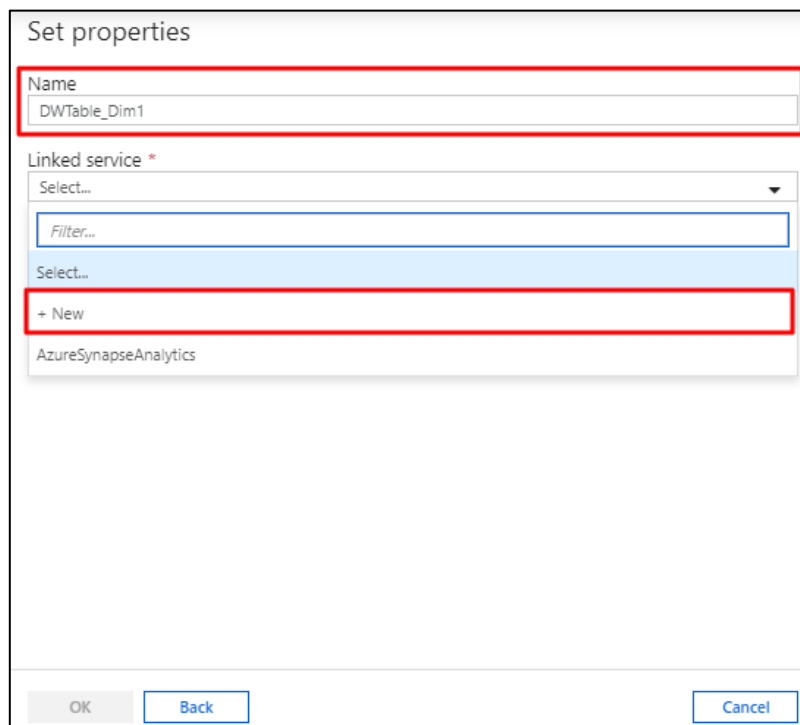
Sink dataset * + New

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



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

✓ Name: **DWTable_Dim1**



10. 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 : **CohoDWXXXX**
- ✓ User name : **demouser**
- ✓ Password : **Demo@pass123**

New linked service (Azure Synapse Analytics (formerly SQL DW))

Name *
AzureSynapseAnalytics

Description

Connect via integration runtime *
AutoResolveIntegrationRuntime

Connection string Azure Key Vault

Account selection method
☒ From Azure subscription ☐ Enter manually

Azure subscription
Azure subscription 1 (35120fde-ae11-40e4-8ea2-8be769fdcc5d)

Server name *
asodwserver0000

Database name *
CohoDW0000

Authentication type *
SQL authentication

User name *
demouser

Password Azure Key Vault

Password *

☒ Connection successful

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

- ✓ Name: **DWTable_Dim1**
- ✓ Linked service : **AzureSynapseAnalytics**
- ✓ Table Name : **dbo.DimCustomer**

Set properties

Name
DWTable_Dim1

Linked service *
AzureSynapseAnalytics

[Edit connection](#)

Table name
dbo.DimCustomer

☐ Edit

Import schema
☒ From connection/store
 ☐ None

▶ Advanced

OK Back Cancel

12. Copy data 액티비티의 **Sink** 탭의 Sink Datasets이 세팅된 것을 확인합니다.

General Source **Sink** Mapping Settings User properties

Sink dataset * DWTable_Dim1 [Open](#) [+ New](#)

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

General Source Sink Mapping **Settings** User properties

i You will be charged # of used DIUs * copy duration * \$0.25/DIU-hour. Local currency and

Data integration unit Auto [Edit](#)

Degree of copy parallelism [Edit](#)

Fault tolerance Abort activity on first incompatible row [Edit](#)

Enable staging ☒ [Add dynamic content \[Alt+P\]](#)

Staging settings

Staging account linked service Select... [+ New](#)

14. 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: **asostorageXXXX** (스토리지 계정 명)

New linked service

Name *
AzureBlobStorage

Description

Connect via integration runtime *
AutoResolveIntegrationRuntime

Authentication method
Account key

Connection string | Azure Key Vault

Account selection method
☒ From Azure subscription ☐ Enter manually

Azure subscription
Azure subscription 1 (35120fde-ae11-40e4-8ea2-8be769fdcc5d)

Storage account name *
asostorage0000

Additional connection properties

Create | **Test connection** | **Cancel**

✓ Connection successful

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

Lookup | **Copy data** | **Copy data**

Check Amazon S3 | Amazon S3 to Azure Gen2 | MySQL to Azure DW 1

General | **Source** | **Sink** | **Mapping** | **Settings** | **User properties**

You will be charged # of used DIUs * copy duration * \$0.25/DIU-hour. Local currency and separate discounting may

Data integration unit
Auto | Edit

Degree of copy parallelism
| Edit

Fault tolerance
Abort activity on first incompatible row

Enable staging
☒

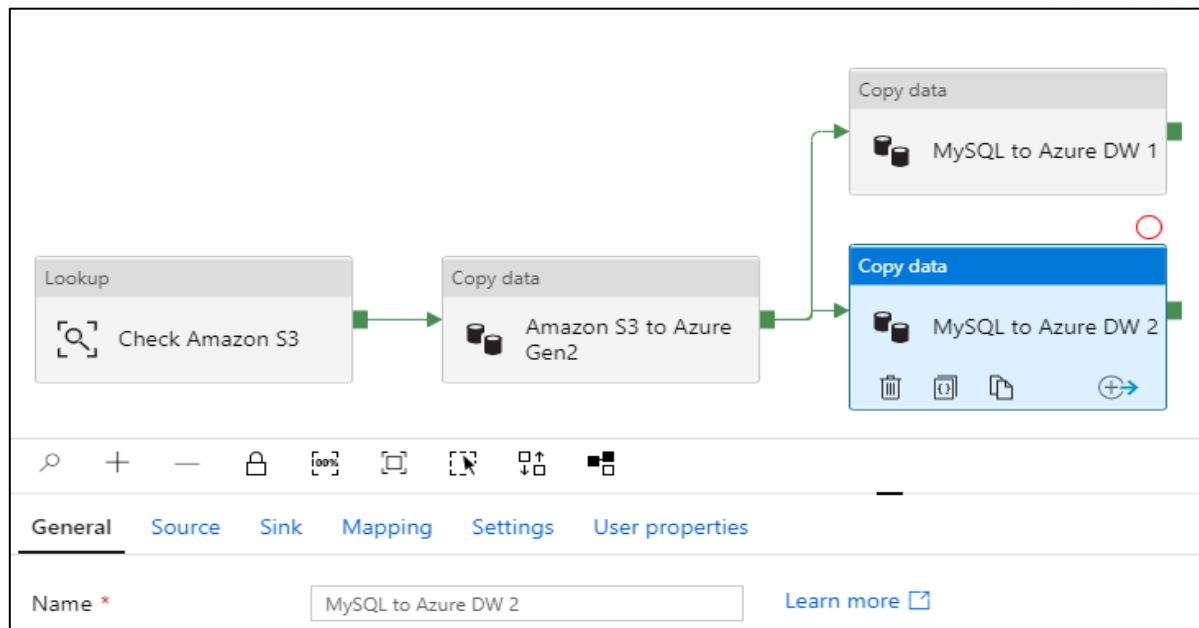
Staging settings

Staging account linked service
AzureBlobStorage

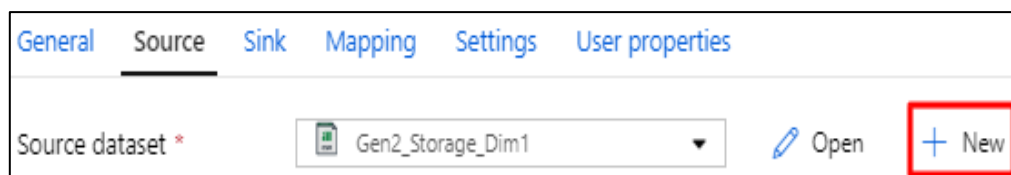
Test connection | **Open**

16. MySQL to Azure DW 1 를 복사 후 붙여넣기 후 하단 **General** 탭에서 다음을 입력합니다.

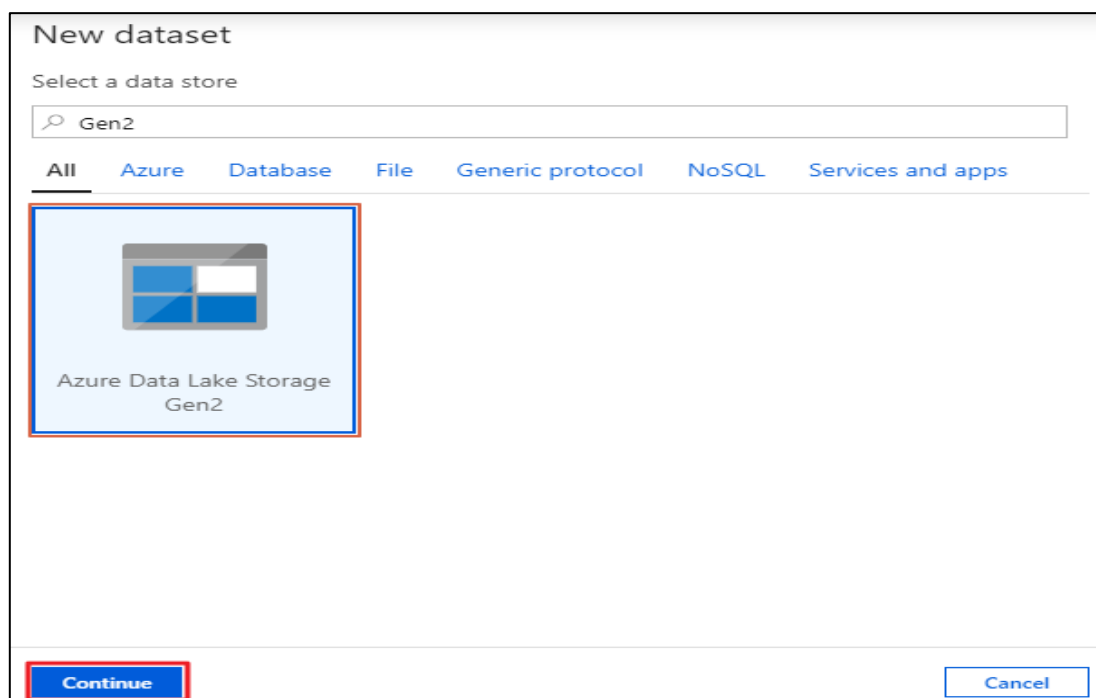
✓ Name: **MySQL to Azure DW 2**



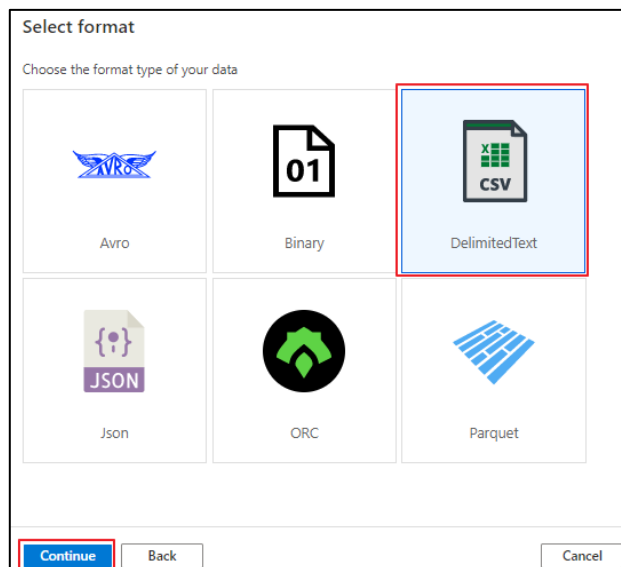
17. Source 탭으로 이동하여 Source dataset의 **+New** 를 클릭합니다.



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

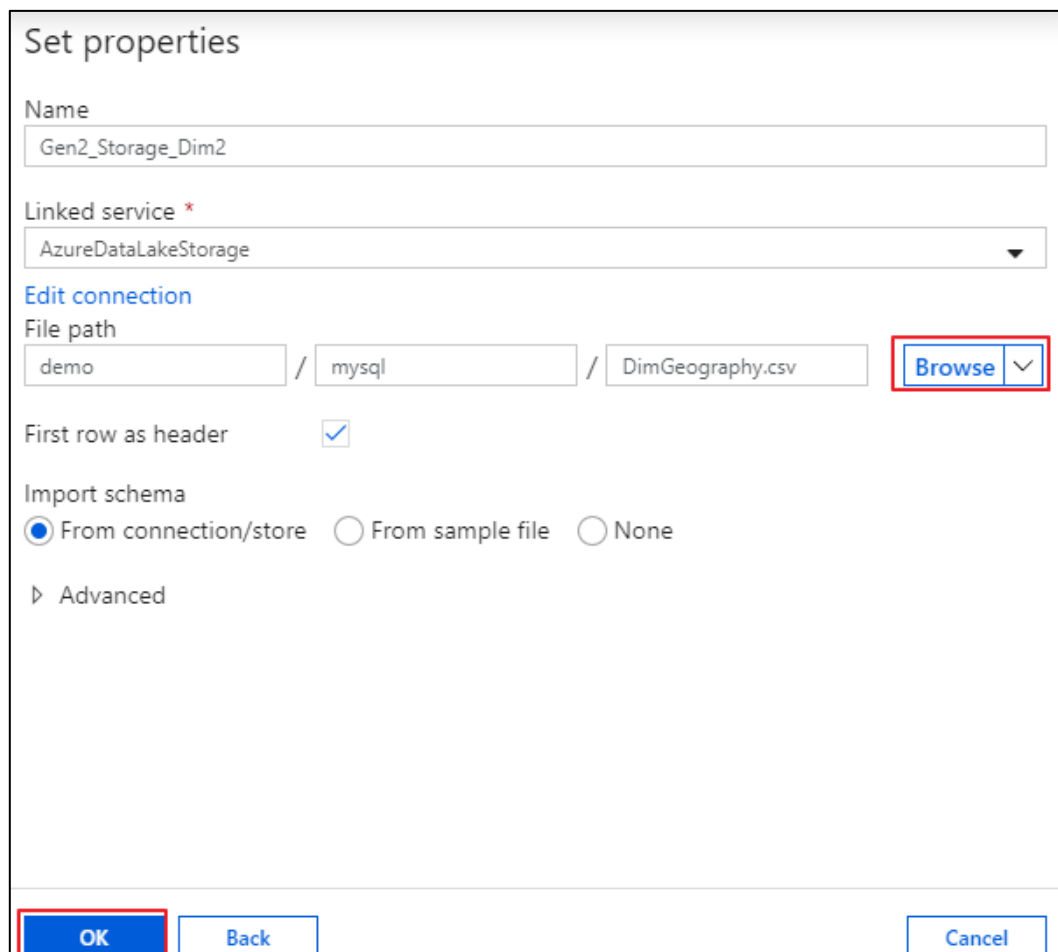


19. format은 **DelimitedText**를 선택합니다

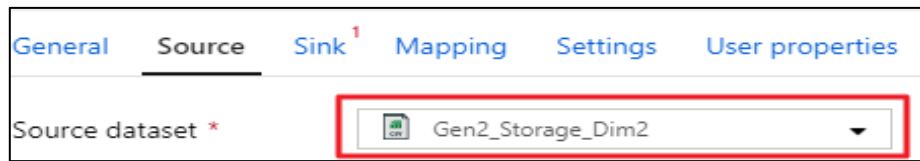


20. Set Properties 블레이드가 열리면 아래의 정보를 입력 후 **OK**버튼을 클릭합니다.

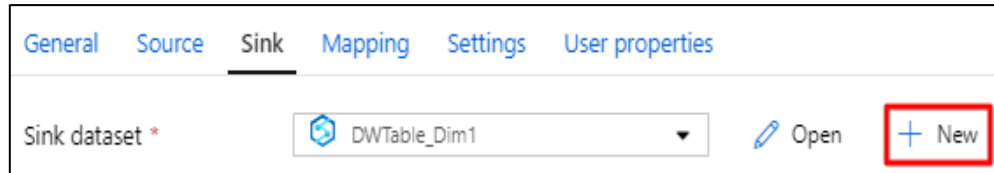
- ✓ Name: **Gen2_Storage_Dim2**
- ✓ Linked service : **AzureDataLakeStorage**
- ✓ File path : **demo / mysql / DimGeography.csv** (Browse 버튼 클릭하여 경로지정)
- ✓ First row as header : **Check**



21. **Source** 탭의 Source Dataset이 세팅된 것을 확인합니다.

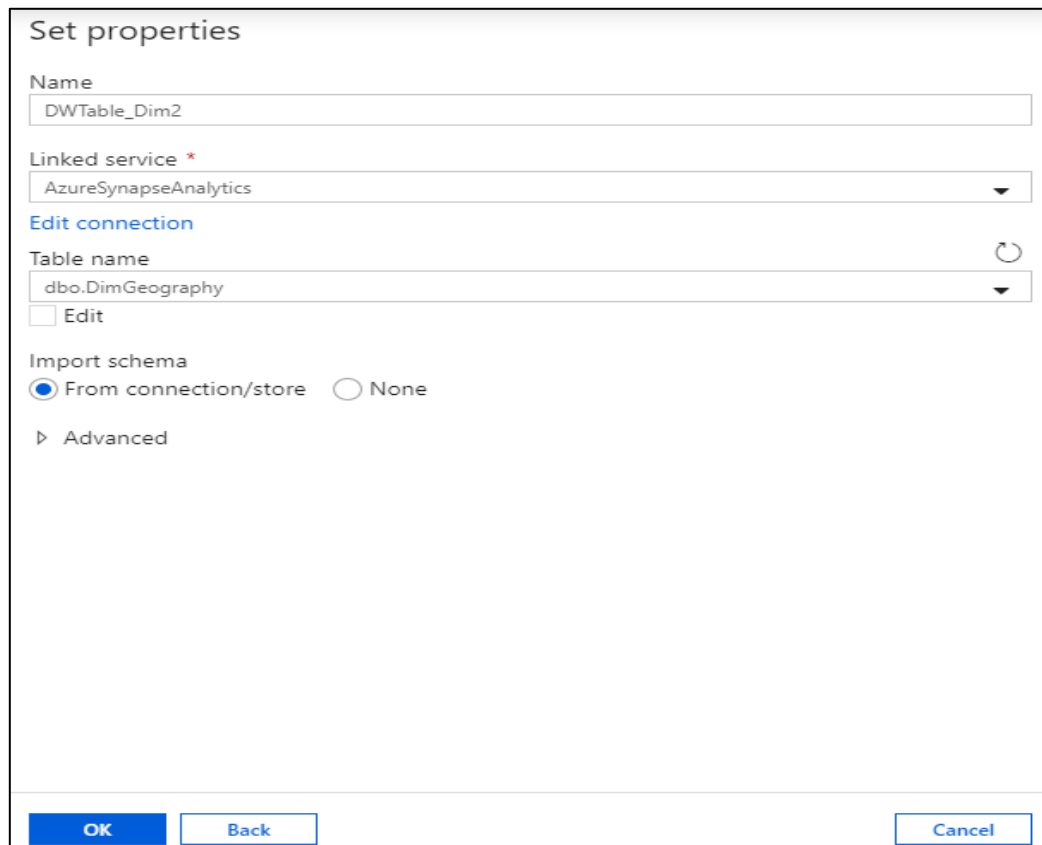


22. **Sink** 탭으로 이동하여 Sink dataset의 **+New**를 클릭합니다.

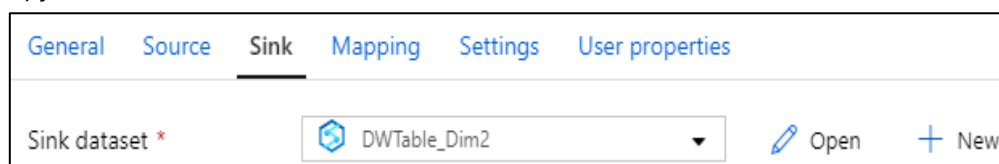


23. Set properties 화면이 열리면 아래와 같이 입력 후 OK 버튼을 클릭합니다.

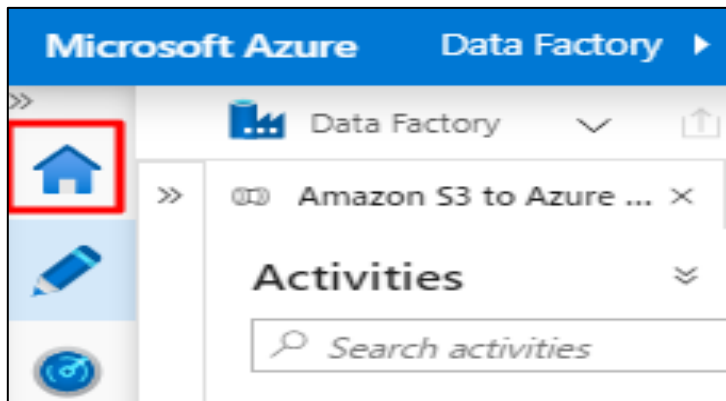
- ✓ Name: **DWTable_Dim2**
- ✓ Linked service : **AzureSynapseAnalytics**
- ✓ Table Name : **dbo. DimGeography**



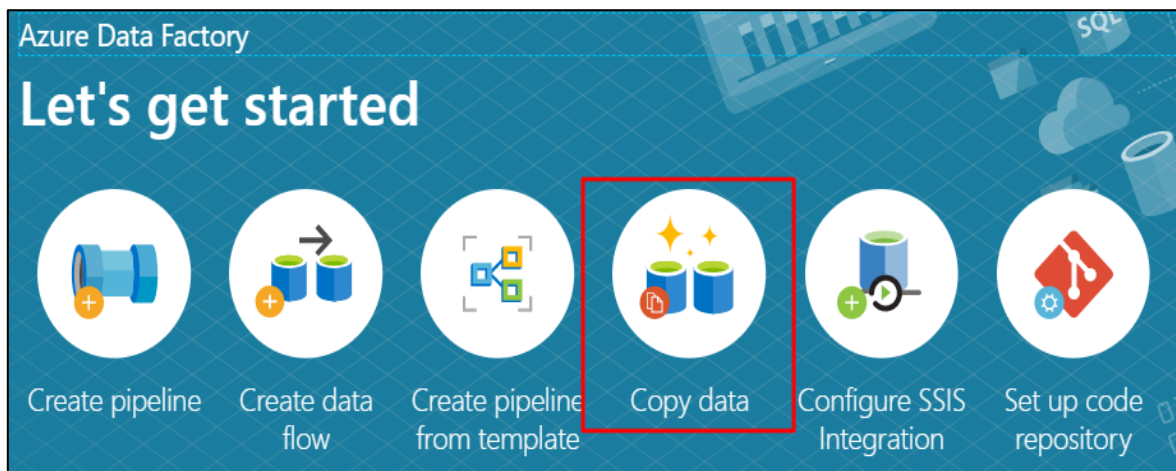
24. Copy data 액티비티의 **Sink** 탭의 Sink Dataset이 세팅된 것을 확인합니다.



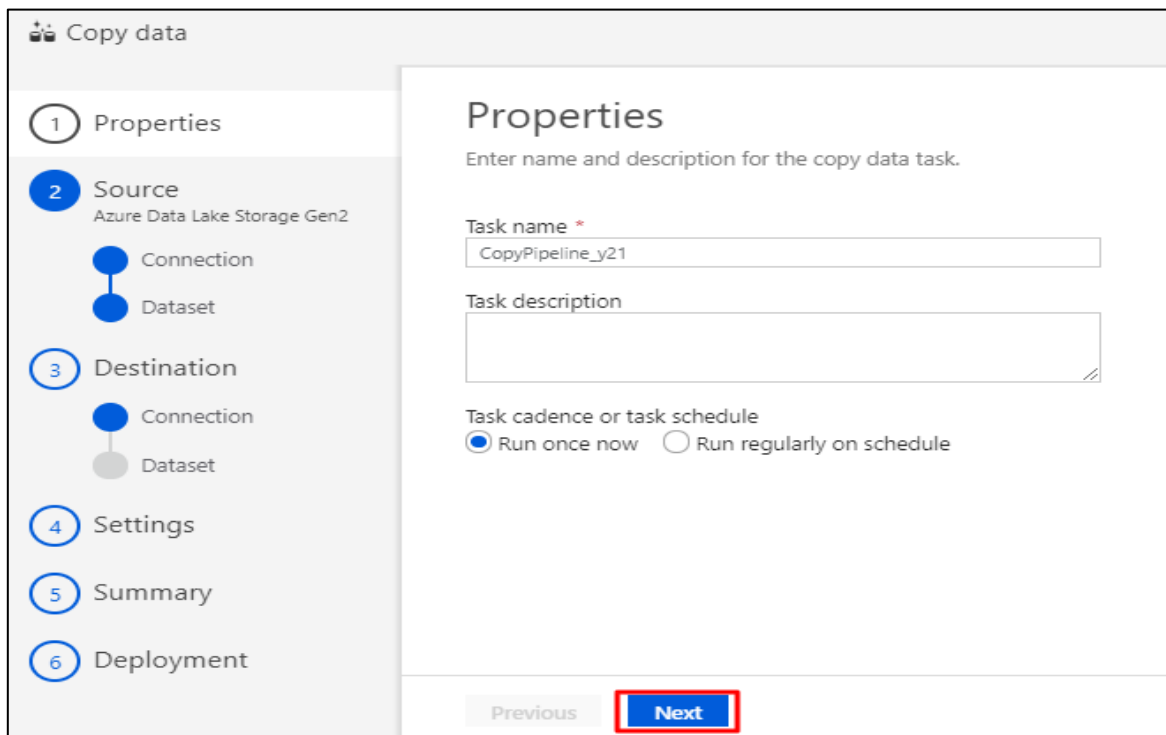
25. 왼쪽 메뉴에서 **Home** 아이콘을 클릭합니다.



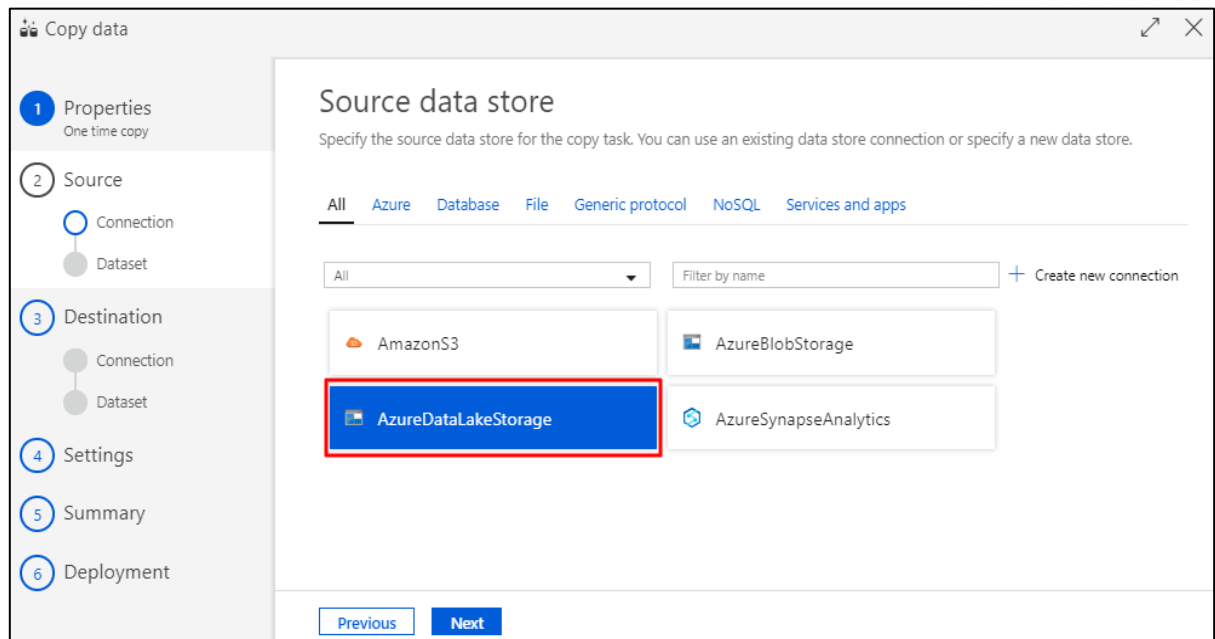
26. Data Factory 포털 홈 화면에서 **Copy Data**를 클릭합니다.



27. Properties 화면에서 **Next** 버튼을 클릭합니다.

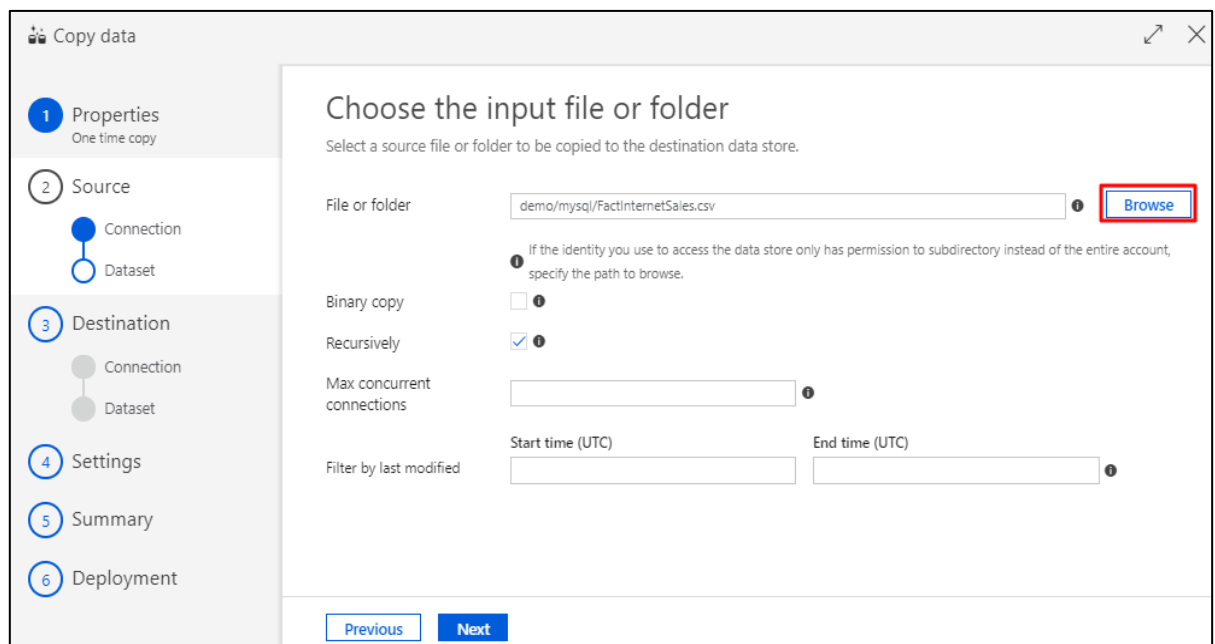


28. 기존에 만들어 놓았던 Gen2 스토리지인 **AzureDataLakeStorage**를 선택 후 Next 버튼을 클릭합니다.



29. Browse 버튼을 클릭하여 다음과 같이 경로를 지정합니다.

✓ File path : **demo / mysql / FactInternetSales.csv**



30. 파일 포맷 및 미리보기를 확인 후 Next 버튼을 클릭합니다.

The screenshot shows the 'Copy data' wizard with the 'File format settings' step selected. The left sidebar shows the progression from Properties to Deployment. The main area displays settings for File format (Text format), Column delimiter (Comma (,)), and Row delimiter (Auto detect). A 'Detect text format' button is present. Below the settings, a 'Preview' tab shows a table of data with columns: CarrierTrackingNumber, CurrencyKey, CustomerKey, CustomerPONumber, DiscountAmount, DueDateKey, and ExtendedAmount. The table contains two rows of data. At the bottom, there are 'Previous' and 'Next' buttons.

CarrierTrackingNumber	CurrencyKey	CustomerKey	CustomerPONumber	DiscountAmount	DueDateKey	ExtendedAmount
19	21768			0	20010713	\$3578.27
39	28389			0	20010713	\$3399.99

31. Destination data store 화면에서 **AzureSynapseAnalytics**를 선택 후 **Next** 버튼을 클릭합니다.

The screenshot shows the 'Copy data' wizard with the 'Destination data store' step selected. The left sidebar shows the progression from Properties to Deployment. The main area displays a list of data stores: AzureBlobStorage, AzureDataLakeStorage, and AzureSynapseAnalytics. The 'AzureSynapseAnalytics' option is highlighted with a red dashed box. At the bottom, there are 'Previous' and 'Next' buttons.

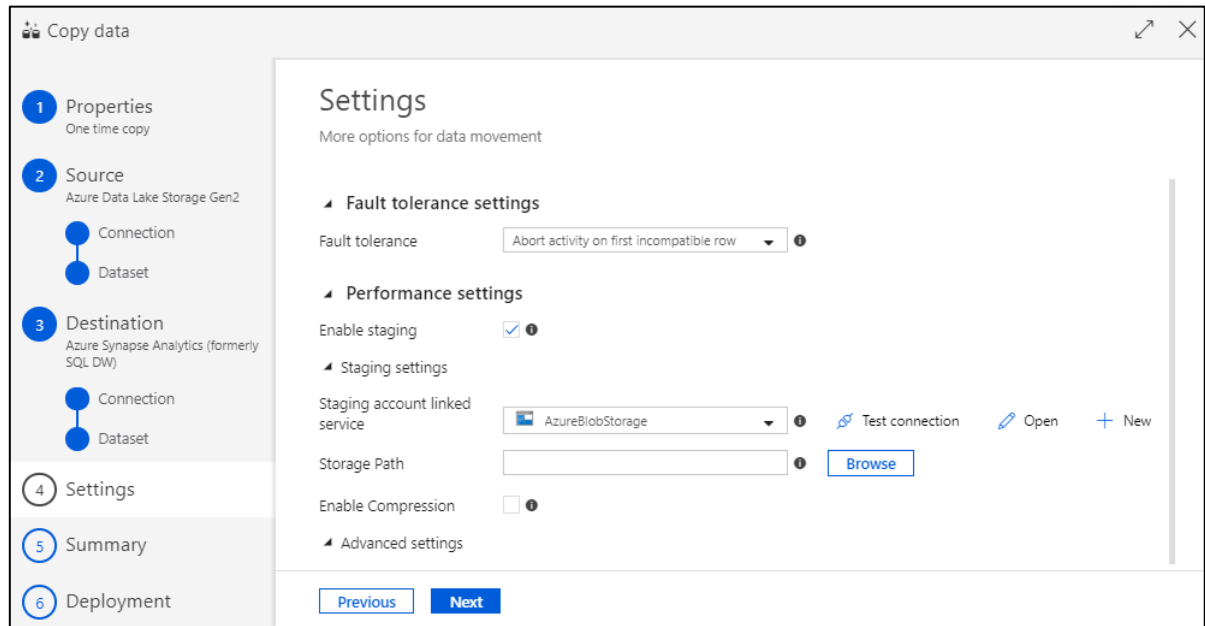
32. Destination 항목의 Use existing table 클릭 후 **dbo.factInternetSales**를 선택하고 **Next** 버튼을 클릭합니다.

The screenshot shows the 'Copy data' wizard in the 'Table mapping' step. The left sidebar has six steps: 1 Properties (One time copy), 2 Source (Azure Data Lake Storage Gen2), 3 Destination, 4 Settings, 5 Summary, and 6 Deployment. Step 3 is currently active. The main area is titled 'Table mapping' and contains the instruction: '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 stored procedure to run at the destination.' Below this, there are two columns: 'Source' and 'Destination'. Under 'Source', it says 'Azure Data Lake Storage Gen2 file'. Under 'Destination', a dropdown menu is open, showing 'dbo.factInternetSales' selected and highlighted with a red rectangle. At the bottom, there is a checkbox labeled 'Skip column mapping for all tables' which is unchecked, and two buttons: 'Previous' and 'Next'.

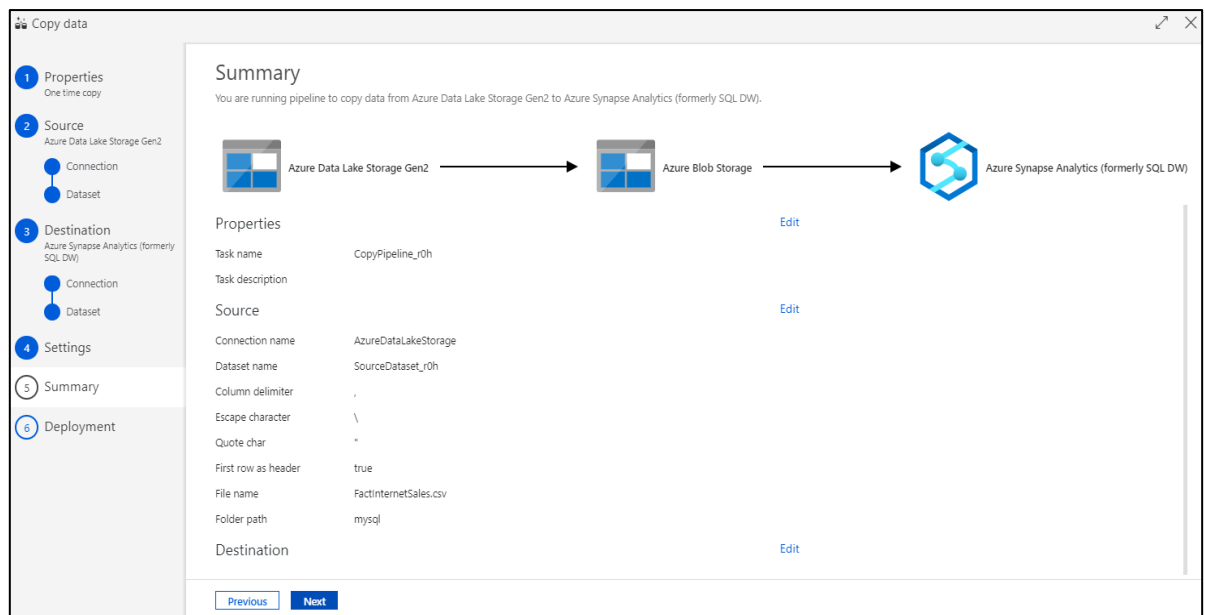
33. 컬럼 매핑 정보를 확인 후 Next 버튼을 클릭합니다.

The screenshot shows the 'Copy data' wizard in the 'Column mapping' step. The left sidebar is the same as in the previous step, with step 3 'Destination' active. The main area is titled 'Column mapping' and contains the instruction: 'Choose how source and destination columns are mapped'. Below this, there are two sections: 'Table mappings (1)' and 'Column mappings'. In 'Table mappings (1)', there is a table with two rows: 'Source' (Azure Data Lake Storage Gen2 file) and 'Destination' (dbo.factInternetSales). In 'Column mappings', there are buttons for '+ New mapping', 'Clear', and 'Delete'. Below these, there is a table with two columns: 'Source' and 'Type'. The first row shows 'ProductKey' in the 'Source' column and 'abc String' in the 'Type' column. Below this, there is a section titled 'Azure Synapse Analytics (formerly SQL DW) sink properties' with two input fields: 'Pre-copy script' and 'Write batch timeout'. At the bottom, there are two buttons: 'Previous' and 'Next'.

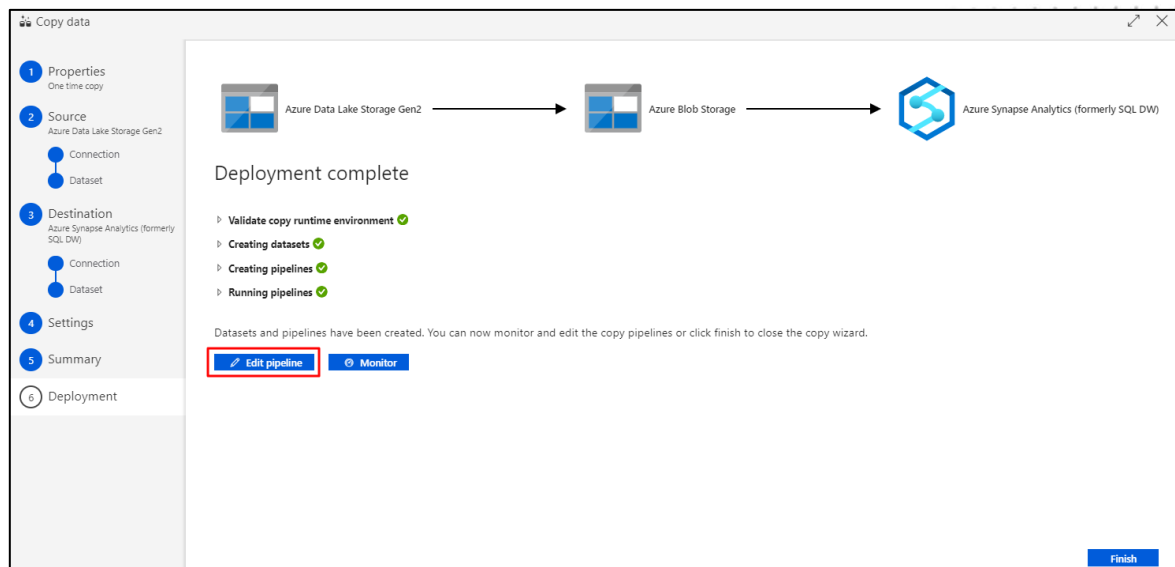
34. Settings 화면의 Enable staging 항목에 체크 후 Staging account linked service 항목의 **AzureBlobStorage**를 선택합니다.



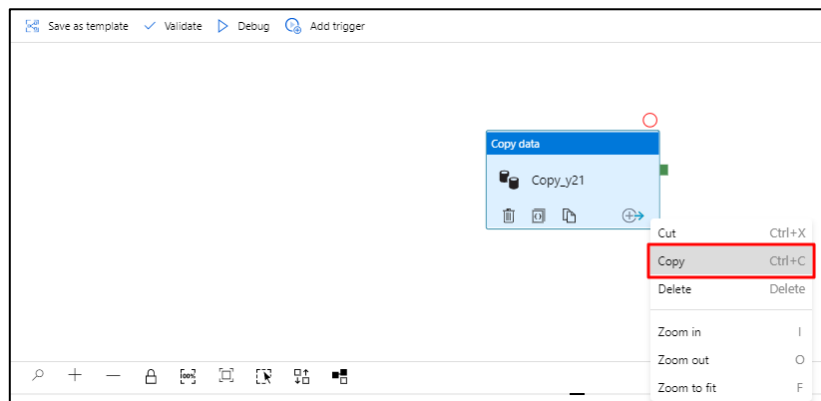
35. 요약된 내용을 확인 후 **Next** 버튼을 클릭합니다.



36. Deploy가 완료되면 **Edit pipeline** 버튼을 클릭하여 pipeline으로 이동합니다.

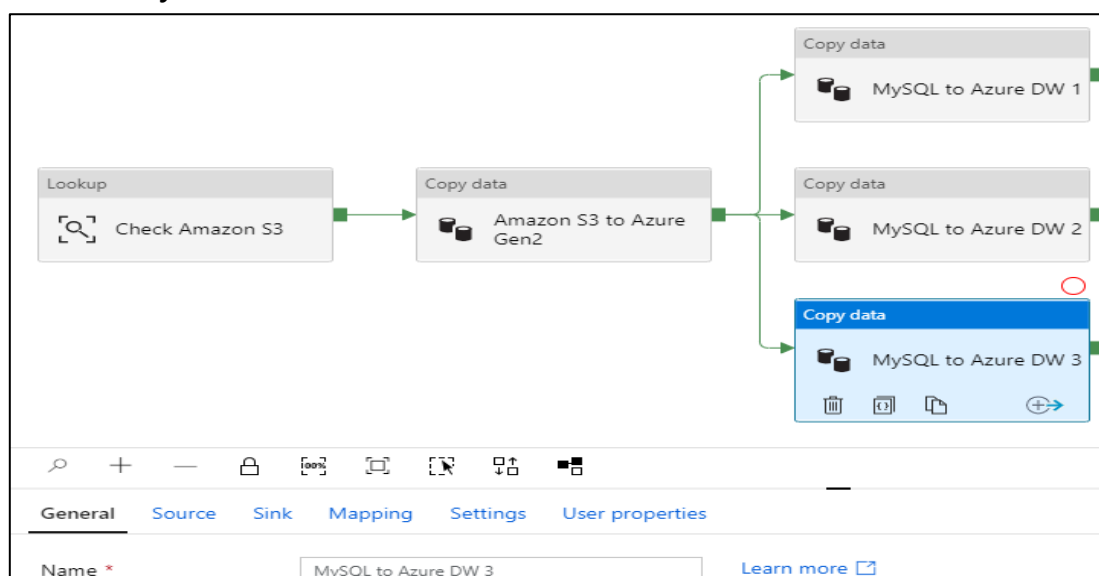


37. 생성된 copy data를 복사합니다.

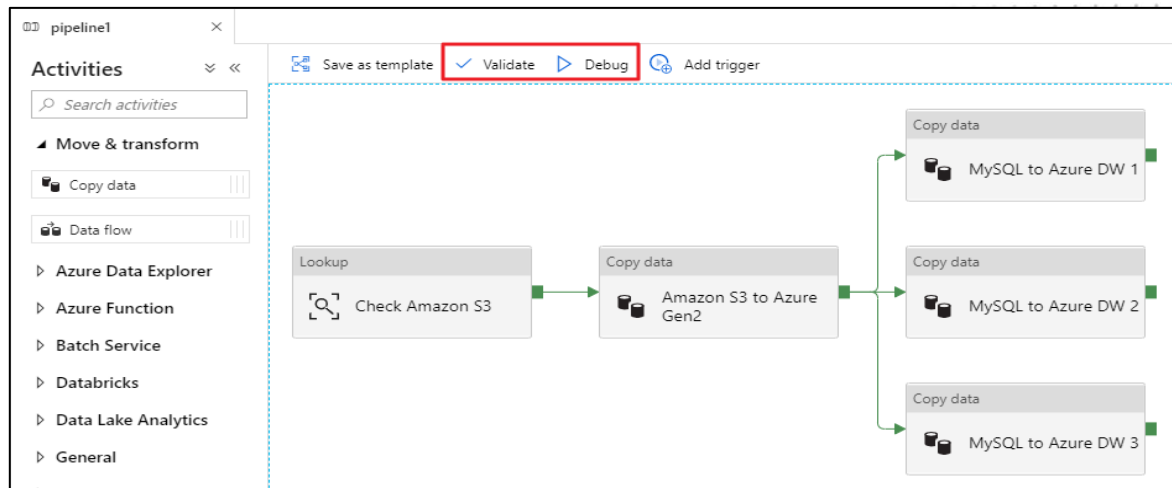


38. 기존에 작업중이던 파이프라인에 붙여넣기 한 후 Amazon S3 to Azure Gen2와 연결하여 **General** 탭에서 다음을 입력합니다.

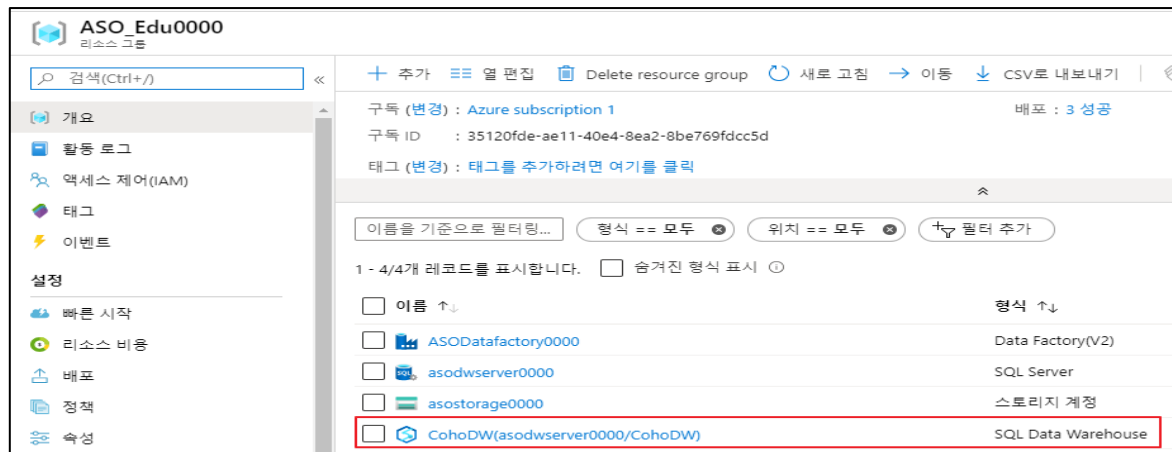
✓ Name: **MySQL to Azure DW 3**



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



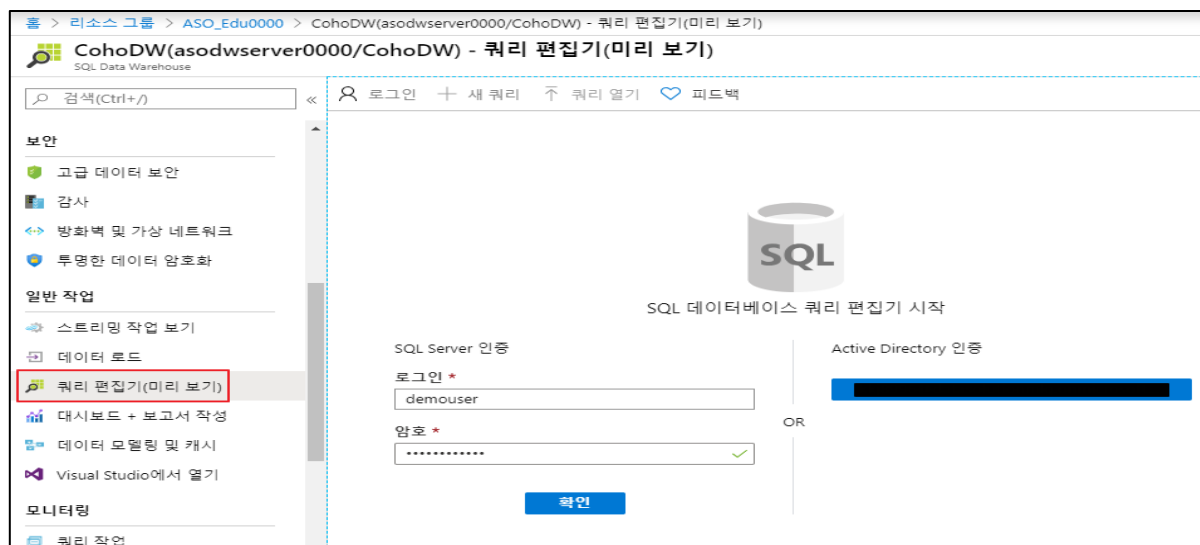
1. 실행 결과를 확인하기 위해 Azure Portal 의 리소스 그룹으로 이동하여 **CohoDWXXXX**를 선택합니다.



2. 왼쪽 메뉴에서 쿼리 편집기를 클릭 후 아래의 정보를 입력하여 로그인합니다.

✓ 로그인: **demouser**

✓ 암호: **Demo@pass123**




3. 다음 쿼리를 1라인씩 블록 지정 하여 실행하여 테이블 별 데이터가 정상적으로 조회되는지 확인 합니다.

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

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

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

여기에 제한된 개체 탐색기를 표시하는 중입니다. 전체 기능을 확인하려면 SSDT를 여세요.

테이블

dbo.DimCustomer

...

dbo.DimGeography

...

dbo.FactInternetSales

...

뷰

저장 프로시저

실행

쿼리 취소

쿼리 저장

데이터를 json(으)로 내보내기

데이터를 .csv(으)로 내보내기

...

1

2

3

4

5

6

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

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

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

결과

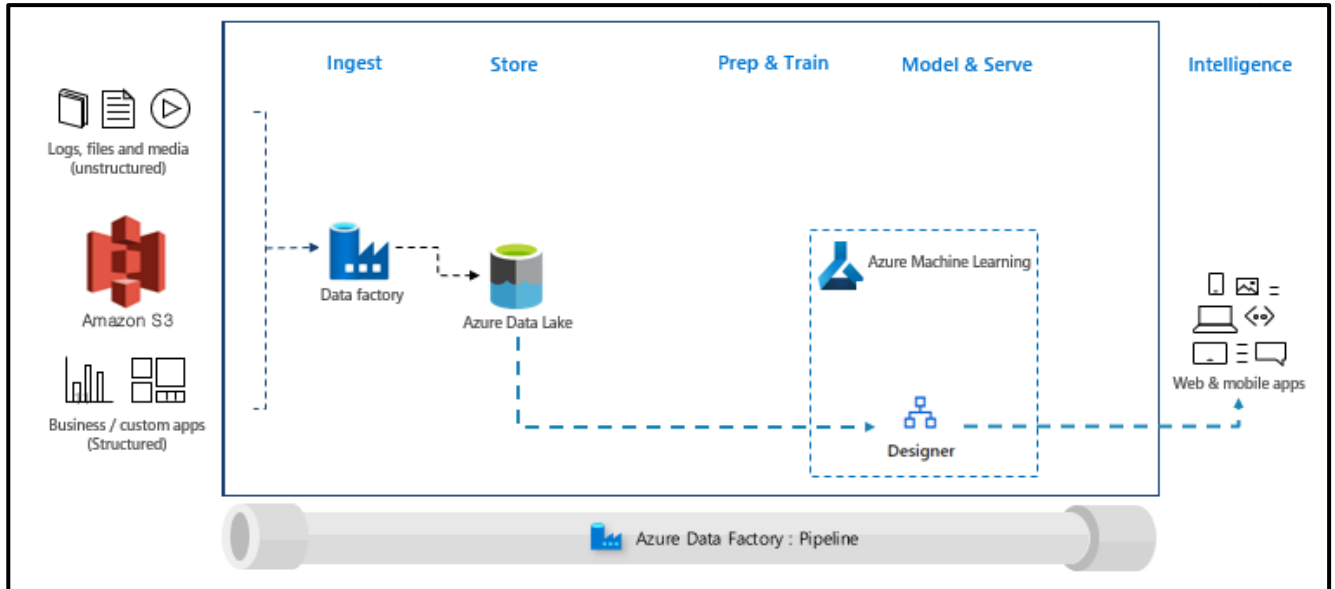
메시지

필터링 항목 검색...

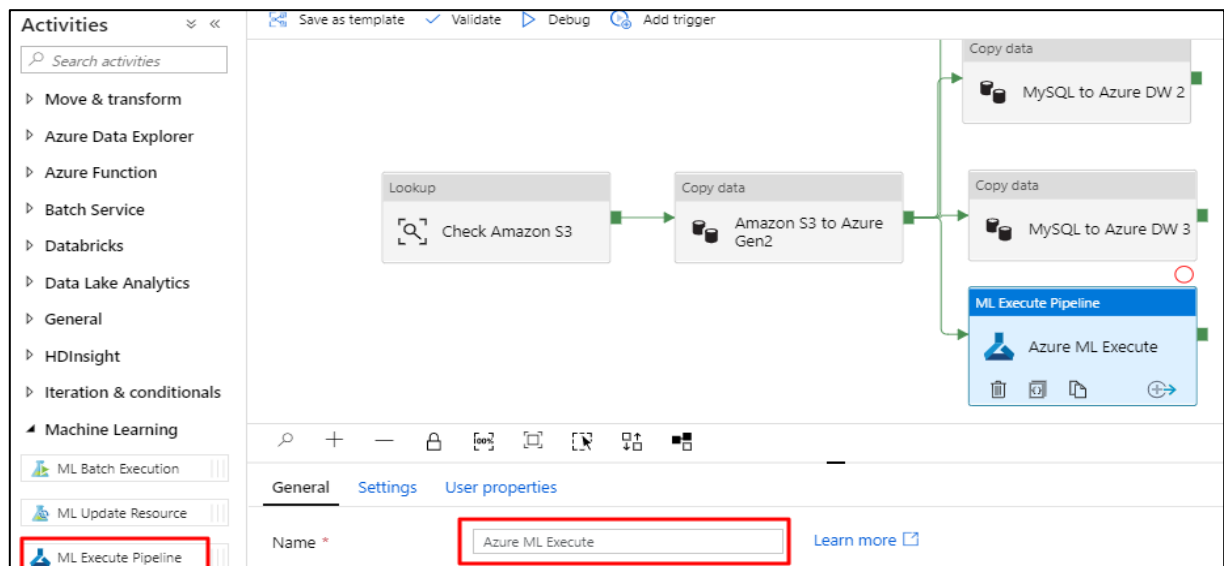
CustomerKey	GeographyKey	CustomerAlternateKey	Title
14988	248	AW00014988	
15946	298	AW00015946	
15096	156	AW00015096	Mr.
18541	71	AW00018541	
11565	229	AW00011565	
12021	16	AW00012021	

4.3. Task 3: Azure ML Execute

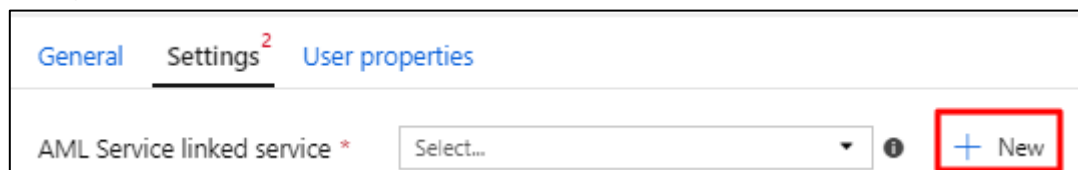
이 작업에서는 Azure ML 서버스에 연결하여 Task 1에서 복사한 파일을 이용하여 Model을 학습하는 ML 파이프라인을 실행시킵니다.



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**



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



4. 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: **Enter manually**
- ✓ Subscription ID : **35120fde-ae11-40e4-8ea2-8be769fdcc5d**
- ✓ Resource group name: **test**
- ✓ Azure ML Service workspace name: **test**
- ✓ Tenant : **56480cdf-0e19-4439-bf83-359f8a87e112**
- ✓ Service principal ID: **13c2c33f-271e-4ee0-ad2a-9cf4f6d7b917**
- ✓ Service principal key : **oKJwUqoilZXWM24u_fbXo_64QDGWlaRBAAd**

New linked service (Azure Machine Learning)

Name *
AzureMLService

Description

Connect via integration runtime *
AutoResolveIntegrationRuntime

Azure Machine Learning workspace selection method
☐ From Azure subscription ☒ Enter manually

Subscription ID *
35120fde-ae11-40e4-8ea2-8be769fdcc5d

Resource group name *
test

Azure Machine Learning workspace name *
test

Tenant *
56480cdf-0e19-4439-bf83-359f8a87e112

Service principal ID *
13c2c33f-271e-4ee0-ad2a-9cf4f6d7b917

Service principal key Azure Key Vault
.....

Annotations
+ New

Name

Advanced

Connection successful

Create Test connection Cancel

5. AML Service linked service 항목이 세팅된 것을 확인 후 아래의 정보를 입력합니다.
- ✓ ML pipeline ID : Add dynamic content 클릭하여 입력 > **cf97b6c3-82e9-46ce-89c0-73316b0ce6c7**
 - ✓ Experiment name : **ASO_DataFactory**

General Settings User properties

Azure Machine Learning linked service AzureMLService Test connection Edit New

Machine Learning pipeline name

Machine Learning pipeline ID cf97b6c3-82e9-46ce-89c0-73316b0ce6c7 Open in Azure Portal

Experiment name ASO_DataFactory

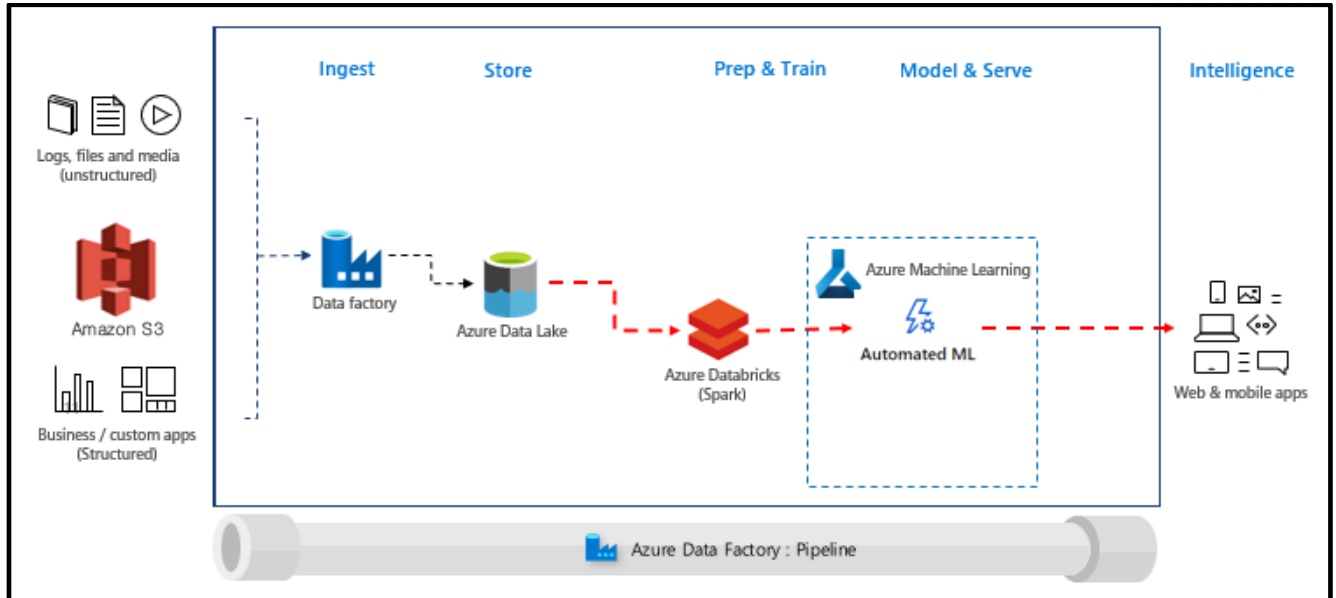
Machine Learning pipeline parameters

Machine Learning parent run ID

Continue on step failure ☐

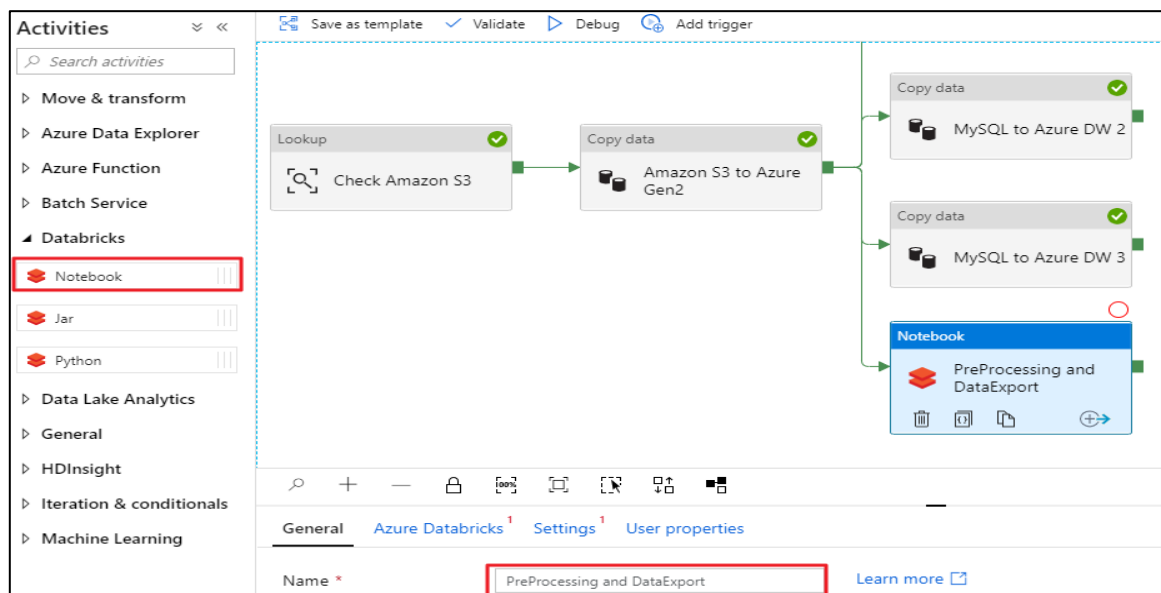
4.4. Task 4: PreProcessing and DataExport

이 작업은 Task 1에서 복사한 파일을 이용하여 Azure Databricks 서비스에 연결하여 Automated ML을 실행하기 위한 데이터셋을 만듭니다.

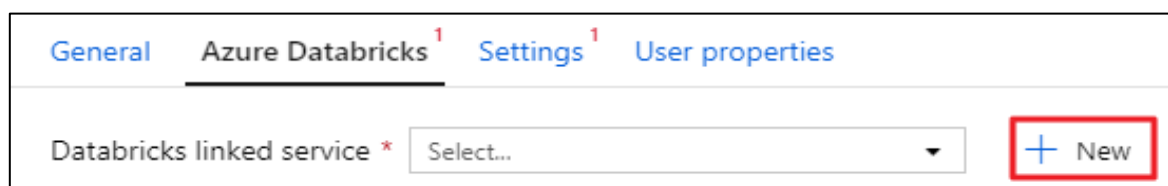


1. **Activities > Databricks > Notebook**을 드래그해서 옮긴 후 하단 **General** 탭에서 다음을 입력합니다.

✓ Name: **PreProcessing and DataExport**



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



3. 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: **Enter manually**
- ✓ Databrick Workspace URL : **https://adb-7283045618948357.17.azuredatabricks.net**
- ✓ Access token : **dapid1fb7b77a6d0e7f9fb2ccf337367a9a5**
- ✓ Select cluster : **Existing interactive cluster**
- ✓ Existing cluster ID: **Cluster_ASO**

4. Databricks linked service 항목이 세팅된 것을 확인합니다.

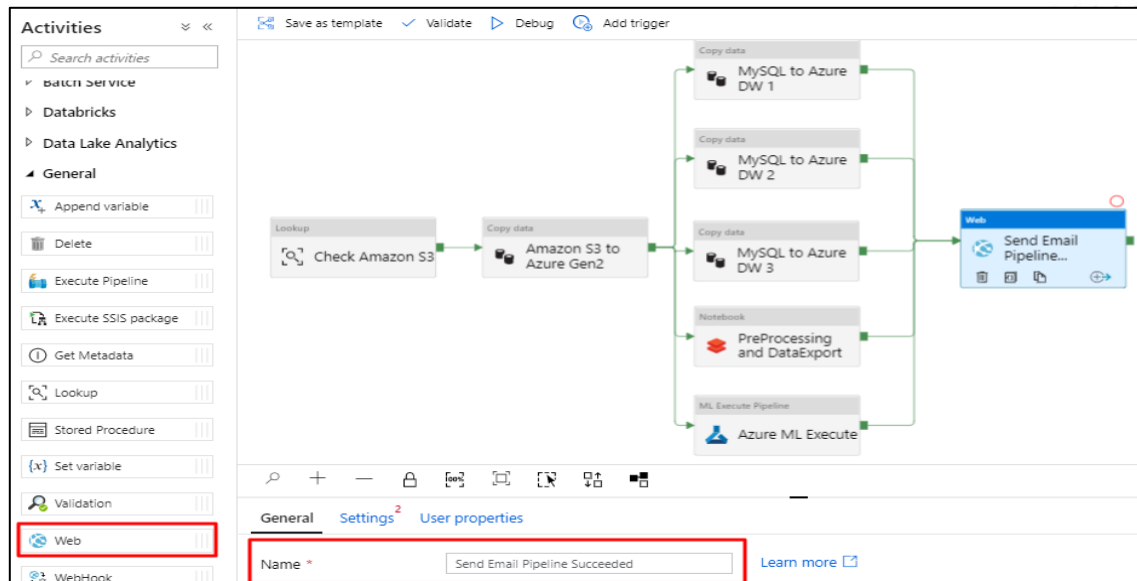
5. Settings 탭으로 이동하여 Notebook path 항목의 Browse 버튼을 클릭하여 다음과 같이 경로를 지정합니다.

- ✓ Notebook path: **/Users/kch8306@datasolutionmlg.onmicrosoft.com/PreProcessing_EDU**

4.5. Task 5: Send Email

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

✓ **Name: Send Email Pipeline Succeeded**



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

✓ URL: **https://prod-**

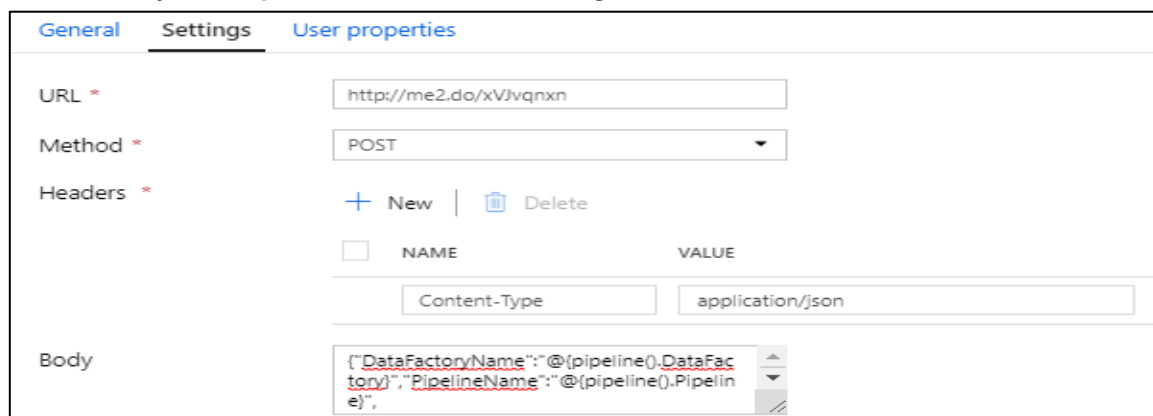
06.koreacentral.logic.azure.com:443/workflows/21f3052675d24be0aa4298a9b0490ce8/t
riggers/manual/paths/invoke?api-version=2016-10-
01&sp=%2Ftriggers%2Fmanual%2Frun&sv=1.0&sig=BfXuXTp05p20wnvDFzkUgi3xbf3X
GSB-1qapyC_EDFA

✓ Method: **POST**

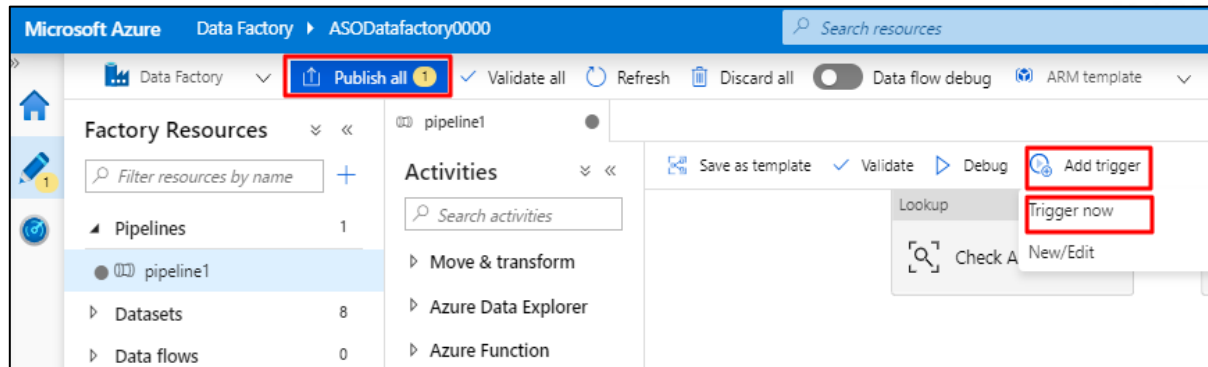
✓ Headers: **+New** 클릭 후 NAME 항목에 **Content-Type**, VALUE 항목에 **application/json** 추가

✓ Body:

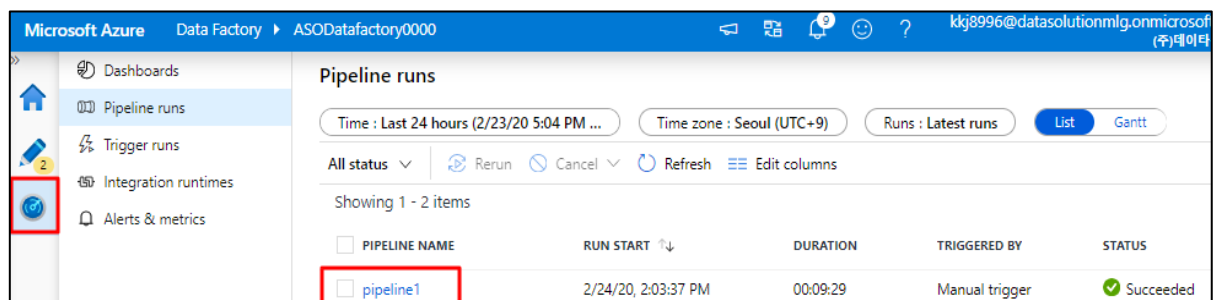
{"DataFactoryName":"@{pipeline().DataFactory}","PipelineName":"@{pipeline().Pipeline}"
,"
"Subject":"Pipeline Succeeded","Message":"Succeeded!","EmailTo":"본인의 메일계정"}



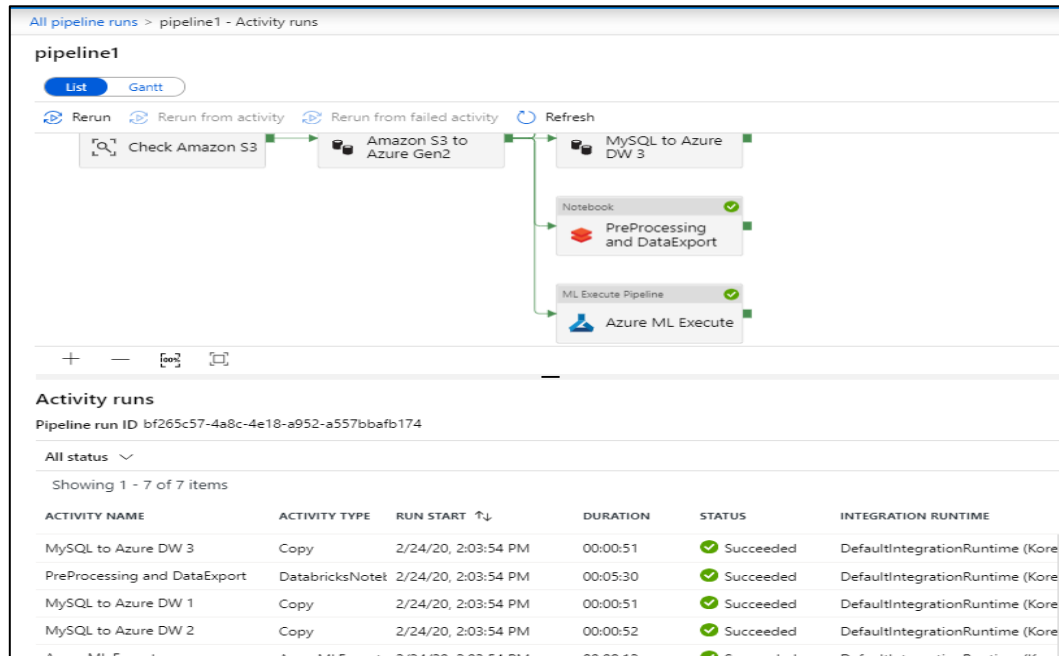
3. 최종 실행을 위해 왼쪽 상단의 **Publish All** 버튼 클릭 후 **Add trigger**의 Trigger now를 클릭합니다.



4. 왼쪽 메뉴의 Monitor 아이콘을 클릭하여 진행 상태를 확인합니다.



5. PIPELINE NAME을 클릭하여 상세화면으로 이동합니다.

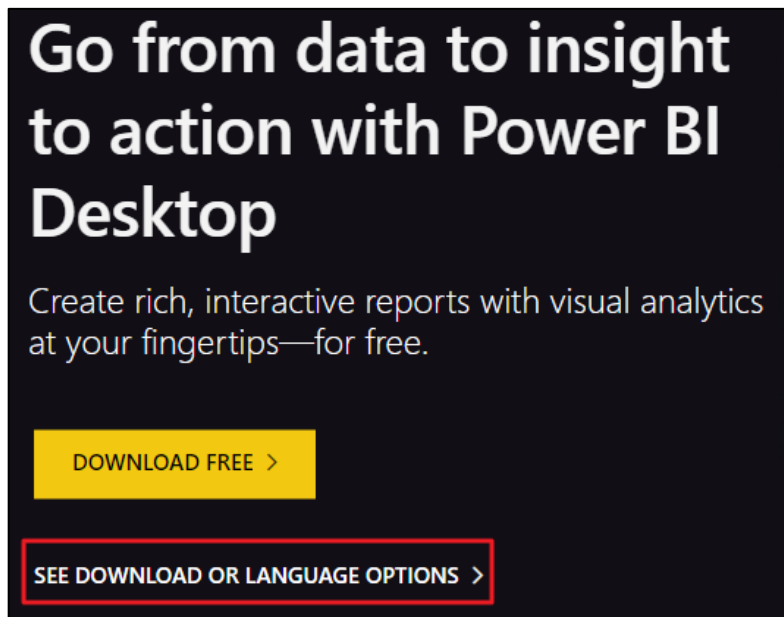


5. Exercise 3: Visualize data with Power BI Desktop

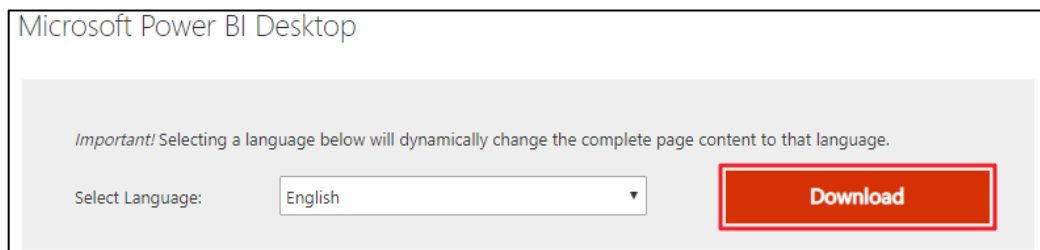
이 실습에서는 Data Warehouse의 데이터를 이용하여 Power BI Desktop과의 통합을 설정합니다.

5.1. Task 1: Install Power BI Desktop

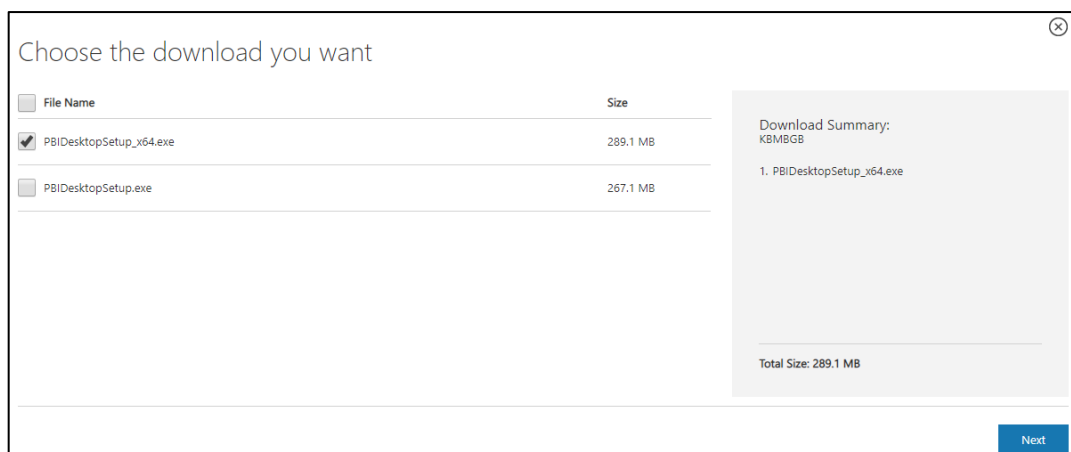
1. 웹 브라우저에서 Power BI Desktop 다운로드 페이지 (<https://powerbi.microsoft.com/en-us/desktop/>)로 이동합니다.
2. **SEE DOWNLOAD OR LANGUAGE OPTIONS** 링크를 선택합니다.



3. **Download** 버튼을 클릭합니다.



4. 파일을 선택 후 Next 버튼을 클릭합니다.

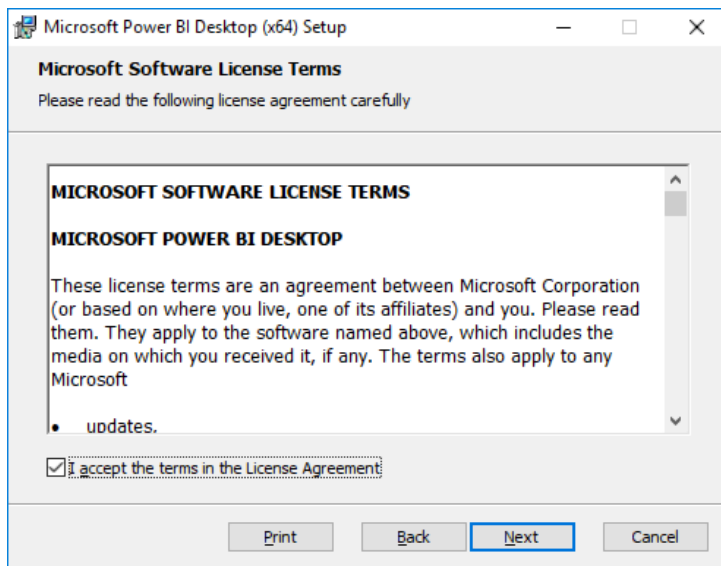


5. 설치파일을 실행합니다.

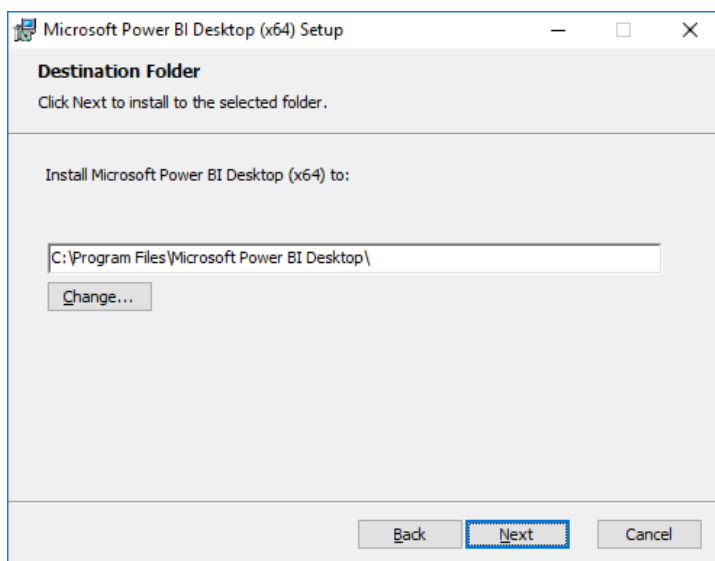
6. Welcome 화면에서 다음을 클릭합니다.



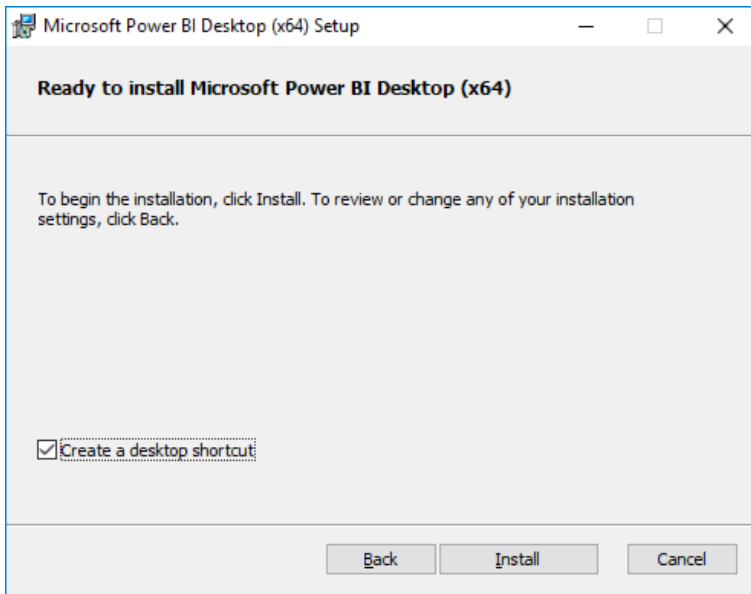
7. 사용권 계약에 동의하고 다음을 선택합니다.



8. 지정 폴더를 디폴트로 두고 다음을 선택합니다.



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

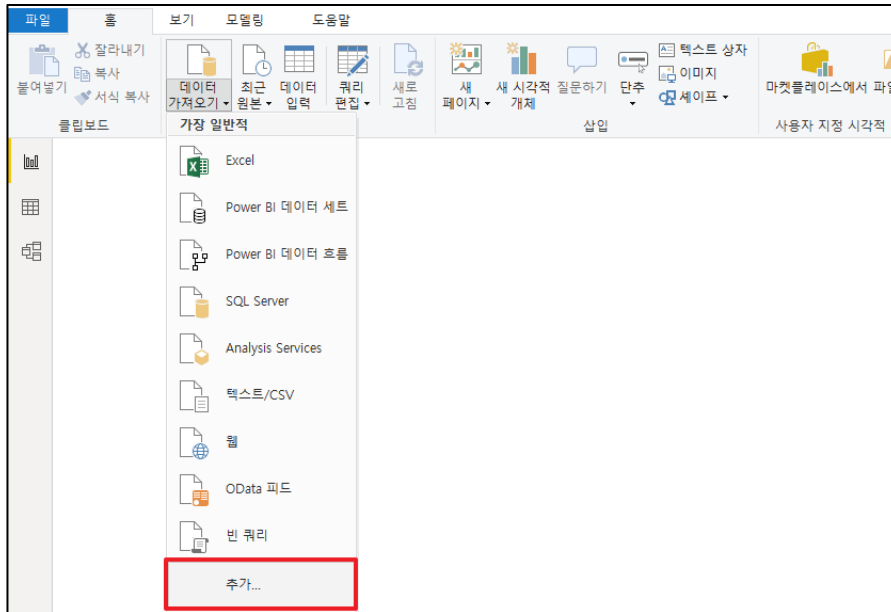


10. Microsoft Power BI Desktop 시작이 선택되어 있는지 확인하고 마침을 선택합니다.

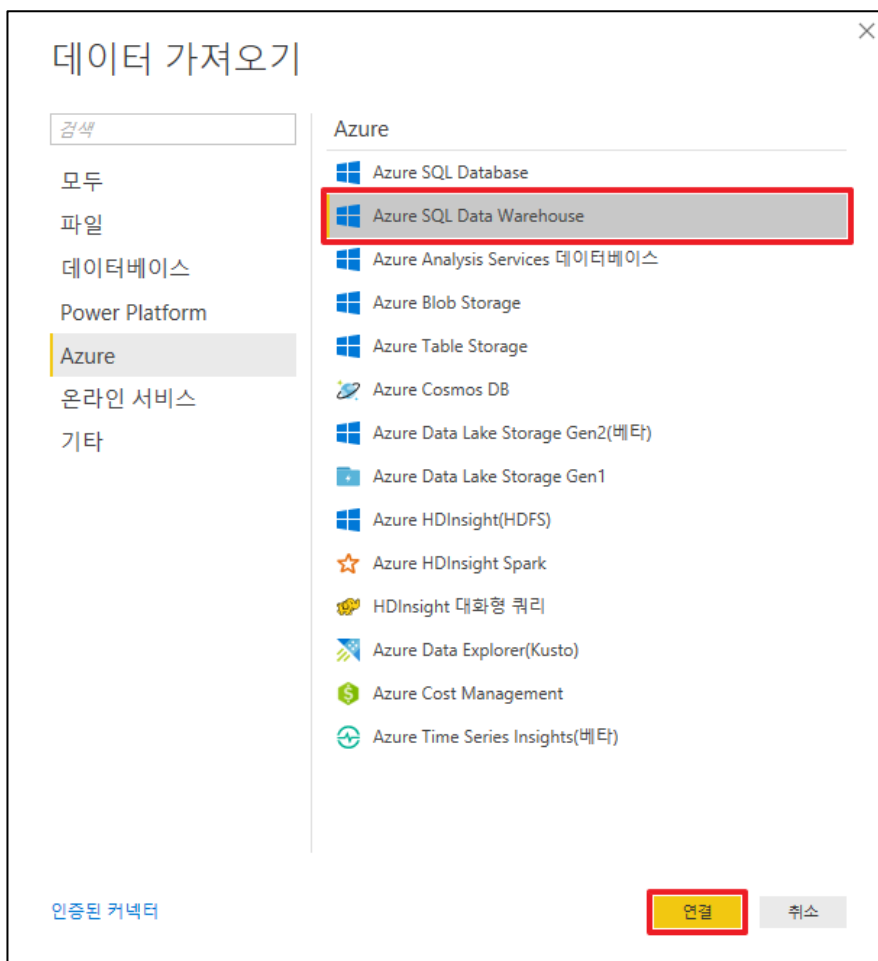


5.2. Task 2: Query data with Power BI Desktop

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.

SQL Server 데이터베이스

서버 ①
asodwserver1738.database.windows.net

데이터베이스(선택 사항)

데이터 연결 모드 ①
☒ 가져오기
☐ DirectQuery

▶ 고급 옵션

확인 취소

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

SQL Server 데이터베이스

Windows
데이터베이스
Microsoft 계정

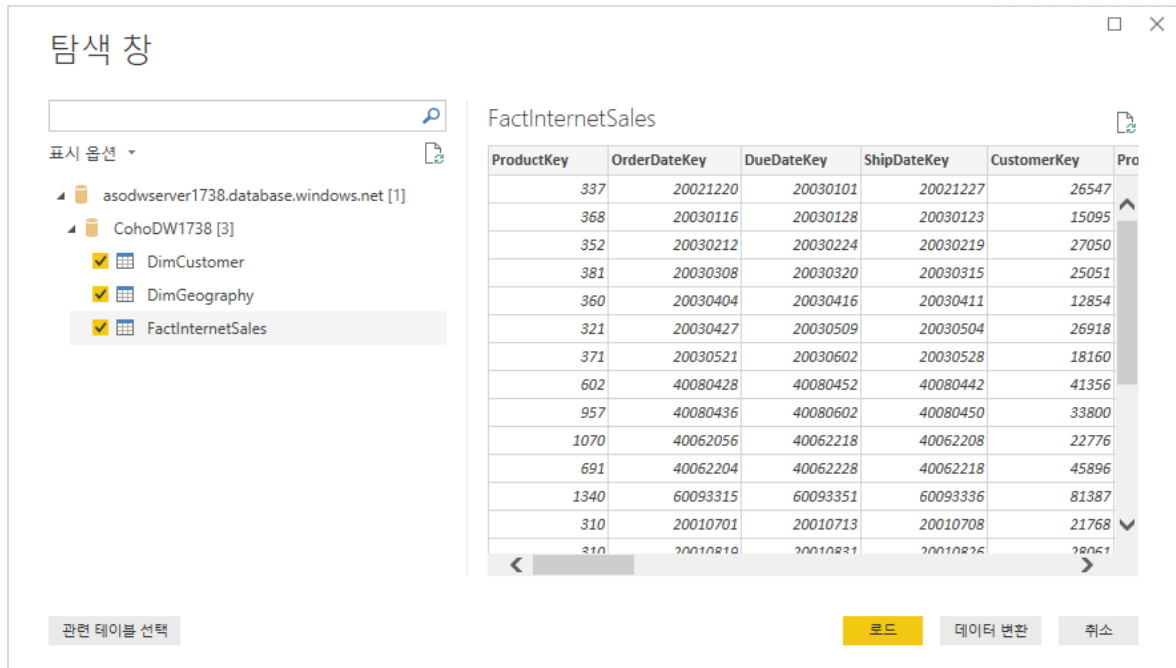
asodwserver1738.database.windows.net

사용자 이름
demouser

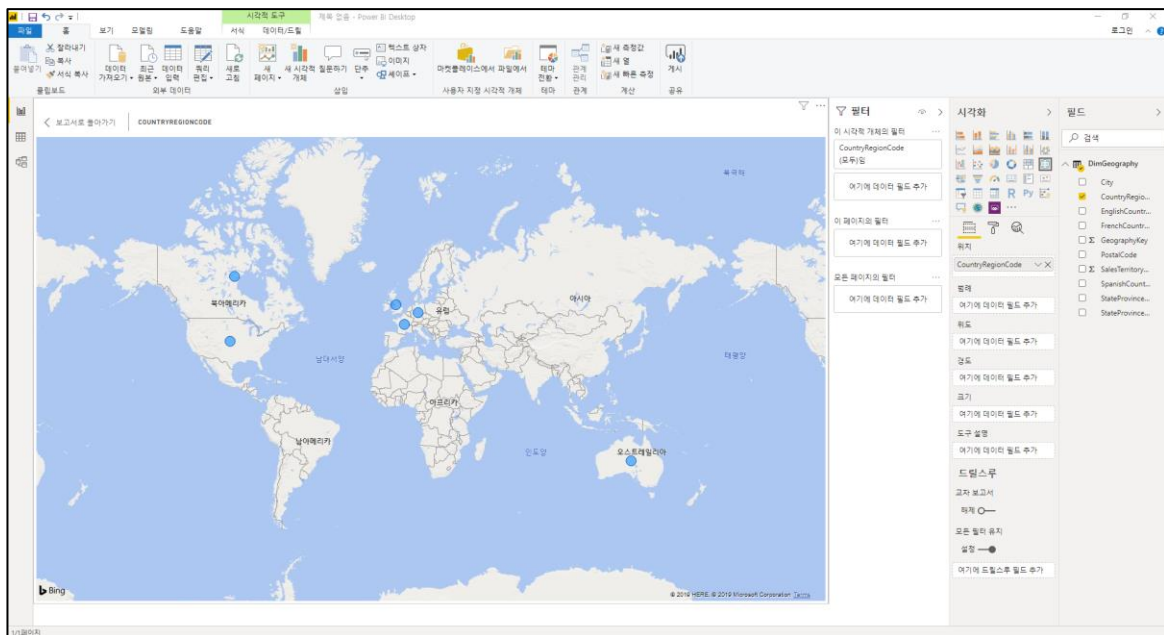
암호
●●●●●●●●

뒤로 연결 취소

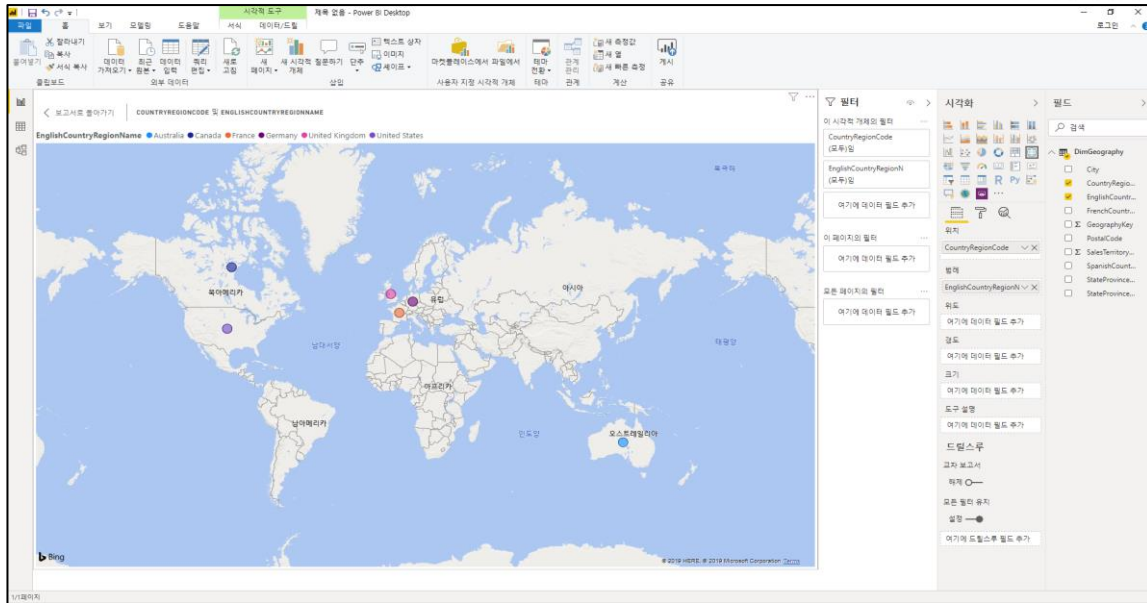
5. 모든 테이블을 선택한 후 **로드** 버튼을 클릭합니다.



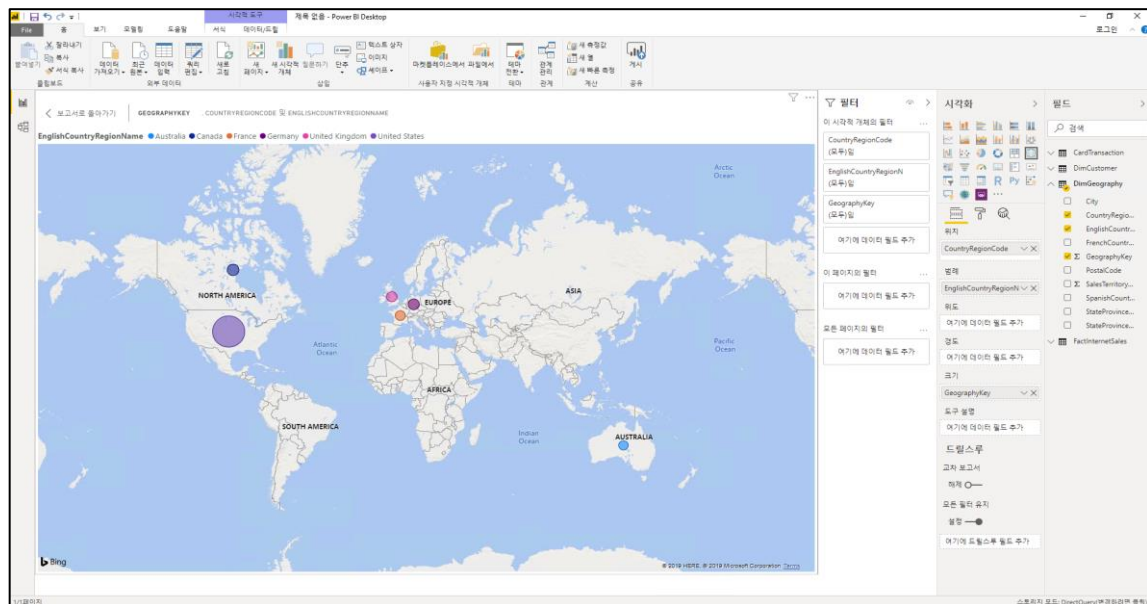
6. 오른쪽에 있는 필드 블레이드에서 DimGeography 테이블을 확장하고 CountryRegionCode 옆에 있는 상자를 선택합니다. Power BI는 지리 데이터를 인식하고 Map 시각화가 자동으로 시작됩니다.



7. 이 원이 실제로 무엇을 의미하는지에 대해 좀 더 구체적인 세부 사항을 보고 싶습니다. 국가/지역을 식별하는 범례를 추가해 봅시다. Legend 아래에서 EnglishCountryRegionName 필드를 드래그합니다.



8. 원의 크기를 변경시켜 보겠습니다. 수치를 가진 필드인 GeographyKey를 드래그해서 '크기' 필드에 추가해봅니다.



실습이 완료되었습니다. 이번 실습을 통해 Azure Data Factory를 이용하여 전체 데이터의 흐름을 관리하는 과정에 대해 알아보았습니다.

- ✓ Azure Data Factory를 이용한 각기 다른 위치의 Data 이동
- ✓ Azure Data Factory에서 Azure에서 제공되고 있는 서비스 연결

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