

클라우드 6주차 과제: 윤요섭(Y2023011)

Azure Synapse Analytics 작업 영역 만들기

Microsoft Azure

리소스, 서비스 및 문서 검색(G+J)

모든 서비스 > Azure Synapse Analytics >

Synapse 작업 영역 만들기 ...

*기본 사항 *보안 네트워킹 태그 검토 + 만들기

Synapse 작업 영역을 만들어 몇 번의 클릭만으로 엔터프라이즈 분석 솔루션을 개발할 수 있습니다.

프로젝트 정보

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

구독 * ①

Azure for Students

리소스 그룹 * ①

(신규) rg-synapse

새로 만들기

관리되는 리소스 그룹 ①

mgd-rg-synapse

작업 영역 정보

작업 영역의 이름을 지정하고, 위치를 선택하고, 로그 및 작업 출력의 기본 위치로 사용할 기본 Data Lake Storage Gen2 파일 시스템을 선택합니다.

작업 영역 이름 *

synapse-wsy2023011

지역 *

Korea Central

Data Lake Storage Gen2 선택 * ①

☒ 구독에서 ☐ URL을 통해 수동으로

계정 이름 * ①

(새로 만들기) datalakekey2023011gen2

새로 만들기

파일 시스템 이름 *

(새로 만들기) fsy2023011

새로 만들기

☒ 작업 영역에서 대화형으로 쿼리할 수 있도록 Data Lake Storage Gen2 계정의 Storage Blob 데이터 기여자 역할을 나에게 할당합니다.

Storage Blob 데이터 Contributor 역할을 사용하여 지정된 Data Lake Storage Gen2 계정에 작업 영역 ID 데이터 액세스 권한을 자동으로 부여합니다. 작업

모든 서비스 > Azure Synapse Analytics >

Synapse 작업 영역 만들기 ...

리소스 그룹 * ①

(신규) rg-synapse

새로 만들기

관리되는 리소스 그룹 ①

mgd-rg-synapse

작업 영역 정보

작업 영역의 이름을 지정하고, 위치를 선택하고, 로그 및 작업 출력의 기본 위치로 사용할 기본 Data Lake Storage Gen2 파일 시스템을 선택합니다.

작업 영역 이름 *

synapse-wsy2023011

지역 *

Korea Central

Data Lake Storage Gen2 선택 * ①

☒ 구독에서 ☐ URL을 통해 수동으로

계정 이름 * ①

(새로 만들기) datalakekey2023011gen2

새로 만들기

파일 시스템 이름 *

(새로 만들기) fsy2023011

새로 만들기

☒ 작업 영역에서 대화형으로 쿼리할 수 있도록 Data Lake Storage Gen2 계정의 Storage Blob 데이터 기여자 역할을 나에게 할당합니다.

Storage Blob 데이터 Contributor 역할을 사용하여 지정된 Data Lake Storage Gen2 계정에 작업 영역 ID 데이터 액세스 권한을 자동으로 부여합니다. 작업 영역을 만든 후 다른 사용자가 이 스토리지 계정을 사용할 수 있도록 하려면 다음 작업을 수행하세요.

- 다른 사용자를 작업 영역의 기여자 역할에 할당
- Synapse Studio를 사용하여 다른 사용자에게 적절한 [Synapse RBAC 역할](#) 할당
- 자신과 다른 사용자를 스토리지 계정의 Storage Blob 데이터 기여자 역할에 할당

자세한 정보

검토 + 만들기

< 이전

다음: 보안 >

Synapse 작업 영역 만들기 ...

작업 영역의 보안 옵션을 구성합니다.

인증

SQL 풀과 같은 작업 영역 리소스에 액세스하기 위한 인증 방법을 선택합니다. 인증 방법은 나중에 변경할 수 있습니다.
[자세한 정보](#)

인증 방법

☒ Use both local and Microsoft Entra ID (ME-ID) authentication

☐ Use only Microsoft Entra ID (ME-ID) authentication

SQL Server 관리자 로그인 *

sqladminuser

✓

SQL 암호

.....

✓

암호 확인

.....

✓

시스템이 할당한 관리 ID 권한

작업 영역 시스템 ID를 사용하여 작업 영역 네트워크 액세스 권한을 Data Lake Storage Gen 2 계정에 부여하려면 선택합니다. [자세한 정보](#)

- ☐ Data Lake Storage Gen2 계정에 네트워크 액세스를 허용합니다.
- ☒ 선택한 Data Lake Storage Gen2 계정이 네트워크 액세스 규칙을 사용하여 네트워크 액세스를 제한하지 않거나, 기본 탭 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 몇몇 관리형 키를 사용하는 인프라 계층에서 암호화를 사용한 이중 암호화가 제공됩니다. [자세한 정보](#)

고객 관리형 키를 사용한 이중 암호화 ☐ 사용 ☒ 사용 안 함

Synapse 작업 영역 만들기 ...

✓ 유효성 검사 성공

* 기본 사항 * 보안 네트워킹 태그 검토 + 만들기

제품 세부 정보

Azure Synapse Analytics 작업 영역 서버리스 SQL 예상 비용/TB
Microsoft 제공
사용 약관 | 개인정보처리방침

5.00 USD

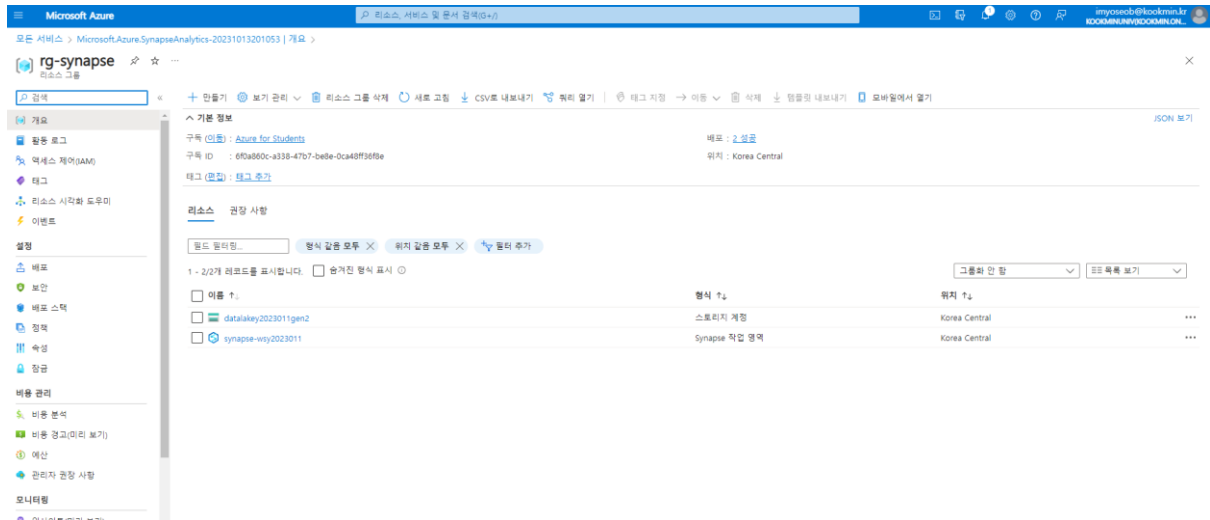
약관

만들기를 클릭함으로써 본인은 (a) 위의 해당 Marketplace 제품과 관련된 약관 및 개인정보처리방침에 동의하고, (b) Microsoft가 현재 결제 방법으로 제품과 관련된 요금을 내 Azure 구독과 동일한 대금 청구 주기로 청구하도록 권한을 부여하는 데 동의합니다. 또한 (c) Microsoft가 지원, 청구 및 기타 거래 목적으로 내 연락처 정보, 트랜잭션 정보 및 사용량 정보를 제품 공급자와 공유할 수 있다는 데 동의합니다. Microsoft는 타사 제품에 대한 권리를 제공하지 않습니다. 자세한 내용은 다음 페이지를 참조하세요. [Azure Marketplace 약관](#)

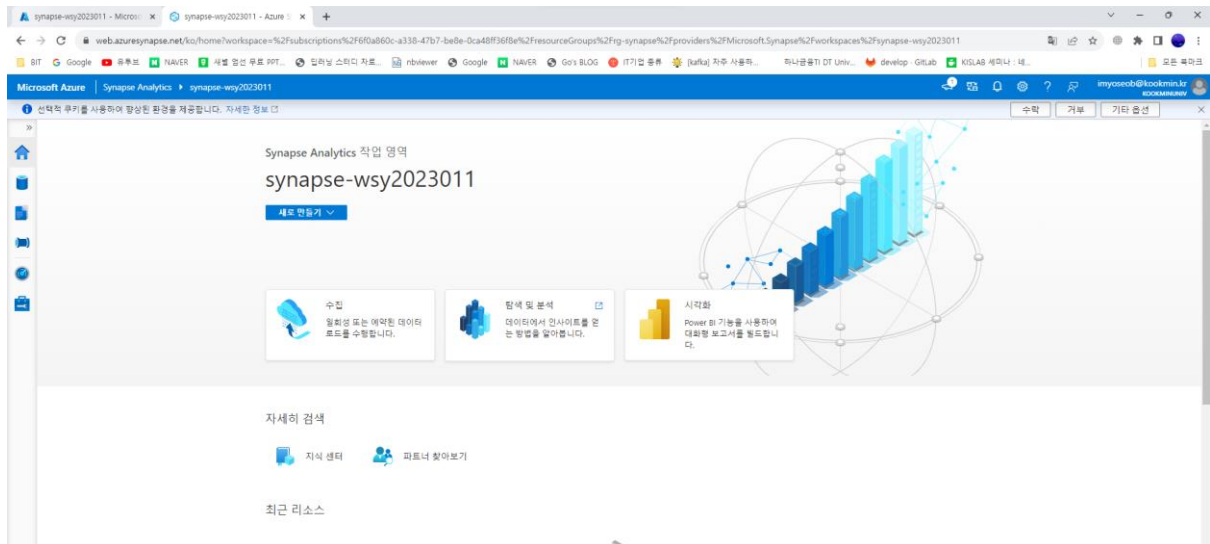
기본 사항

구독	Azure for Students
리소스 그룹	(새로 만드는 중) rg-synapse
지역	Korea Central
작업 영역 이름	(새로 만드는 중) synapse-wsy2023011
Data Lake Storage Gen2 계정	(새로 만드는 중) https://datalakey2023011gen2.dfs.core.windows.net
Data Lake Storage Gen2 파일 시스템	(새로 만드는 중) fsy2023011
관리되는 리소스 그룹	mgd-rg-synapse
역할 할당	Storage Blob 데이터 기여자 역할이 지정된 Data Lake Storage Gen2에서 작업 영역 관리 ID 및 현재 사용자에게 할당됩니다.

보안



Synapse Analytics 작업 영역 생성 완료 증적



데이터 세트 및 파이프라인 만들기

데이터 복사 도구

배출 완료

배출 단계	상태
복사 컨테이너 환경 검증	성공
> 데이터 세트 만들기	성공
> 파이프라인 만들기	성공
> 파이프라인 실행	성공

데이터 세트 및 파이프라인을 만들었습니다. 이제 복사 파이프라인을 모니터링하고 편집하거나 [다침]을 클릭하여 데이터 복사 도구를 닫을 수 있습니다.

선택적 쿼리를 사용하여 향상된 환경을 제공합니다. 자세한 정보

데이터

이름	마지막으로 수정된 날짜	관련도 형식	크기
products.csv	2023. 10. 13. 오후 5:13:52		15.9 KB

CSV 파일에서 TOP 100 쿼리 실행해보기

The screenshot shows the Microsoft Azure Synapse Analytics workspace 'synapse-ws2023011'. The 'Data' tab is active, displaying a list of datasets. The 'fsy2023011' dataset is selected, and the 'SQL script 1' tab is open. The SQL query is as follows:

```
1 -- This is auto-generated code
2 SELECT
3     TOP 100 *
4 FROM
5     OPENROWSET(
6         BULK 'https://datalake2023011gen2.dfs.core.windows.net/fsy2023011/products.csv',
7         FORMAT = 'CSV',
8         PARSE_VERSION = '2.0'
9     ) AS [result]
10
```

The '결과' (Results) tab shows the execution results in a table format:

C1	C2	C3	C4
ProductID	ProductName	Category	ListPrice
771	Mountain-100 Silver, 38	Mountain Bikes	3399.9900
772	Mountain-100 Silver, 42	Mountain Bikes	3399.9900
773	Mountain-100 Silver, 44	Mountain Bikes	3399.9900
774	Mountain-100 Silver, 48	Mountain Bikes	3399.9900
775	Mountain-100 Black, 38	Mountain Bikes	3374.9900
776	Mountain-100 Black, 42	Mountain Bikes	3374.9900
777	Mountain-100 Black, 44	Mountain Bikes	3374.9900

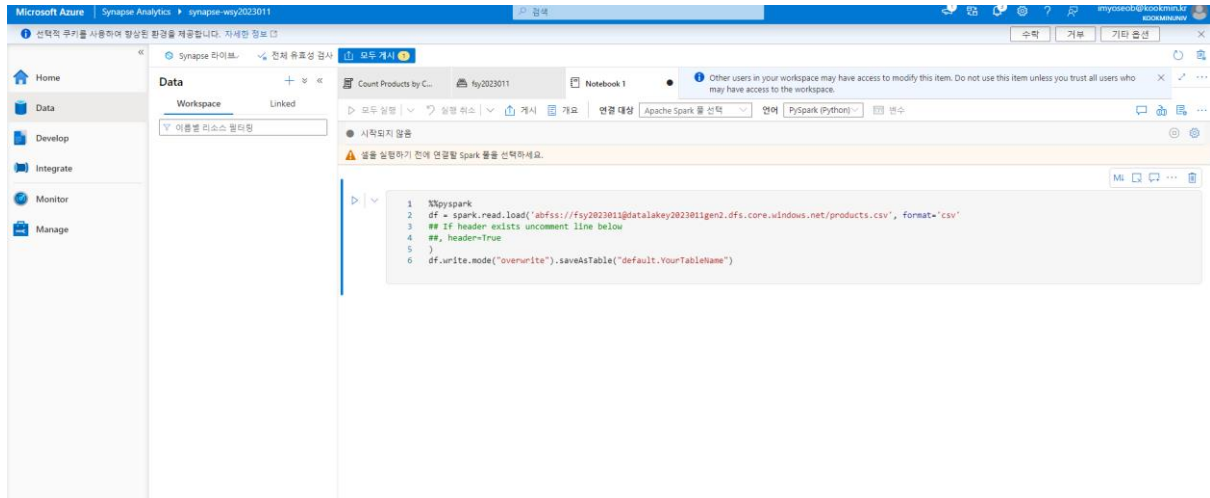
The screenshot shows the same Microsoft Azure Synapse Analytics workspace. The 'SQL script 1' tab is open, displaying a new SQL query:

```
1 SELECT Category
2     , COUNT(*) AS ProductCount
3 FROM OPENROWSET(
4     BULK 'https://datalake2023011gen2.dfs.core.windows.net/fsy2023011/products.csv',
5     FORMAT = 'CSV',
6     PARSE_VERSION = '2.0', HEADER_ROW=TRUE
7 ) AS [result]
8 GROUP BY Category;
9
```

The '결과' (Results) tab shows the execution results in a table format:

Category	ProductCount
Bib-Shorts	3
Bike Racks	1
Bike Stands	1
Bottles and Cages	3
Bottom Brackets	3
Brakes	2
Caps	1
Chainrings	1

Spark Pool 만들기



Spark 쿼리 실행 후, 결과 보기

