

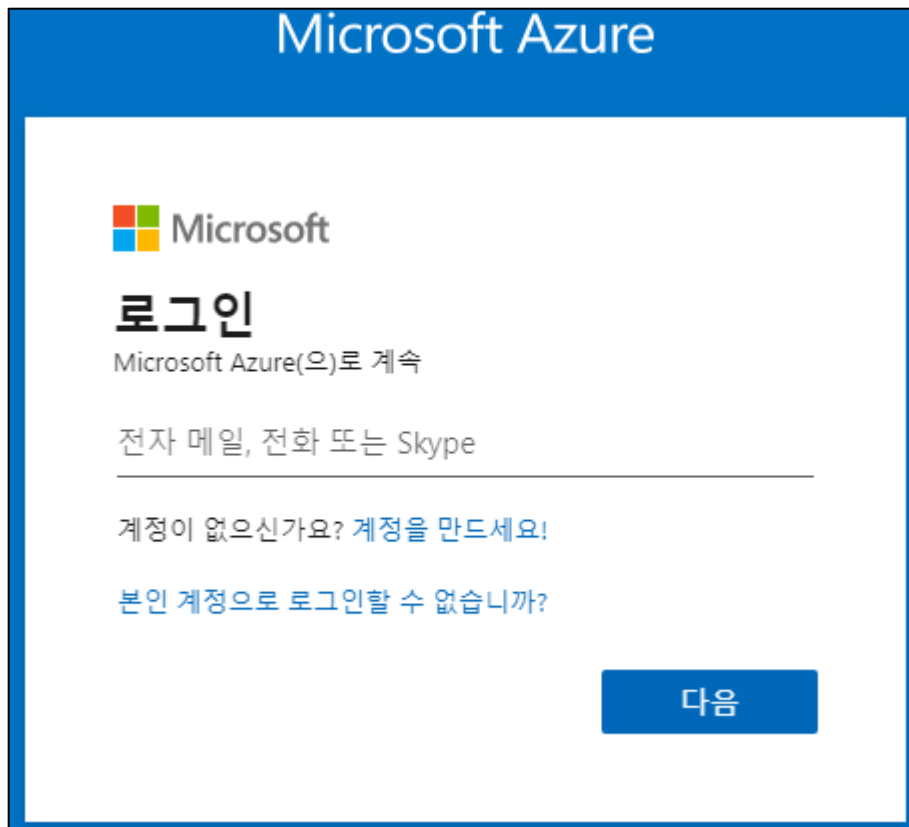
Lab 0 – 실습 환경 설정

요구 사항

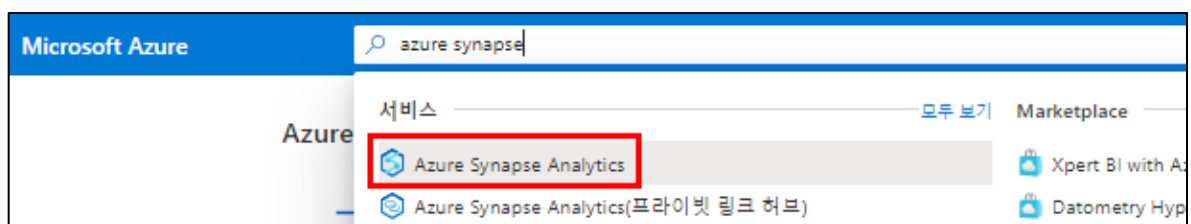
1. Microsoft Azure 구독 (Microsoft 이외의 구독은 유료 구독이어야 합니다.)

Task 1 : Azure Synapse Workspace 생성

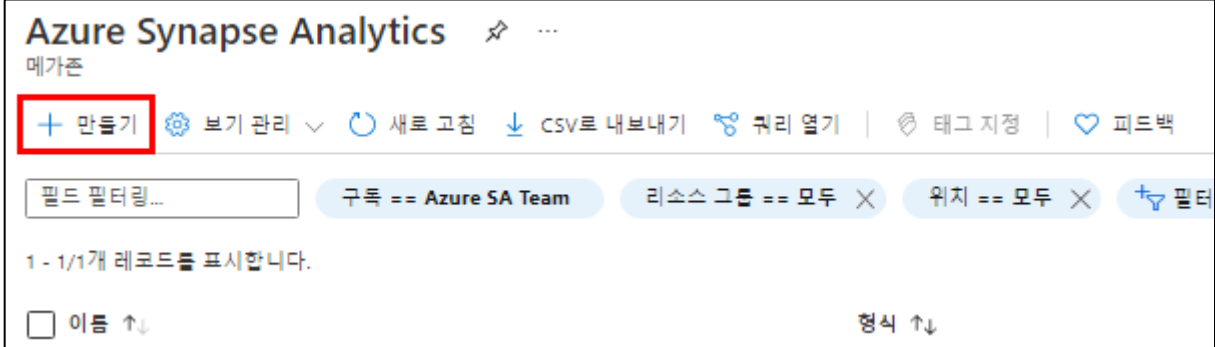
1. Azure Portal(<https://portal.azure.com/>)에 접속하여 개인 혹은 제공된 아이디로 로그인 합니다.



2. 상단의 돋보기에서 **Azure Synapse Analytics** 를 입력하여 선택합니다.

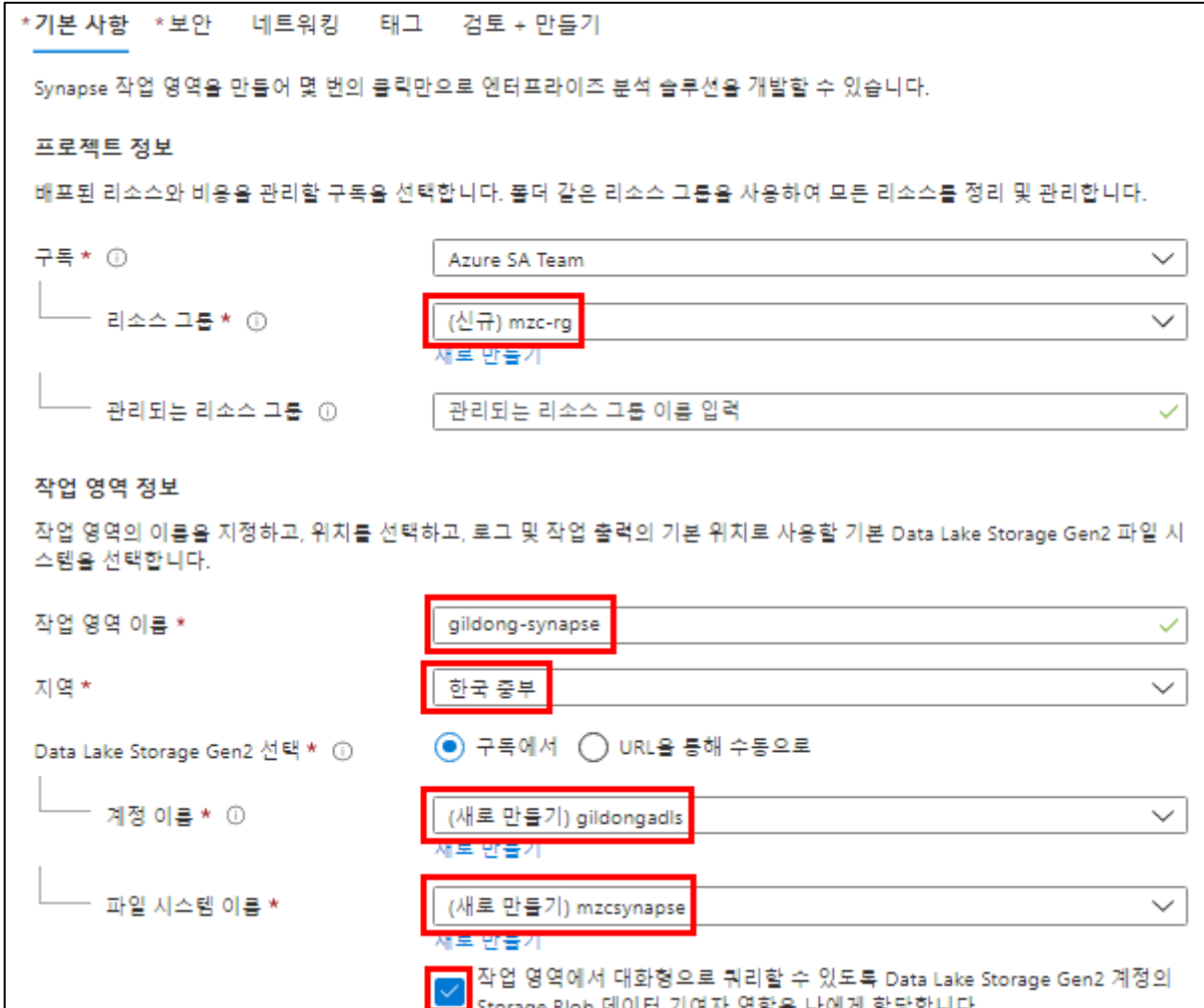


3. Azure Synapse Analytics **만들기**를 클릭하여 생성할 리소스의 정보를 입력합니다.



The screenshot shows the Azure Synapse Analytics '만들기' (Create) button highlighted in a red box. The interface includes a search bar, filters for 'Azure SA Team' and '모두' (All), and a list of records.

4. 기본 사항에 필요한 정보를 입력합니다.



The screenshot shows the '기본 사항' (Basic) tab of the Azure Synapse Analytics '만들기' (Create) form. The following fields are highlighted in red boxes:

- 구독 ***: Azure SA Team
- 리소스 그룹 ***: (신규) mzc-rg
- 작업 영역 이름 ***: gildong-synapse
- 지역 ***: 한국 중부
- Data Lake Storage Gen2 계정 이름 ***: (새로 만들기) gildongadls
- 파일 시스템 이름 ***: (새로 만들기) mzc-synapse

Additional information visible in the form includes the '작업 영역에서 대화형으로 쿼리할 수 있도록 Data Lake Storage Gen2 계정의 Storage Blob 데이터 기여자 역할을 나에게 할당합니다.' checkbox, which is checked.

리소스 그룹 : mzc-rg (새로 만들기)

작업 영역 이름 : <username>-synapse (고유 이름값)

지역 : 한국 중부

Data Lake Storage Gen2 계정 이름 : <username>adls (새로 만들기)(고유 이름값)
Data Lake Storage Gen2 파일 시스템 이름 : mzcsynapse (새로 만들기)

5. 보안에서 **SQL admin 관리자 계정 정보를 입력합니다.**

*기본 사항
*보안
네트워킹
태그
검토 + 만들기

작업 영역의 보안 옵션을 구성합니다.

SQL 관리자 자격 증명

작업 영역의 SQL 풀에 대한 관리자 액세스에 사용할 수 있는 자격 증명을 제공합니다. 암호를 제공하지 않으면 자동으로 생성됩니다. 나중에 암호를 변경할 수 있습니다.

SQL Server 관리자 로그인 * ①

SQL 암호 ①

암호 확인

시스템이 할당한 관리 ID 권한

작업 영역의 시스템이 할당한 ID에 할당하려는 권한을 선택합니다. [자세한 정보](#)

☒ 파이프라인(작업 영역의 시스템 할당 ID로 실행됨)에서 SQL 풀에 액세스할 수 있도록 허용합니다. ①

☐ Data Lake Storage Gen2 계정에 네트워크 액세스를 허용합니다. ①

① 선택한 Data Lake Storage Gen2 계정이 네트워크 액세스 규칙을 사용하여 네트워크 액세스를 제한하지 않거나, 기본 템 아래의 URL을 통해 스토리지 계정을 수동으로 선택했습니다. [자세한 정보](#)

작업 영역 암호화

⚠ 작업 영역을 만들 때 고객 관리형 키 사용에 동의한 후에는 이중 암호화 구성을 변경할 수 없습니다.

사용자가 관리하는 키(고객 관리형 키)를 사용하여 작업 영역의 미사용 데이터를 모두 암호화하도록 선택합니다. 그러면 플랫폼 관리형 키를 사용하는 인프라 계층에서 암호화를 사용한 이중 암호화가 제공됩니다. [자세한 정보](#)

고객 관리형 키를 사용한 이중 암호화 ☐ 사용 ☒ 사용 안 함

SQL Server 관리자 로그인 : sqladminuser

SQL 암호 : 각자 입력

6. 검토 + 만들기를 클릭 후 하단의 **만들기**로 생성합니다.

Task 2 : Machine Learning 생성

-

- 기계 학습
메가존

+ 만들기
보기 관리
새로 고침
CSV로 내보내기
쿼리 열기
태그 지정
삭제

필드 필터링...
구독 == Azure SA Team
리소스 그룹 == 모두
위치 == 모두

1 - 1/1개 레코드를 표시합니다.

☐ 이름 ↑↓
리소스 그룹

3. 기본에서 필요한 정보를 입력합니다.

기본 네트워킹 고급 태그 검토 + 만들기

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

구독 * ① Azure SA Team ▼

리소스 그룹 * ① **mzc-rg** ▼
새로 만들기

작업 영역 세부 정보
 작업 영역의 이름 및 지역을 지정하세요.

작업 영역 이름 * ① **mzcmlworkspace** ✓

지역 * ① 한국 중부 ▼

스토리지 계정 * ① (신규) mzcmlworkspace2774648598 ▼
새로 만들기

키 자격 증명 모음 * ① (신규) mzcmlworkspace2409213624 ▼
새로 만들기

Application Insights * ① (신규) mzcmlworkspace2243890366 ▼
새로 만들기

컨테이너 레지스트리 * ① 없음 ▼
새로 만들기

리소스 그룹 : mzc-rg

작업 영역 이름 : mzcmlworkspace

지역 : 한국 중부

스토리지 계정 : 작업 영역 이름 입력 시 자동 신규 생성

키 자격 증명 모음 : 작업 영역 이름 입력 시 자동 신규 생성

Application Insights : 작업 영역 이름 입력 시 자동 신규 생성

컨테이너 레지스트리 : 없음

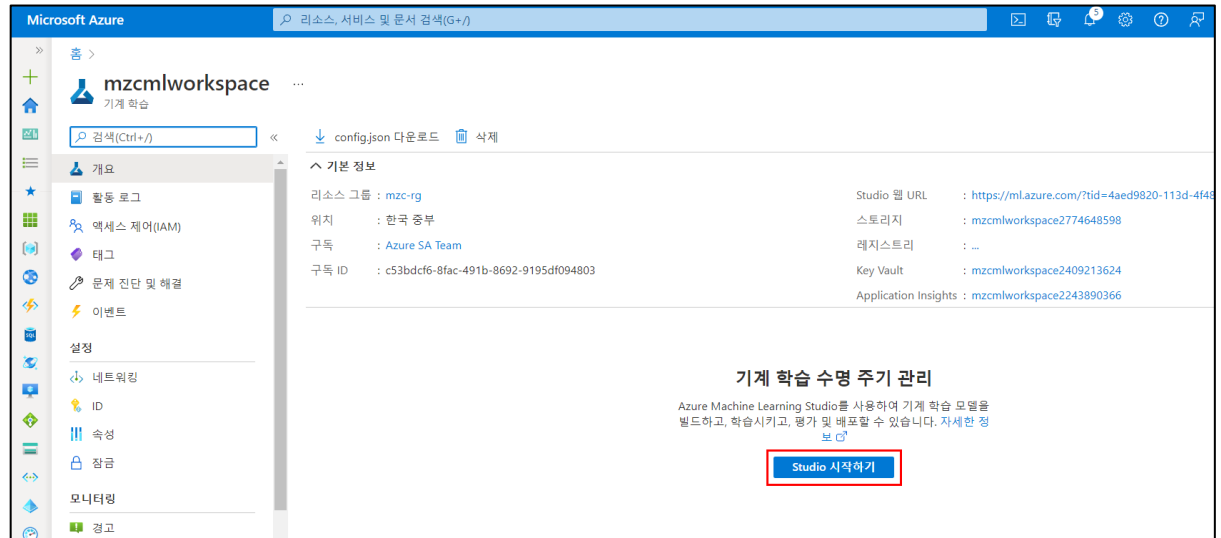
4. 검토 + 만들기를 클릭 후 하단의 만들기로 생성합니다.

✓ 유효성 검사 통과

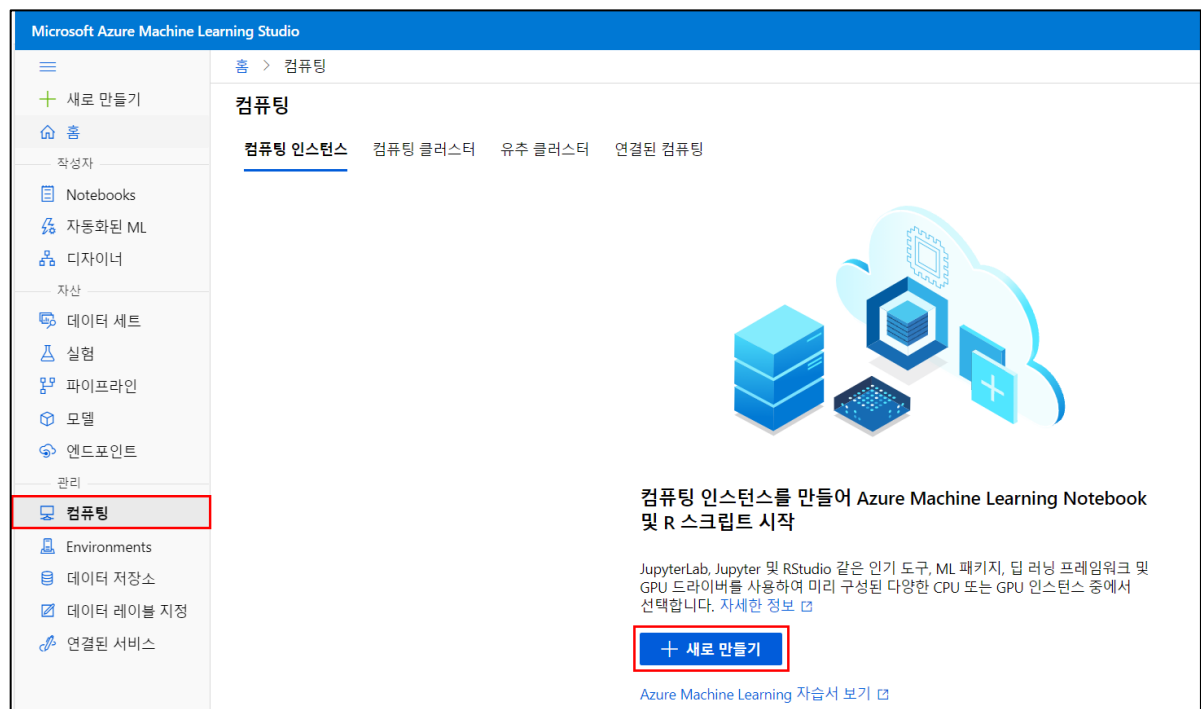
기본 네트워킹 고급 태그 **검토 + 만들기**

* 빨간 박스 외에 리소스 설정은 기본 값이므로 수정하지 마세요.

5. 방금 생성했던 **ML(Machine Learning)** 워크스페이스로 검색을 통해 이동하여 **Studio** 시작하기를 클릭합니다.



6. 접속한 스튜디오에서 왼쪽 블레이드의 **컴퓨팅**으로 이동합니다. 아래와 같이 페이지가 뜨면 **새로 만들기**를 클릭합니다.



7. 컴퓨팅 이름, 가상머신 유형, 가상 머신 크기를 선택합니다. 만들기를 클릭합니다.

필수 설정

고급 설정

필수 설정 구성
컴퓨팅 인스턴스에 사용할 이름 및 가상 머신 크기를 선택합니다. 컴퓨팅 인스턴스는 공유할 수 없습니다. 게 할당되며 고급 설정 섹션에서 다른 사용자로 변경할 수 있습니다.

컴퓨팅 이름 *

gildongmlcompute

위치

koreacentral

가상 머신 유형

CPU GPU

가상 머신 크기

권장 옵션에서 선택 모든 옵션에서 선택

사용 가능한 총 할당량: 코어 296개

이름 ↑	범주	워크로드 유형	사용 가능한 할...	비용
Standard_DS11_v2 코어 2개, 14GB RAM, 28GB 스토리지	메모리 최적화	Notebook(또는 다른 IDE)에서 개발 및 경량 테스트	코어 296개	\$0.20/...
Standard_DS3_v2 코어 4개, 14GB RAM, 28GB 스토리지	범용	클래식 ML 모델 학습, AutoML 실행, 파이프라인 실행(기본 컴퓨팅)	코어 296개	\$0.33/...
Standard_DS12_v2 코어 4개, 28GB RAM, 56GB 스토리지	메모리 최적화	대규모 데이터 세트(1GB 초과) 병렬 실행 단계, 일괄 처리 추론에 대한 학습	코어 296개	\$0.40/...
Standard_F4s_v2 코어 4개, 8GB RAM, 32GB 스토리지	컴퓨팅 최적화	실시간 추론 및 기타 대기 시간이 중요한 작업	코어 296개	\$0.19/...

만들기

뒤로

다음: 고급 설정

취소

컴퓨팅 이름: <username>mlcompute

가상 머신 유형: CPU

가상 머신 크기: Standard_DS3_v2

8. 컴퓨팅 인스턴스가 실행 중으로 표시됩니다. 과금 방지를 위해 인스턴스를 선택하고 상단의 중지 버튼을 눌러 인스턴스를 중지시켜 놓습니다.

컴퓨팅 인스턴스

컴퓨팅 클러스터

유추 클러스터

연결된 컴퓨팅

+

새로 만들기

↺

새로 고침

▶

시작

⏻

중지

↻

다시 시작

🗑

삭제

📄

열 편집

🔍

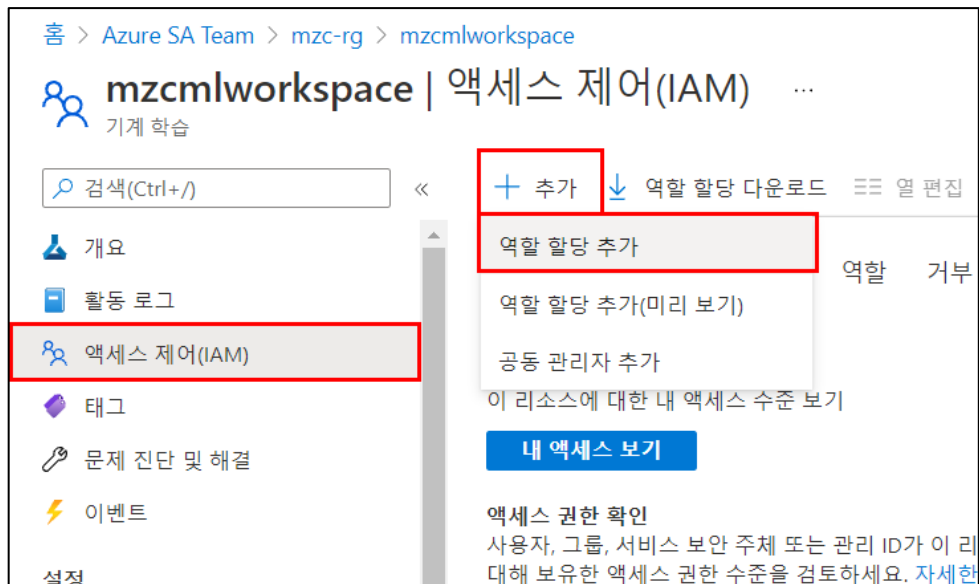
보기 다시 설정

🔍

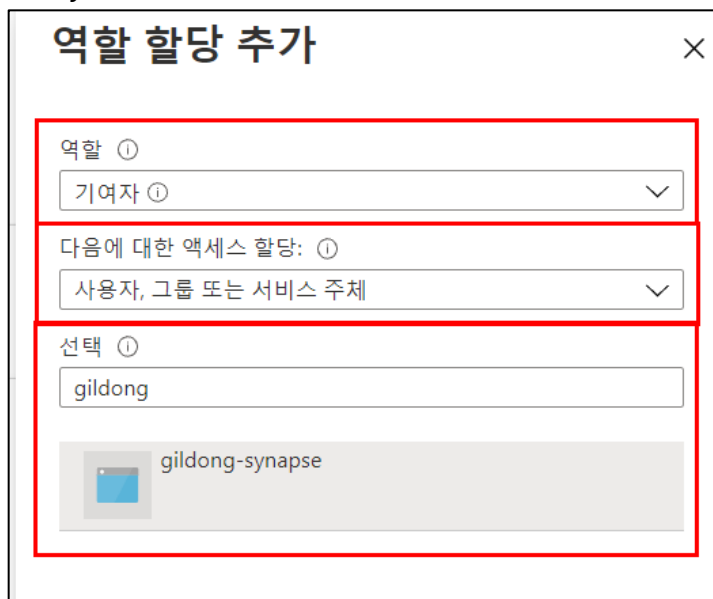
검색

이름	상태	애플리케이션
gildongmlcompute	▶ 실행 중	JupyterLab Jupyter VS Code RStudio 터미널

9. Studio를 닫고 **ML(Machine Learning)** 워크스페이스로 이동합니다. 왼쪽 블레이드의 **액세스 제어(IAM)**으로 이동하여 **역할 할당**을 추가합니다.

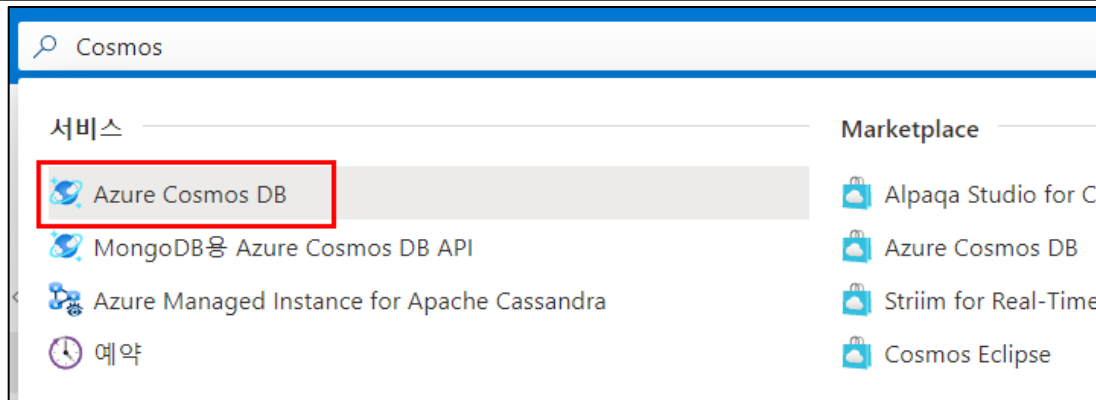


10. **역할**은 기여자, **선택** 부분에 Synapse Workspace의 이름을 검색합니다. Synapse Analytics를 선택합니다. **저장**을 눌러 역할을 할당합니다.

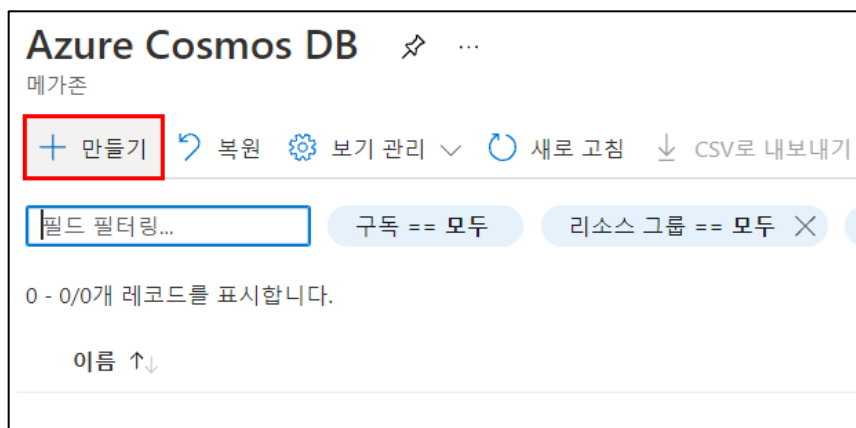


Task 3 : Cosmos DB 생성

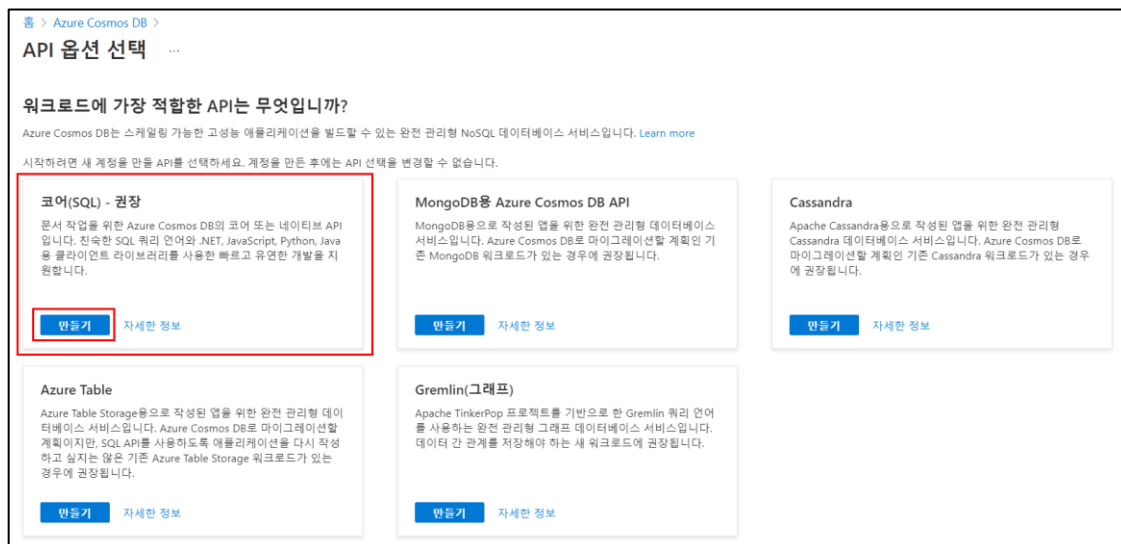
1. 상단의 돋보기에서 **Comsmos DB** 를 입력하여 선택합니다.



2. **Azure Cosmos DB 만들기를 클릭하여** 생성할 리소스의 정보를 입력합니다.



3. API 중 **코어(SQL) API** 를 선택합니다.



4. 기본에서 필요한 정보를 입력합니다. 검토 + 만들기를 클릭 후 하단의 만들기로 생성합니다.

Azure Cosmos DB 계정 만들기 - 코어(SQL) ...

Azure Cosmos DB는 확장 가능한 고성능 애플리케이션을 구축하기 위한 완전 관리형 NoSQL 데이터베이스 서비스입니다. 는 프로덕션 환경을 이용하세요. [자세한 정보](#)

프로젝트 세부 정보

배포된 리소스 및 비용을 관리할 구독을 선택합니다. 폴더와 같은 리소스 그룹을 사용하여 모든 리소스를 구성하고 관리할

구독 *

Azure SA Team

리소스 그룹 *

mzc-rg

[새로 만들기](#)

인스턴스 세부 정보

계정 이름 *

gildongcosmos

위치 *

(Asia Pacific) 한국 중부

용량 모드 ①

☒ 프로비저닝된 처리량
 ☐ Serverless

[용량 모드에 대한 자세한 정보](#)

With Azure Cosmos DB free tier, you will get the first 1000 RU/s and 25 GB of storage for free in an account. You can ena

무료 계층 할인 적용

☒ 적용
 ☐ 적용 안 함

검토 + 만들기

이전

다음: 전역 배포

리소스 그룹 : mzc-rg

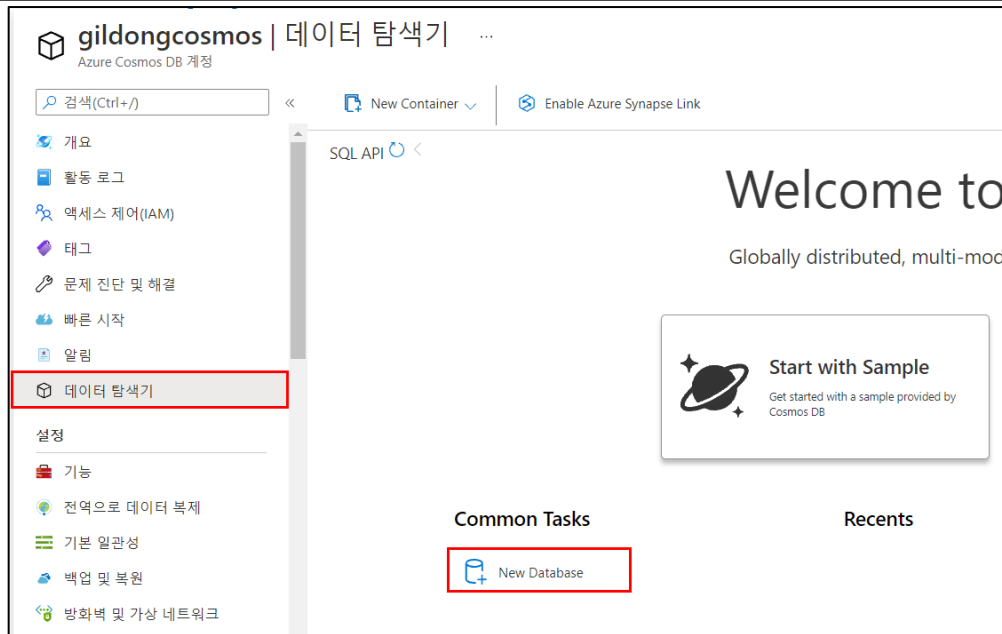
리소스 이름 : <username>cosmos (고유 이름값)

지역 : 한국 중부

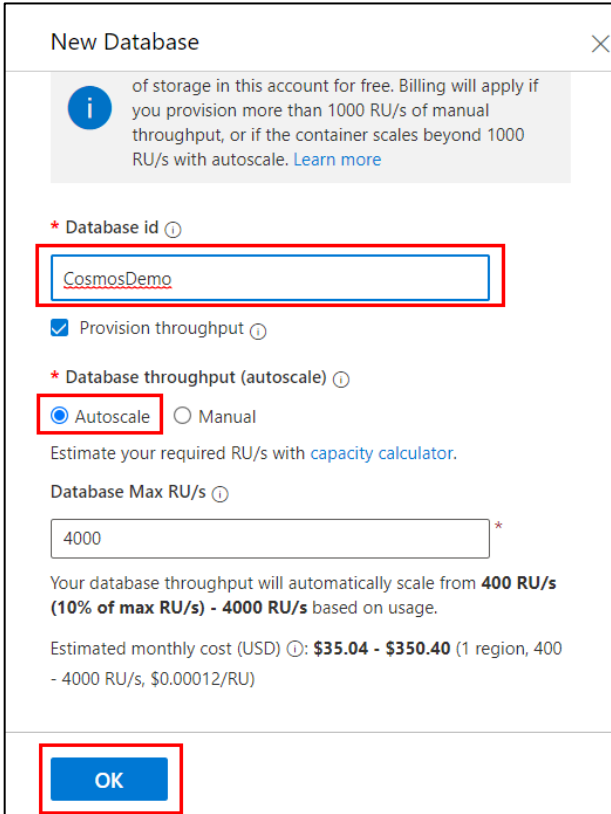
용량 모드 : 프로비저닝된 처리량

무료 계층 할인 적용 : 적용

5. 만들어진 **Cosmos DB 계정**으로 이동합니다. 왼쪽 메뉴에서 **데이터 탐색기**를 선택합니다. **New Database**를 클릭합니다.



6. **Database id**를 아래와 같이 입력합니다. **Autoscale**을 설정하고 **Database Max RU/s**는 기본값인 4000으로 설정합니다. **OK**를 눌러 데이터베이스를 생성합니다.



New Database

of storage in this account for free. Billing will apply if you provision more than 1000 RU/s of manual throughput, or if the container scales beyond 1000 RU/s with autoscale. [Learn more](#)

* Database id ⓘ

CosmosDemo

☒ Provision throughput ⓘ

* Database throughput (autoscale) ⓘ

☒ Autoscale ☐ Manual

Estimate your required RU/s with [capacity calculator](#).

Database 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.

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

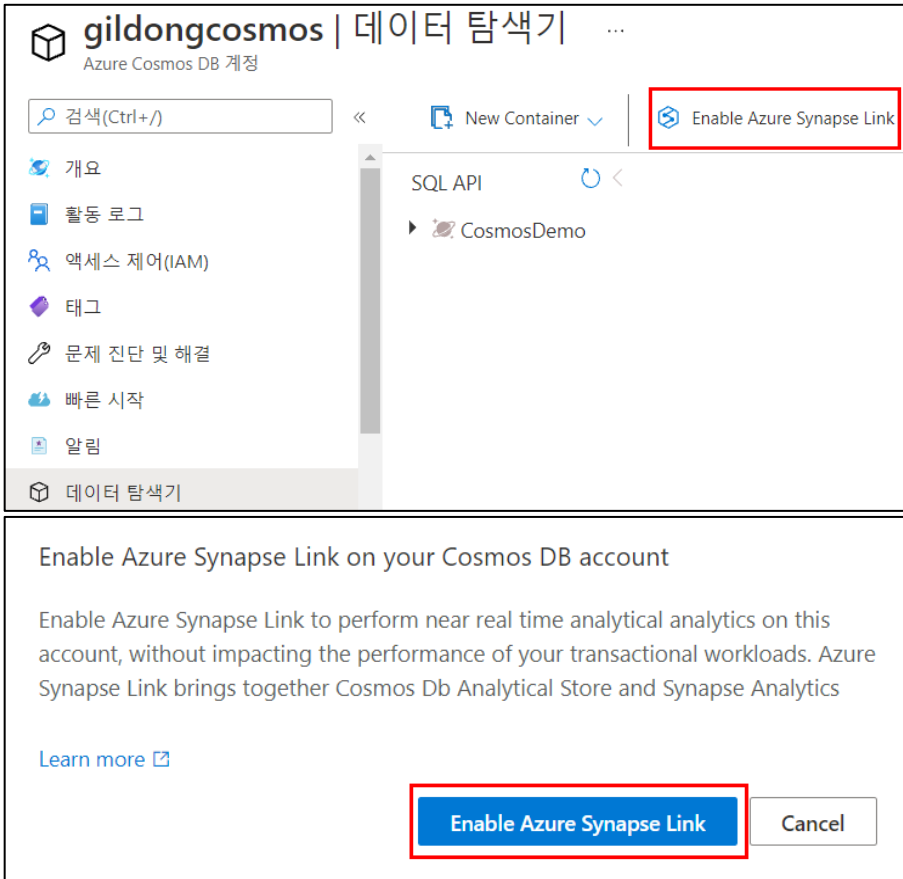
OK

Database id: CosmosDemo

Database Max RU/s : 4000

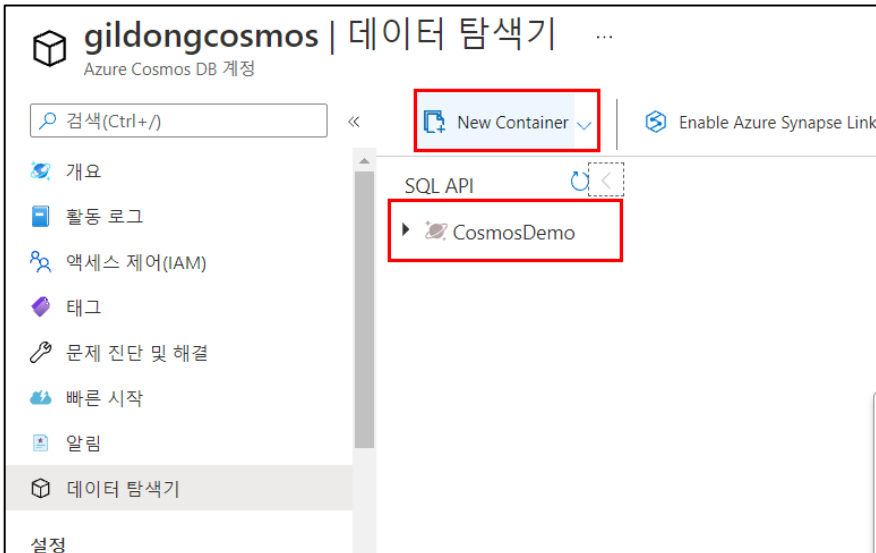
7. 아래와 같이 **데이터베이스가 생성된 것**을 볼 수 있습니다. 상단의 **Enable Azure Synapse**

Link를 클릭하여 Synapse Link를 허용합니다. (시간이 조금 소요됩니다)



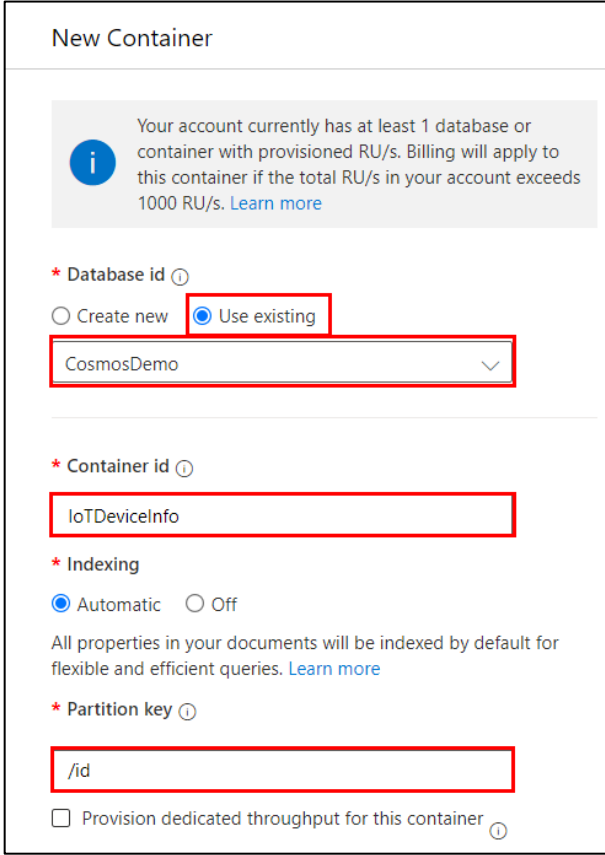
The screenshot shows the Azure Cosmos DB portal for the account 'gildongcosmos'. The left sidebar contains navigation links: 개요, 활동 로그, 액세스 제어(IAM), 태그, 문제 진단 및 해결, 빠른 시작, 알림, and 데이터 탐색기. The main area displays the 'SQL API' section with a 'CosmosDemo' container. A red box highlights the 'Enable Azure Synapse Link' button in the top right corner. Below the main area, a modal dialog is open with the title 'Enable Azure Synapse Link on your Cosmos DB account'. The dialog contains text explaining that enabling Synapse Link allows for near real-time analytical analytics without impacting transactional workloads. At the bottom of the dialog, there is a blue 'Enable Azure Synapse Link' button (highlighted with a red box) and a 'Cancel' button.

8. 상단의 **New Container**를 클릭하여 컨테이너를 생성합니다.



The screenshot shows the same Azure Cosmos DB portal interface. A red box highlights the 'New Container' button in the top navigation bar. Another red box highlights the 'CosmosDemo' container in the main area. The 'SQL API' section is also visible.

9. **Database id**는 **Use existing**을 선택하여 기존에 생성한 **CosmosDemo**를 선택합니다. **Container id**와 **Partition key**를 아래와 같이 생성합니다.



New Container

Your account currently has at least 1 database or container with provisioned RU/s. Billing will apply to this container if the total RU/s in your account exceeds 1000 RU/s. [Learn more](#)

*** Database id** ⓘ

☐ Create new ☒ Use existing

CosmosDemo

*** Container id** ⓘ

IoTDeviceInfo

*** Indexing**

☒ Automatic ☐ Off

All properties in your documents will be indexed by default for flexible and efficient queries. [Learn more](#)

*** Partition key** ⓘ

/id

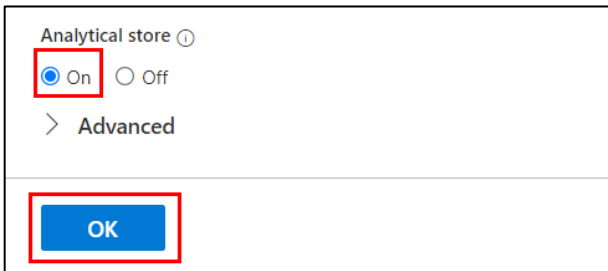
☐ Provision dedicated throughput for this container ⓘ

Database id : CosmosDemo

Container id : IoTDeviceInfo

Partition key : /id

10. **Analytical store**를 **On**으로 설정합니다. **OK**를 클릭하여 컨테이너를 선택합니다.



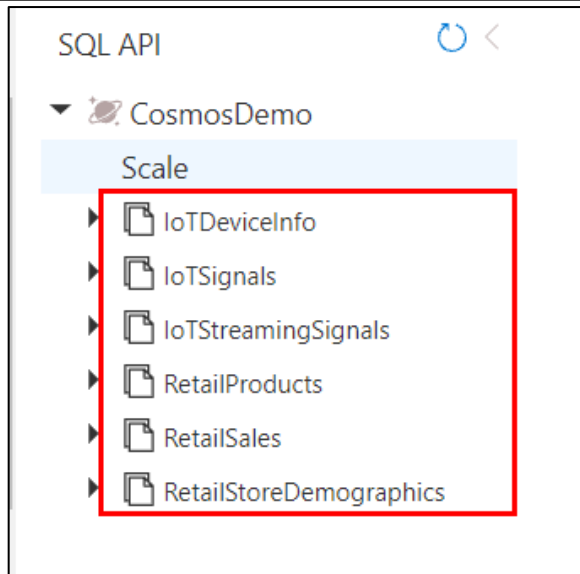
Analytical store ⓘ

☒ On ☐ Off

> Advanced

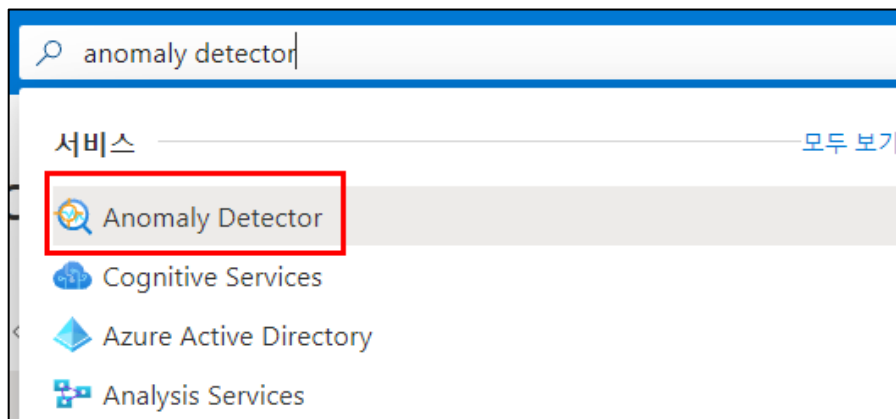
OK

11. 같은 방법으로 **IoTSignals**, **IoTStreamingSignals**, **RetailProducts**, **RetailSales**, **RetailStoreDemographics** 컨테이너를 생성합니다. **Partition Key**는 동일하게 **/id**를 입력합니다. **Analytical Store**의 경우, **IoTStreamingSignals**만 **Off**를 선택합니다. 나머지는 **On**을 선택합니다. 최종 모습은 아래와 같습니다.

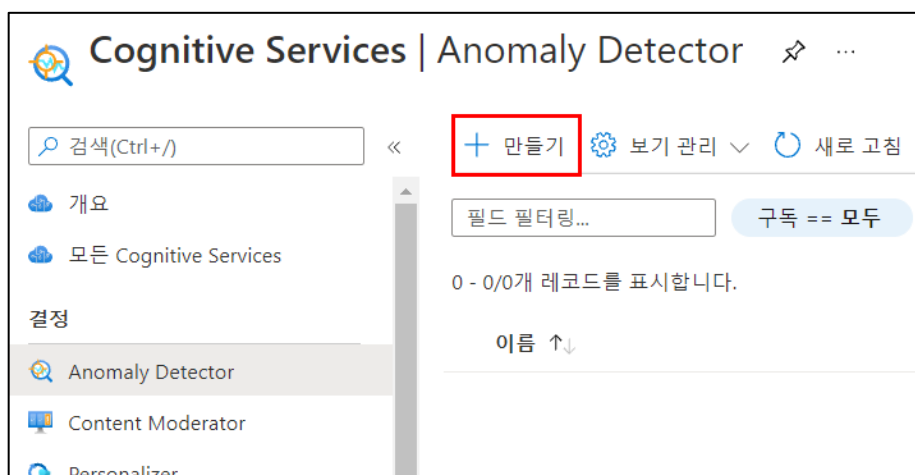


Task 4 : Anomaly Detector (Cognitive Service) 생성

1. 상단의 돋보기에서 **Anomaly Detector** 를 입력합니다.



2. **+만들기**를 클릭하여 생성합니다.



3. 기본에서 필요한 정보를 입력합니다. 검토 + 만들기를 클릭 후 하단의 만들기로 생성합니다.

프로젝트 정보

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

구독 * ①

Azure SA Team

리소스 그룹 * ①

mzc-rg

새로 만들기

인스턴스 정보

지역 * ①

한국 중부

이름 * ①

gildonganomalydetect

가격 책정 계층 * ①

무료 F0(호출 10회/초, 트랜잭션 20,000개/월)

[전체 가격 책정 세부 정보 보기](#)

검토 + 만들기

< 이전

다음: 가상 네트워크 >

리소스 그룹 : mzc-rg

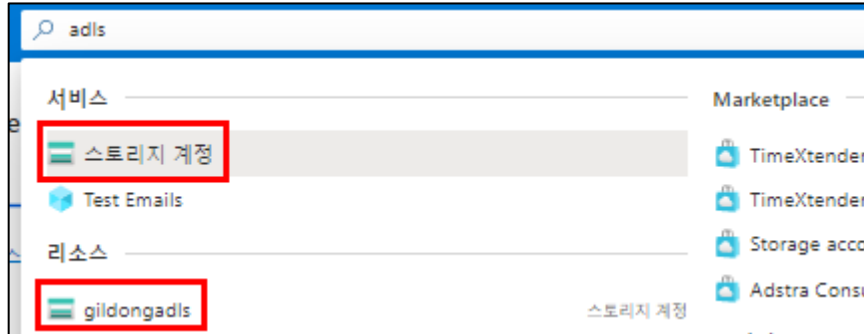
지역 : 한국 중부

리소스 이름: <username>anomalydetect

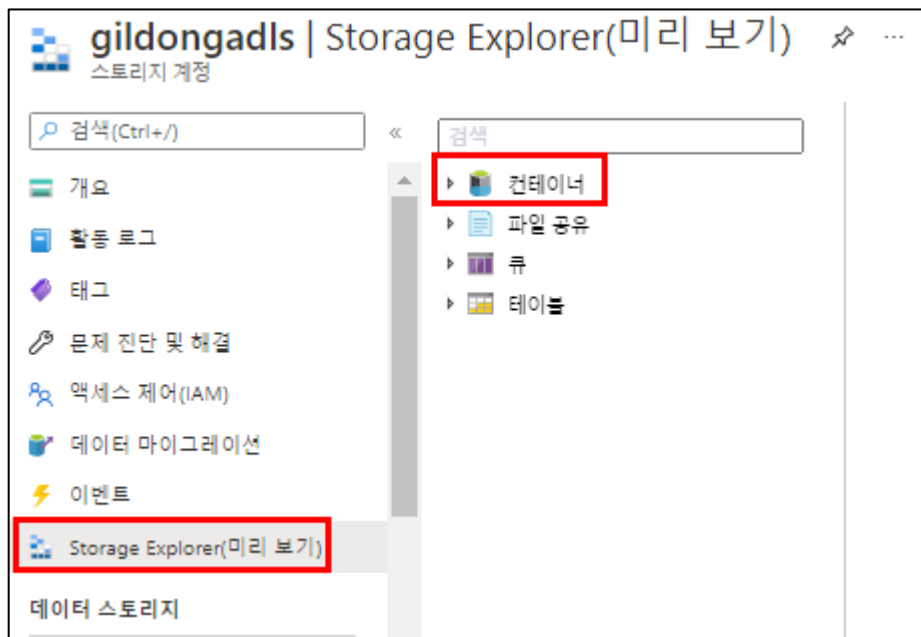
가격 책정 계층 : 무료 F0

Task 5 : ADLS Gen2 에 컨테이너 및 데이터 생성

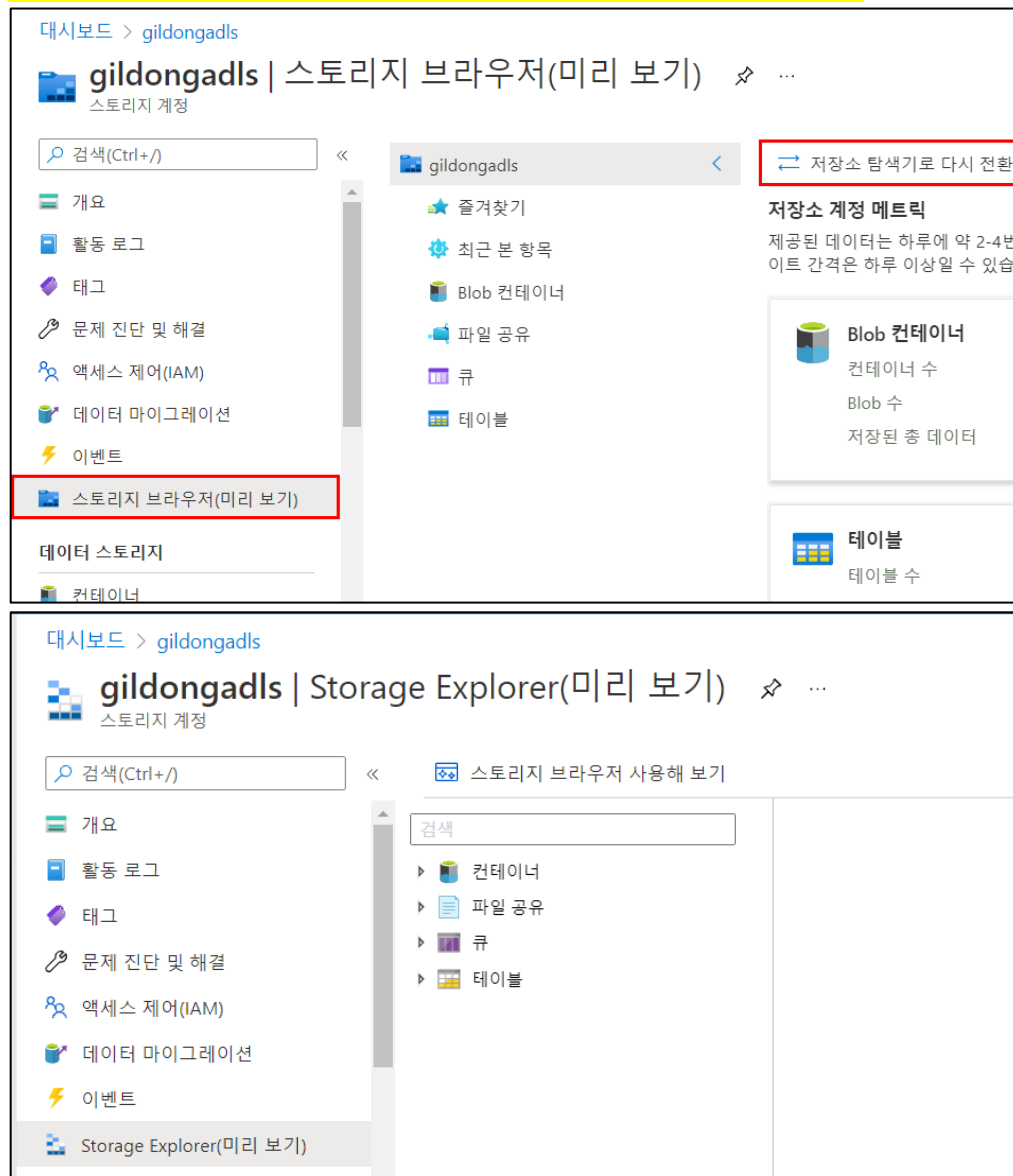
1. 상단의 돋보기에서 **ADLS** 를 입력하여 앞에서 생성한 <username>adls 를 선택합니다.



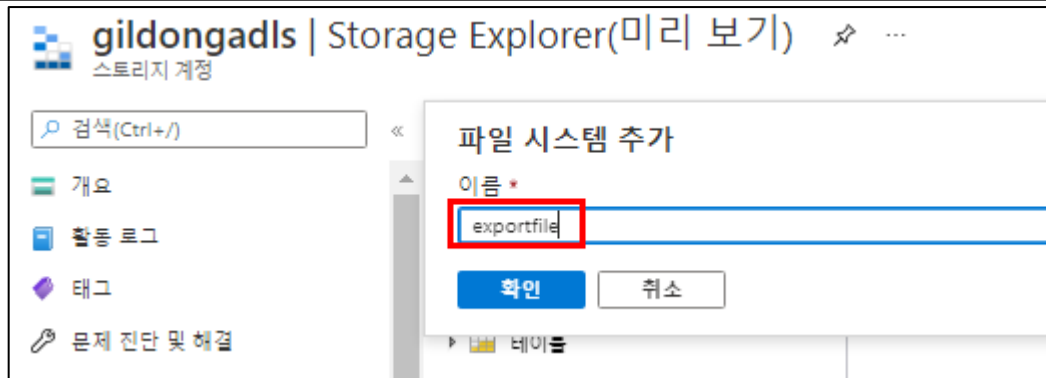
2. 왼쪽 블레이드에서 **Storage Explorer** 를 클릭하고, 컨테이너 클릭합니다. 앞에서 생성한 mzcynapse 파일 시스템이 보이는지 확인합니다.



*아래와 같이 스토리지 브라우저(미리 보기)가 적용되는 경우, 스토리지 브라우저 클릭 후 상단의 저장소 탐색기로 다시 전환을 클릭합니다.



3. 컨테이너에서 마우스 오른쪽 클릭을 하여 파일 시스템 만들기를 클릭합니다.
4. 컨테이너를 추가하기 위해 이름을 입력하고 확인을 클릭합니다. (4 개의 컨테이너를 추가합니다.)



exportfile

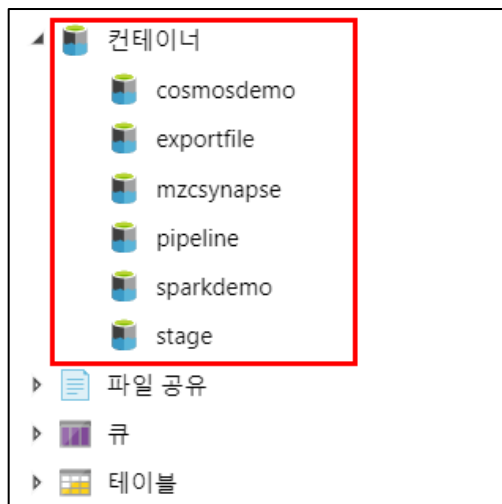
pipeline

stage

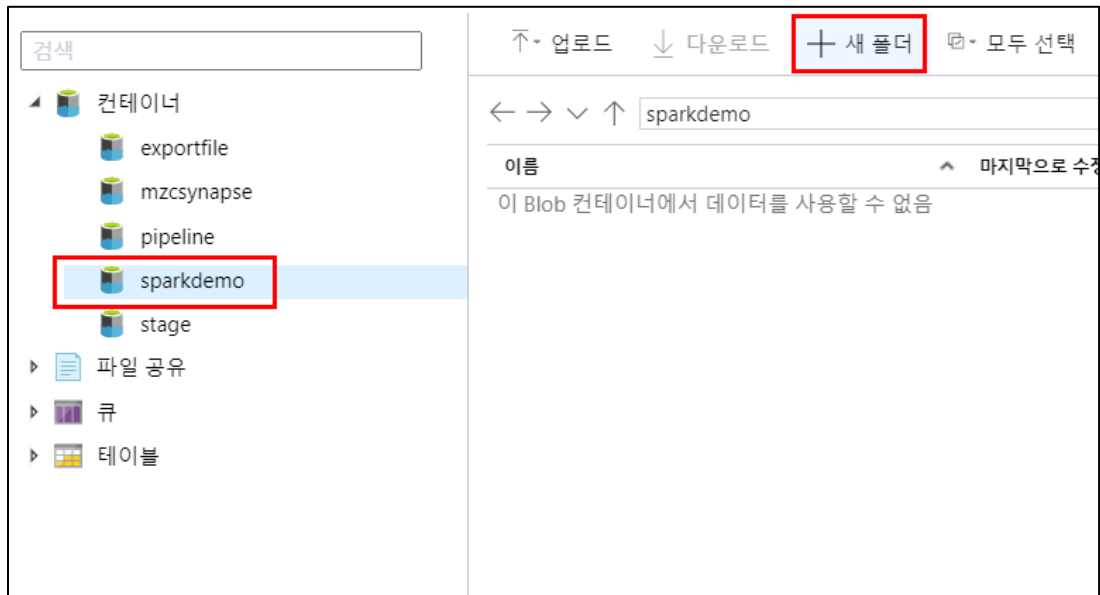
sparkdemo

cosmosdemo

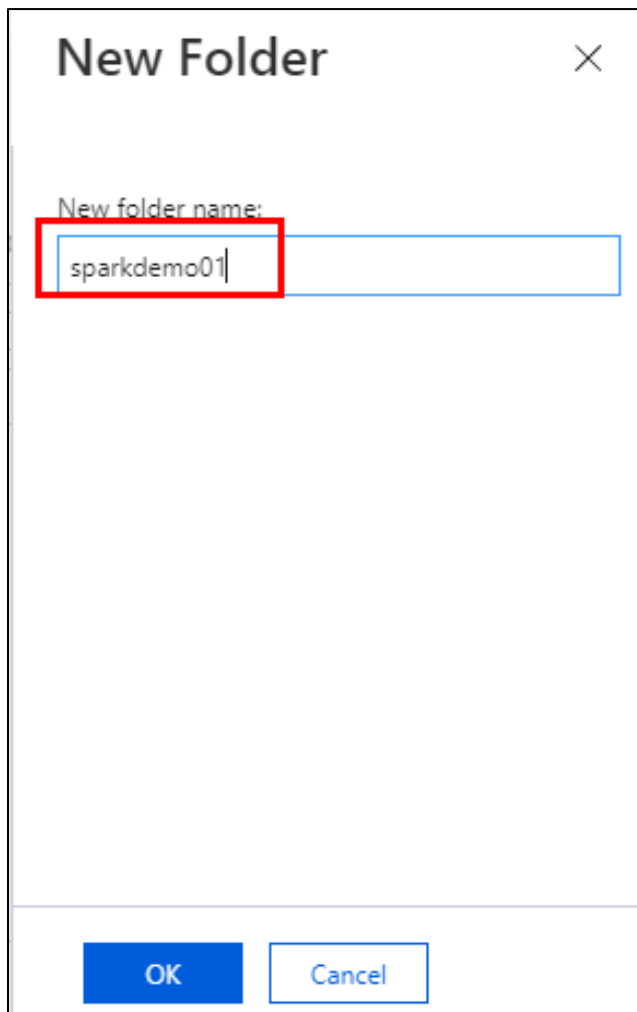
5. 총 5 개의 컨테이너가 구성되었는지 확인합니다.



6. sparkdemo 컨테이너를 클릭합니다. +새 폴더를 클릭합니다.

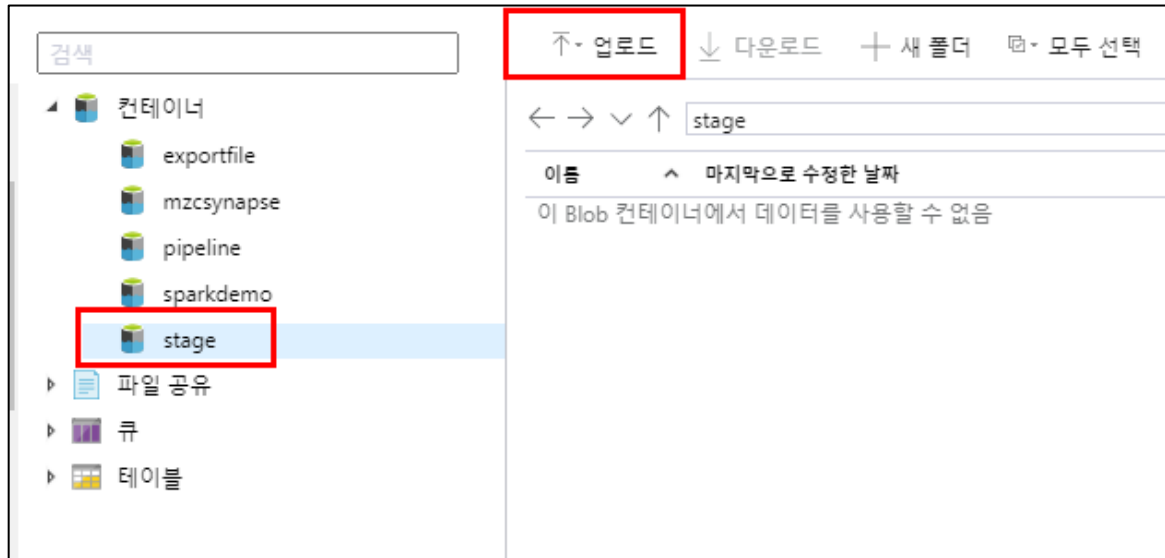


7. sparkdemo01 이라는 폴더를 추가합니다.



New folder name : sparkdemo01

8. 구성된 컨테이너 중에 하나를 선택하여 **업로드**를 클릭합니다.

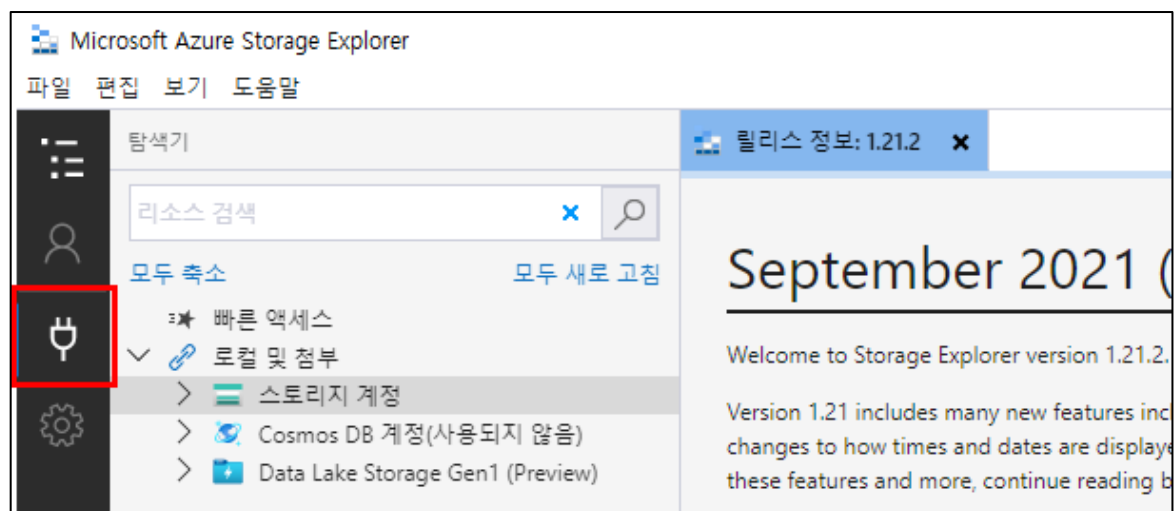


9. 파일을 업로드하기 위해 **Azure Storage Explorer**를 다운로드하여 설치합니다.

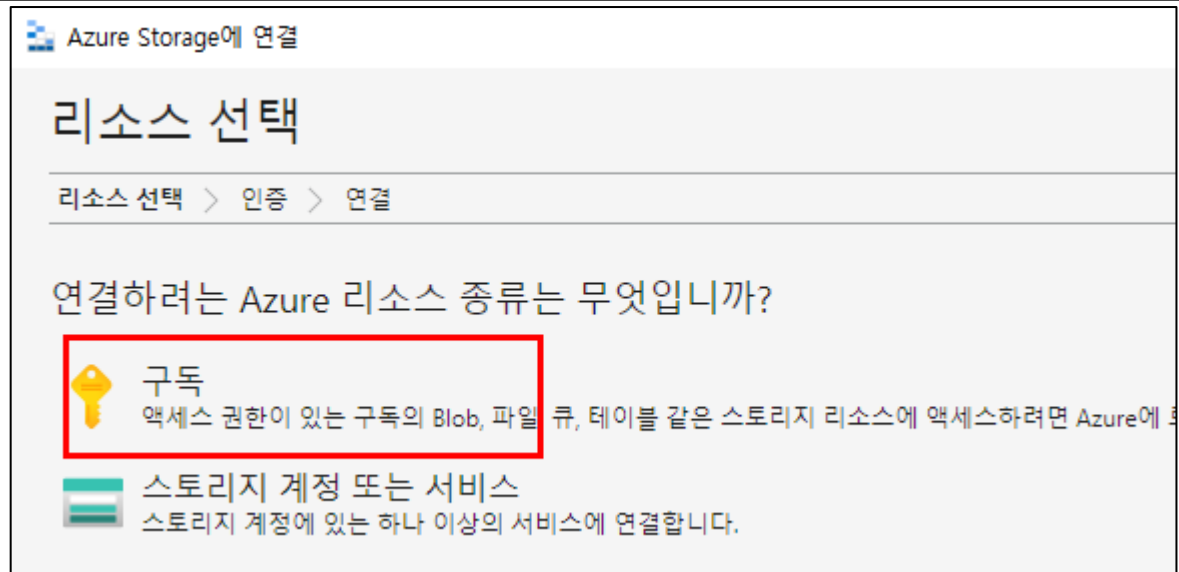




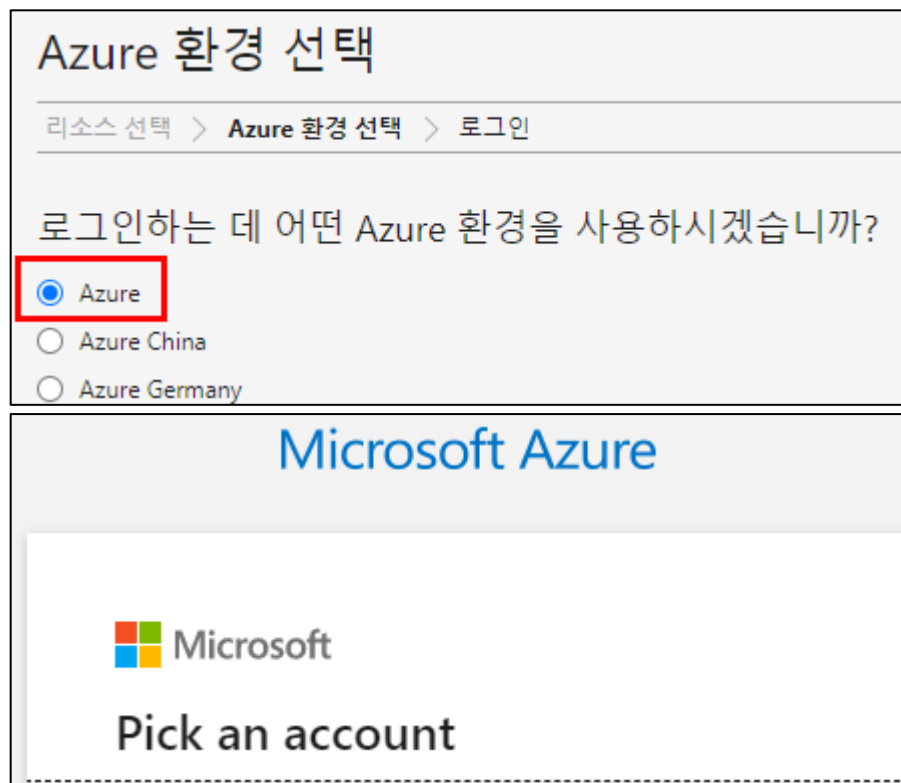
10. 계정 연결을 위해 콘센트 모양의 버튼을 클릭합니다.



11. 리소스 선택에서 구독을 클릭합니다.



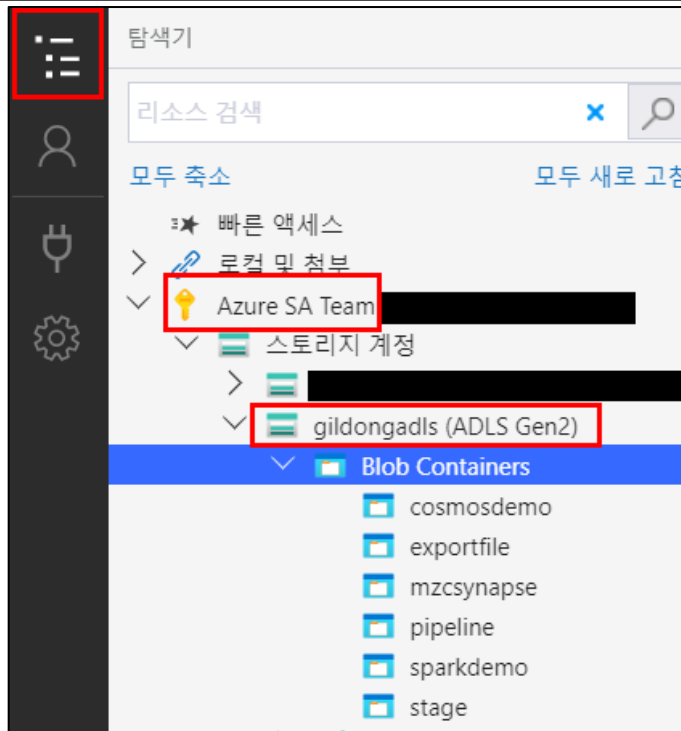
12. **Azure** 환경 선택에서 **Azure** 선택 후 다음을 클릭하여 계정을 연결합니다.



* 연결이 완료되면 아래와 같은 문구가 나타납니다.

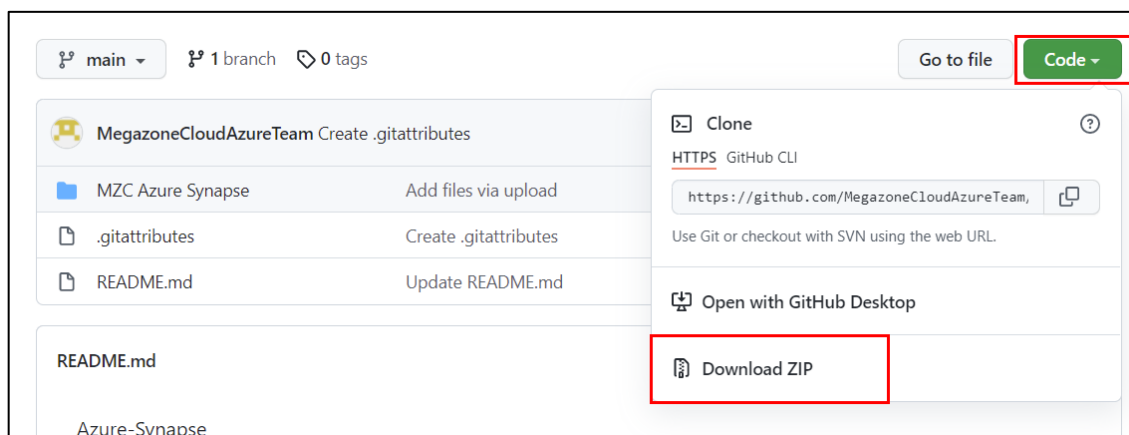
Authenticated. You can return to Storage Explorer. You might need to authenticate again if you close this browser tab.

13. 맨 위의 왼쪽 상단을 클릭하여 개별 구독 아래에 스토리지 계정에서 앞에서 생성했던 **ADLS Gen2** 가 보이는지 확인합니다.

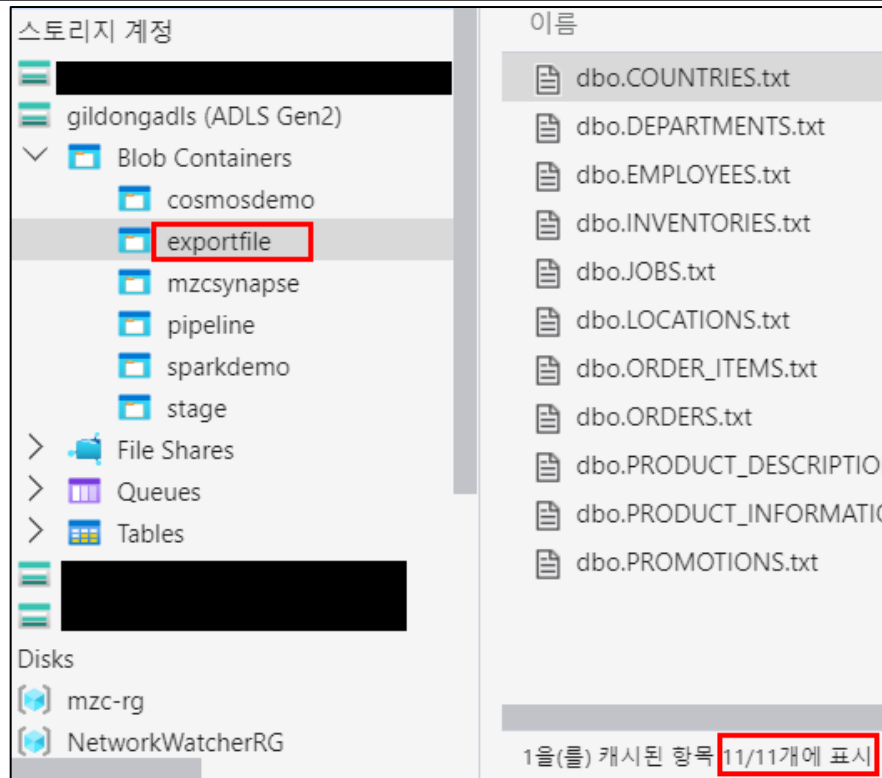


14. **Github** 에서 실습에서 사용될 자료를 다운받습니다. 이미 자료를 받았다면 압축을 풀어서 **MZC Azure Synapse** 폴더를 확인합니다.

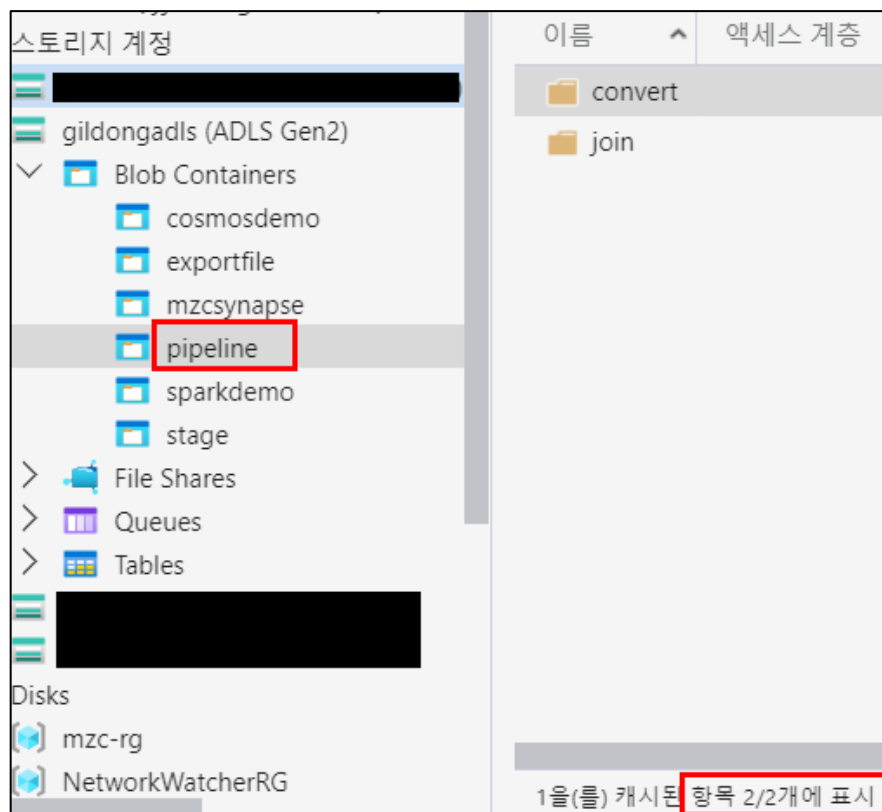
(<https://github.com/MegazoneCloudAzureTeam/Azure-Synapse>)



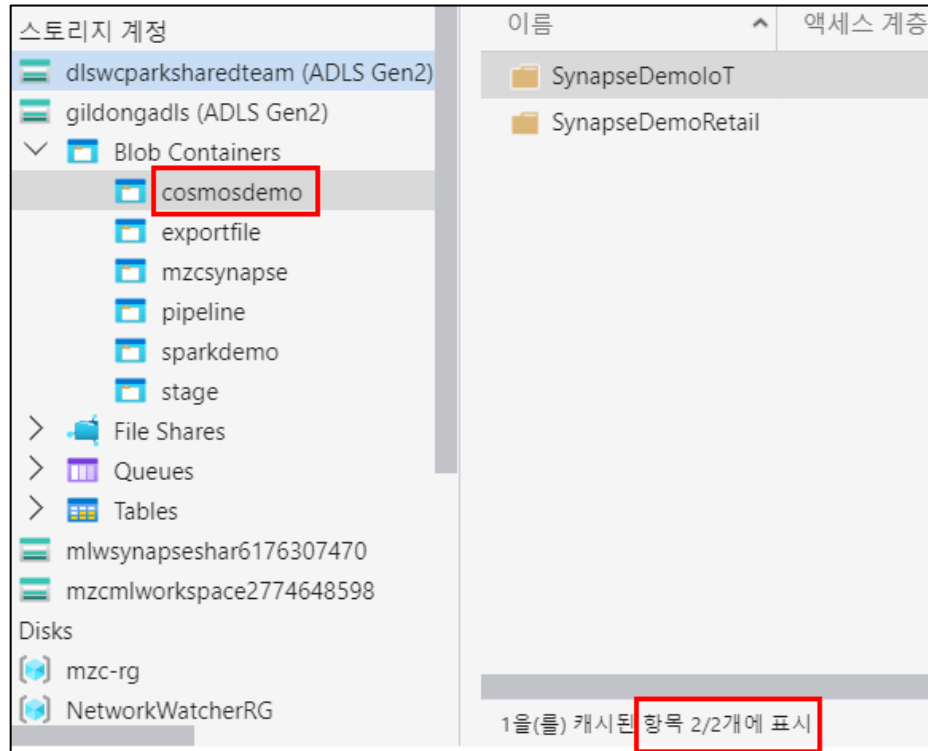
15. 다운로드한 자료에서 **MZC Azure Synapse** LAB FileWexportfile 안의 파일들을 **exportfile** 컨테이너에 **Drag & Drop** 하여 넣어줍니다. 11 개의 텍스트 파일이 업로드 됩니다.



16. MZC Azure SynapseLAB FileWpipeline 안의 폴더들을 pipeline 컨테이너에 Drag & Drop 하여 넣어줍니다. 2 개의 폴더가 업로드 됩니다.

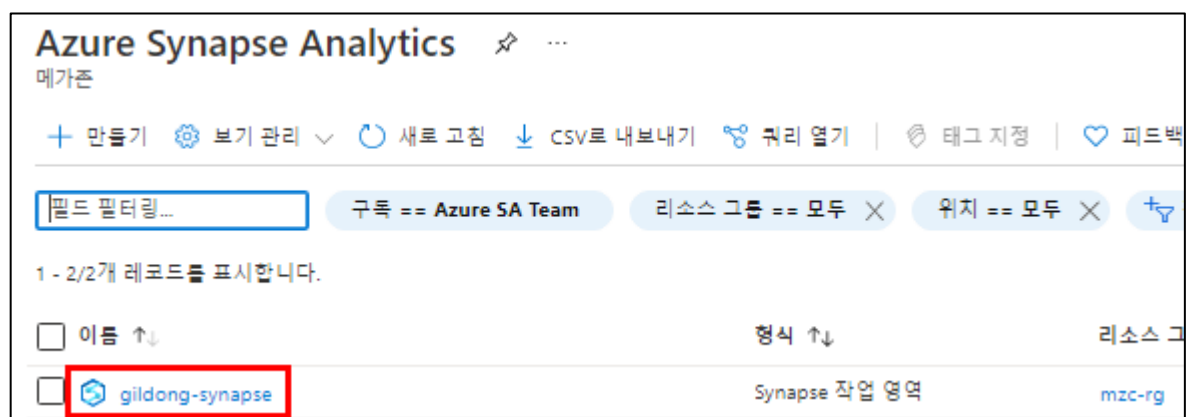


17. MZC Azure SynapseWLAB FileWcosmosdemo 안의 폴더들을 cosmos 컨테이너에 Drag & Drop 하여 넣어줍니다. 2 개의 폴더가 업로드 됩니다.

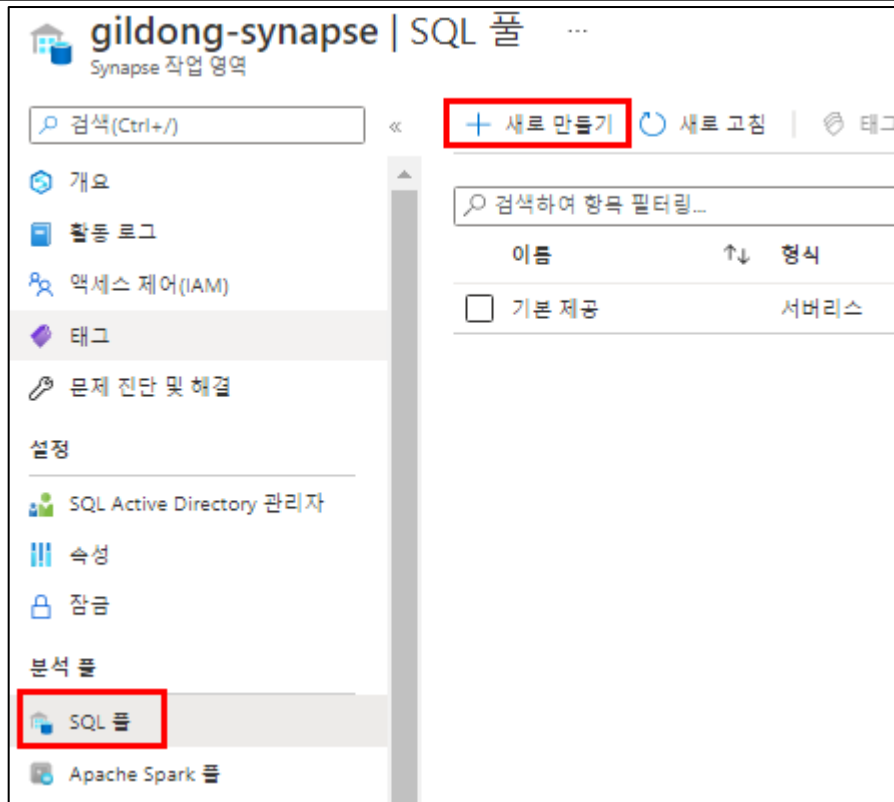


Task 6 : Azure Synapse Analytics 에서 SQL Pool 및 Spark 생성

1. 포털 상단 돋보기를 통해 Azure Synapse Analytics 를 검색하여 생성된 리소스를 클릭해서 이동합니다.



2. 왼쪽 블레이드에서 SQL 풀로 이동하여 새로 만들기를 클릭합니다.



3. 기본 사항에서 SQL 풀 세부 정보를 입력합니다.

*기본 사항 *추가 설정 태그 검토 + 만들기

원하는 구성으로 전용 SQL 풀을 만듭니다. [기본] 탭을 완료한 후 [검토 + 만들기]로 이동하여 적절한 기본값으로 프리비저닝하거나, 각 탭을 방문하여 사용자 지정합니다. [자세한 정보](#)

전용 SQL 풀 세부 정보

전용 SQL 풀의 이름을 지정하고 초기 설정을 선택합니다.

전용 SQL 풀 이름 *

성능 수준 ①

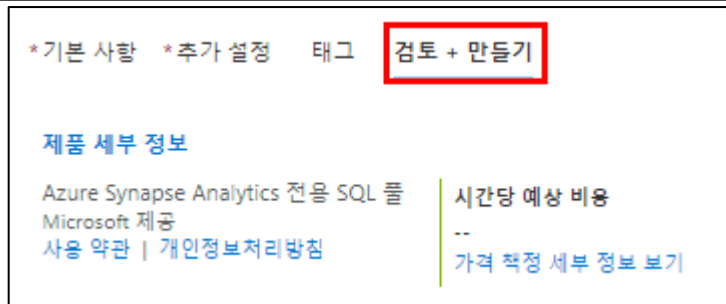
예상 가격 ①

시간당 예상 비용
--
[가격 책정 세부 정보 보기](#)

전용 SQL 풀 이름 : sqlpool

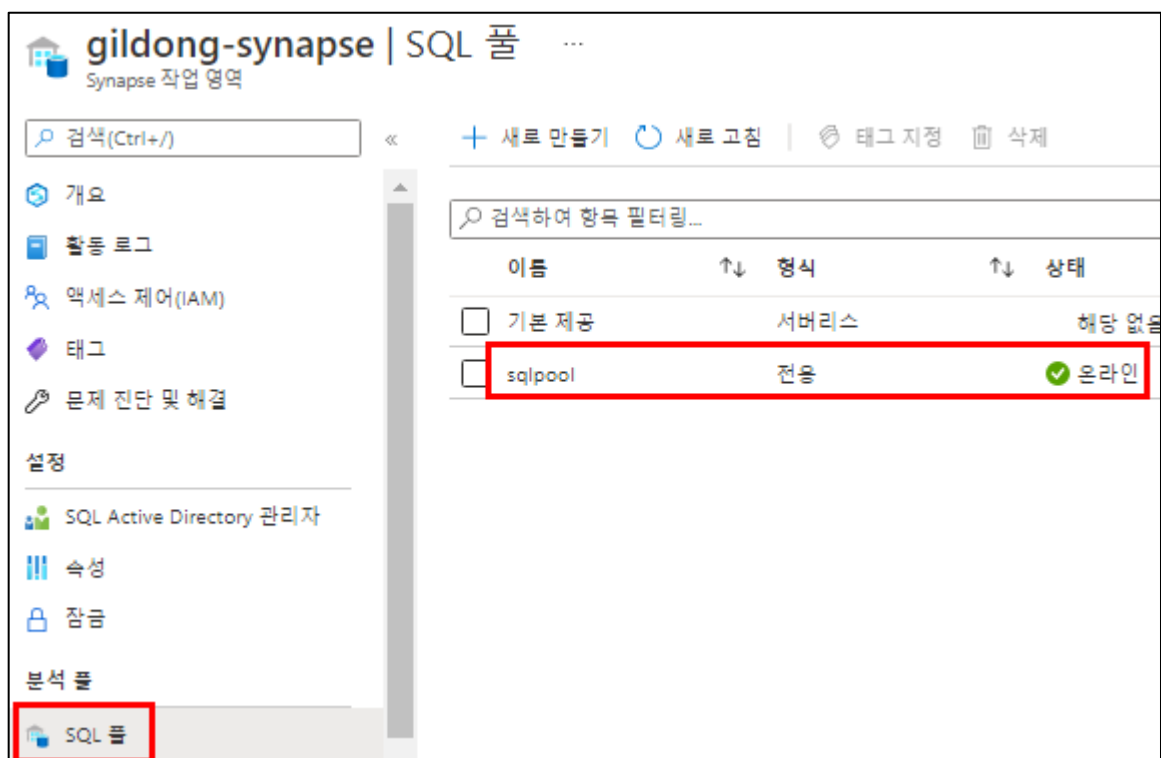
성능 수준 : DW100c

4. 검토 + 만들기를 클릭 후 하단의 만들기로 생성합니다.

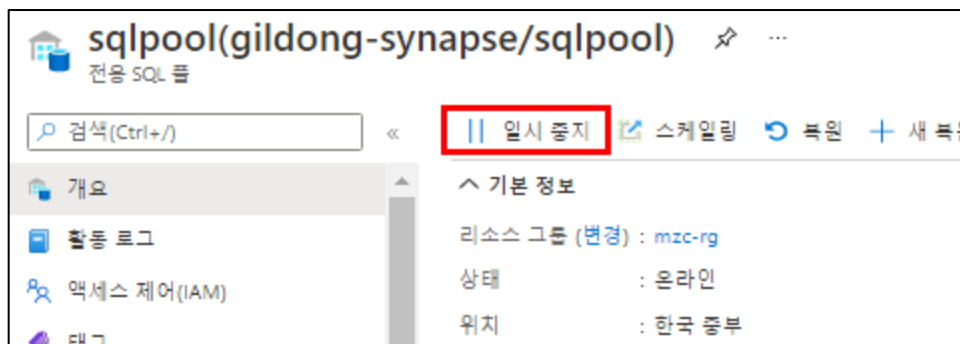


* 그 외에 리소스 설정은 기본 값이므로 수정하지 마세요.

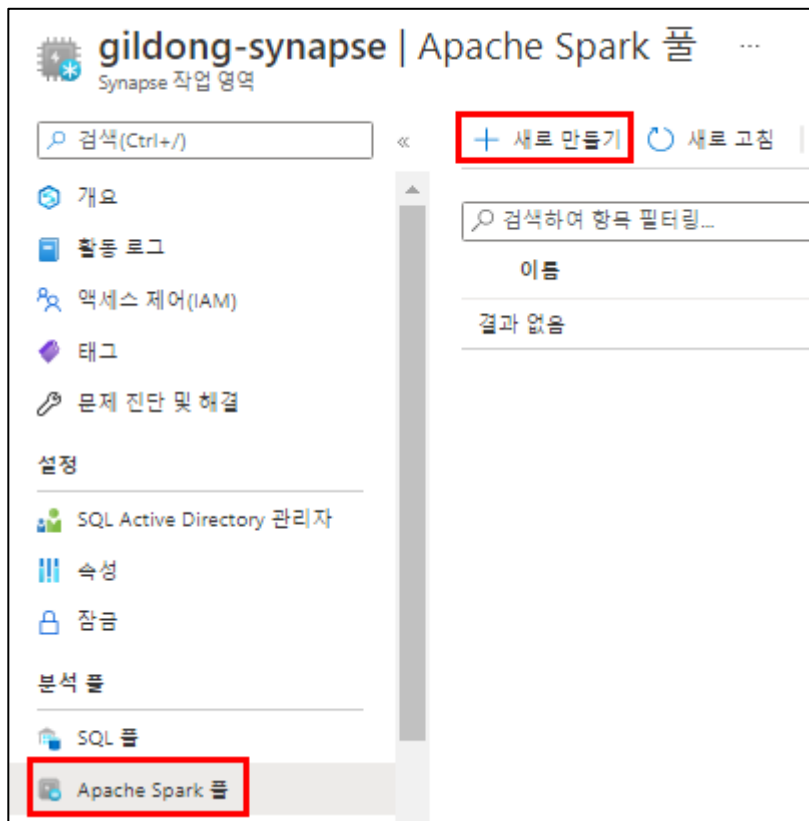
5. 리소스 생성이 완료되면 다시 **Azure Synapse Analytics** 리소스로 돌아갑니다.
SQL 풀을 선택하여 **sqlpool** 이 보이는지 확인합니다.



* sqlpool 을 클릭하면 전용 SQL 풀에 대한 일시 중지가 가능합니다. 사용하지 않을 시 중지하여 비용 발생을 줄일 수 있습니다.



6. 왼쪽 블레이드에서 **Apache Spark 풀**로 이동하여 **새로 만들기**를 클릭합니다.



7. 기본 사항에서 **Apache Spark 풀** 세부 정보를 입력합니다.

*기본 사항 *추가 설정 태그 검토 + 만들기

원하는 구성으로 Synapse Analytics Apache Spark 풀을 만듭니다. [기본] 탭을 완료한 후 [검토] + [만들기]로 가서 적절한 기본 값으로 프리비저닝하거나, 각 탭을 방문하여 사용자 지정합니다.

Apache Spark 풀 세부 정보

Apache Spark 풀의 이름을 지정하고 초기 설정을 선택합니다.

Apache Spark 풀 이름 *

노드 크기 패밀리 MemoryOptimized

노드 크기 *

자동 크기 조정 * ☒ 사용 ☐ 사용 안 함

노드 수 *

예상 가격

예상 hour당 비용
현재 청구 데이터를 이용할 수 없습니다.
[가격 책정 세부 정보 보기](#)

Apache Spark 풀 이름 : sparkpool

노드 크기 : Medium

자동 크기 조정 : 사용

노드 수 : 3 ~ 10

8. 검토 + 만들기를 클릭 후 하단의 만들기로 생성합니다.

* 그 외에 리소스 설정은 기본 값이므로 수정하지 마세요.

9. 리소스 생성이 완료되면 다시 **Azure Synapse Analytics** 리소스로 돌아갑니다.
SQL 풀을 선택하여 **sparkpool** 이 보이는지 확인합니다.

gildong-synapse | Apache Spark 풀 ...

Synapse 작업 영역

검색(Ctrl+/)

개요

활동 로그

액세스 제어 (IAM)

태그

문제 진단 및 해결

설정

SQL Active Directory 관리자

속성

잠금

분석 풀

SQL 풀

Apache Spark 풀

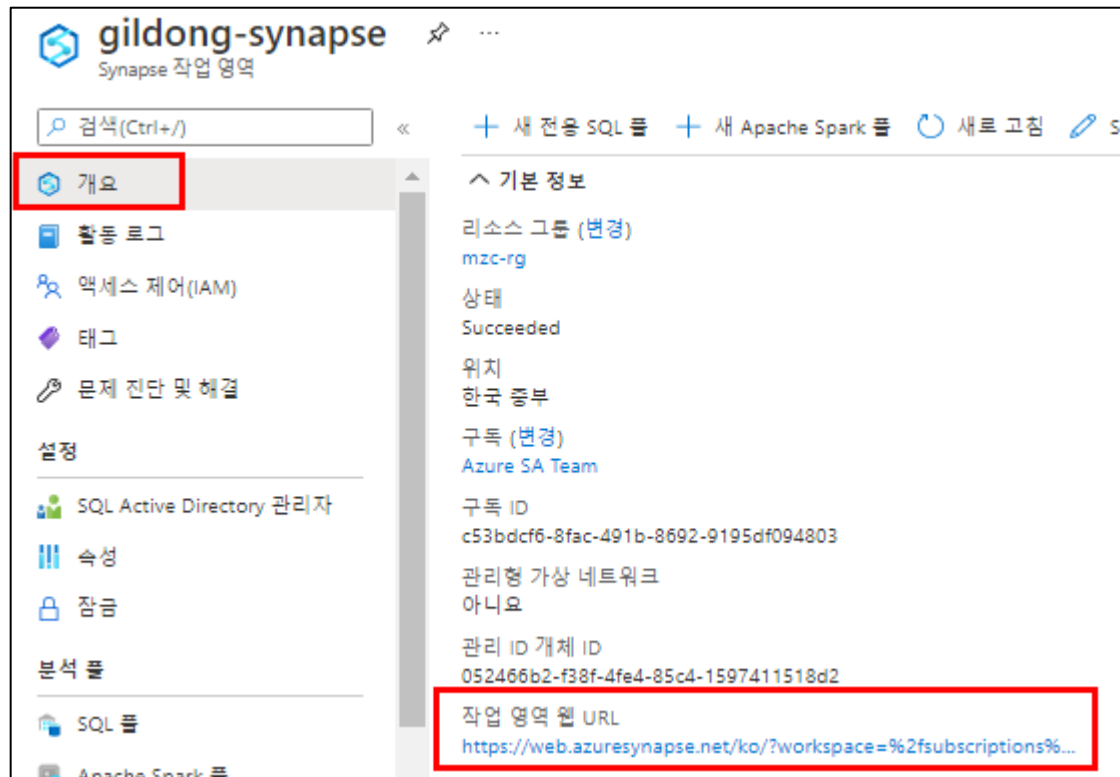
새로 만들기 새로 고침 태그 지정 삭제

검색하여 항목 필터링...

이름	노드 제품군	메모리 최적화
<input type="checkbox"/> sparkpool		

Task 7 : Azure Synapse Workspace Studio 접속

1. Azure Synapse Analytics 리소스의 개요에서 작업 영역 웹 URL 을 클릭합니다.



2. Azure Synapse Studio 웹 페이지가 나타납니다.

