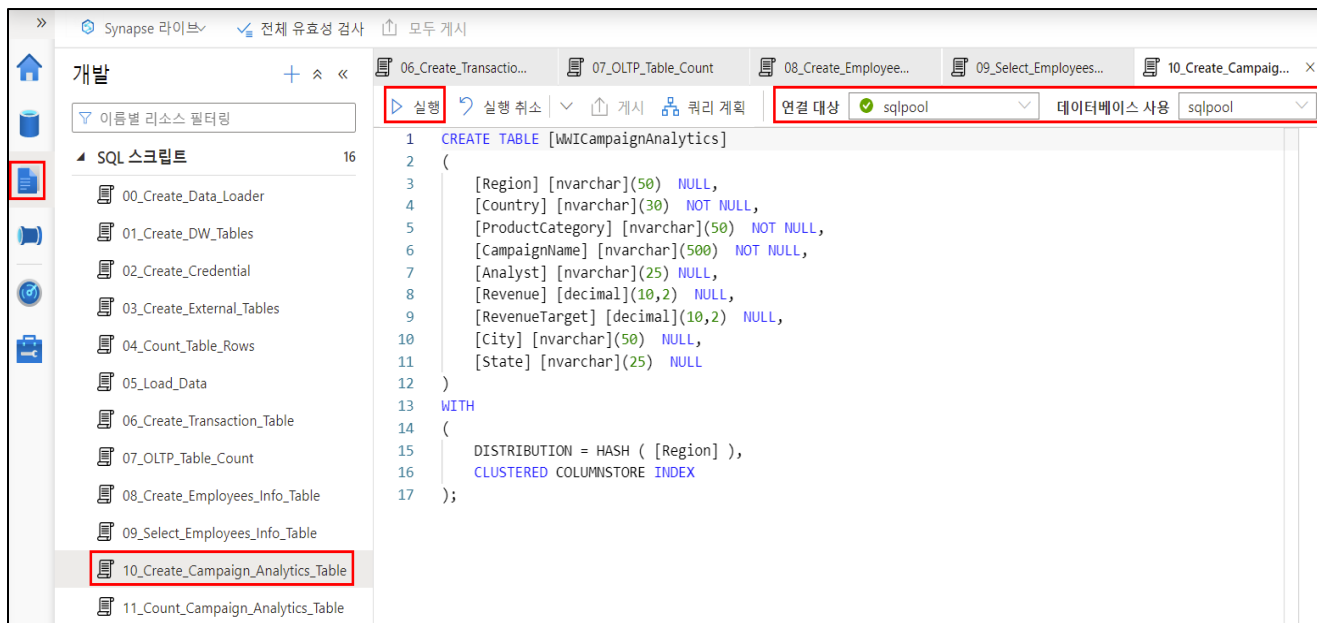


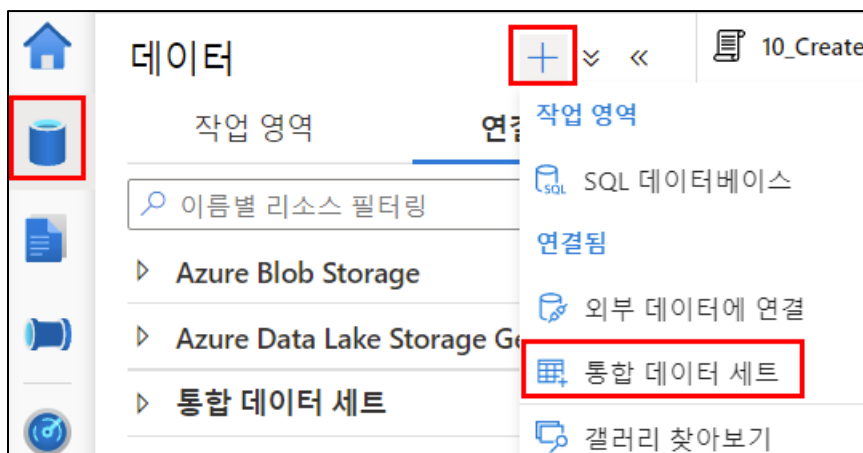
Lab 5 – Transform 변환

Task 1 : Create & Populate Campaign Analytics Table

1. 왼쪽 메뉴의 “개발” 메뉴를 선택하고 **Create_Campaign_Analytics_Table** 스크립트를 열어 실행합니다.



2. “데이터” 메뉴로 이동하여 + 버튼을 누르고 “통합 데이터 세트”를 생성합니다.



3. “새 통합 데이터 세트”에 Azure Data Lake Storage Gen2를 선택, “형식 선택”에 DelimitedText를 선택합니다.

4. 아래와 같이 속성 설정을 입력하고 확인 버튼을 누릅니다.

Field	Value
이름	asamcw_campaign_analytics_csv
연결된 서비스	<ADLS Gen2 Account Name> (각자 가지고 있는 스토리지 계정 선택)
파일 경로 – Container	pipeline
파일 경로 – Directory	convert
파일 경로 – File	campaignAnalytics.csv
첫 번째 행을 머리글로	Unchecked
스키마 가져오기	연결/저장소에서

속성 설정

이름

연결된 서비스 *

통합 런타임을 통해 연결 * ⓘ

파일 경로
 / /

첫 번째 행을 머리글로 ☐

스키마 가져오기
☒ 연결/저장소에서 ☐ 샘플 파일에서 ☐ 없음

5. 다시 + 를 누르고 “통합 데이터 세트”를 생성합니다.

6. “새 통합 데이터 세트”에 Azure Synapse 전용 SQL pool을 선택합니다.

7. 아래와 같이 속성 설정을 입력하고 확인 버튼을 누릅니다.

Field	Value
이름	asamcw_campaign_analytics_asa
SQL 풀	Select sqlpool
테이블 이름	Select dbo.WWICampaignAnalytics
스키마 가져오기	Select 연결/저장소에서

속성 설정

이름

asamcw_campaign_analytics_asa

SQL 풀 * ①

☒ sqlpool

테이블 이름

dbo.WWICampaignAnalytics

↻

☐ 편집

스키마 가져오기

☒ 연결/저장소에서
 ☐ 없음

8. "모두 게시" 버튼을 누르고 저장합니다.

9. "개발" 메뉴로 돌아와서 +를 누르고 "데이터 흐름"을 생성합니다. 속성값에 있는 이름 란에 **Campaign_Analytics_Data**를 입력합니다.

소스 추가

+

○

—

속성

일반

관련

이름 *

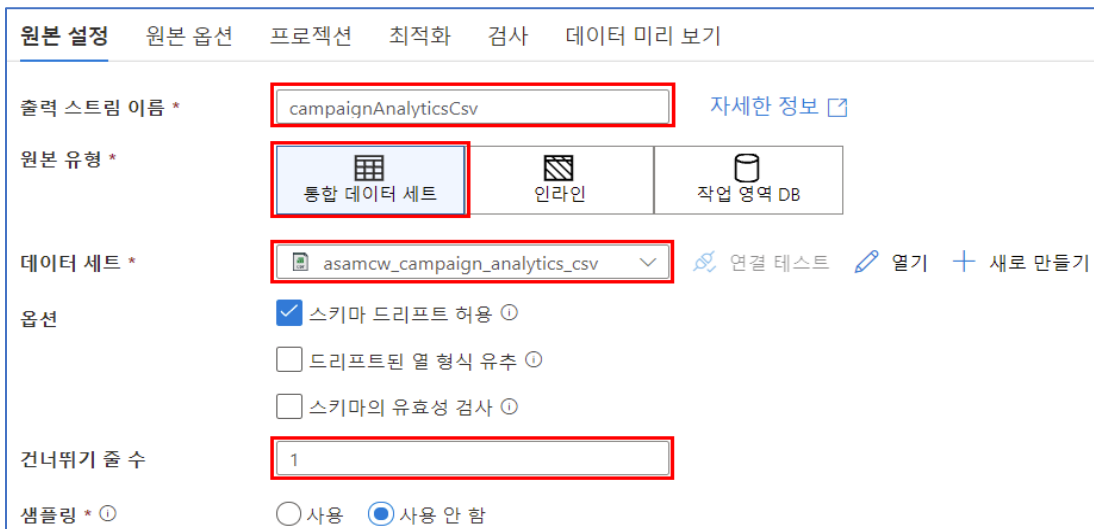
Campaign_Analytics_Data

설명

10. 화면에 보이는 "소스 추가" 를 클릭합니다.

11. 아래와 같이 Property 를 입력합니다.

Field	Value
출력 스트림 이름	Enter campaignAnalyticsCsv
원본 유형	Select 통합 데이터 세트
데이터 세트	Select asamcw_campaign_analytics_csv
건너뛰기 줄 수	Enter 1



원본 설정 원본 옵션 프로젝트 최적화 검사 데이터 미리 보기

출력 스트림 이름 * campaignAnalyticsCsv [자세한 정보](#)

원본 유형 * 통합 데이터 세트 인라인 작업 영역 DB

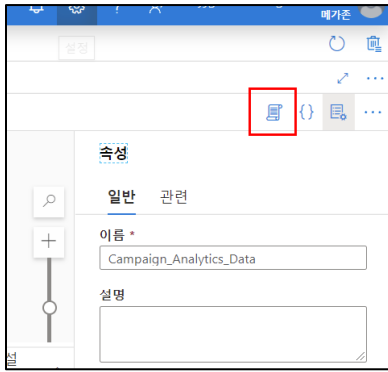
데이터 세트 * asamcw_campaign_analytics_csv [연결 테스트](#) [열기](#) [새로 만들기](#)

옵션 ☒ 스키마 드리프트 허용 [?](#)
☐ 드리프트된 열 형식 유추 [?](#)
☐ 스키마의 유효성 검사 [?](#)

건너뛰기 줄 수 1

샘플링 * [?](#) ☐ 사용 ☒ 사용 안 함

12. Data Flow 를 만들 때 데이터 미리보기 및 스키마 가져오기(프로젝션)와 같은 디버그를 설정하여 특정 기능을 사용할 수 있습니다만 시간적, 환경적 제약으로 인해 이런 과정은 생략하도록 합니다. Data Source 에는 일반적으로 정의해줘야 하는 Schema가 있으며, Data Flow의 Designer Toolbar 도구모음에서도 Script 로 구성이 가능합니다.



13. 스크립트 안의 내용을 아래 스크립트로 수정합니다.

Note: 소스파일의 헤더가 잘못되어 매핑을 변경하는 시나리오입니다.

[변경 전]

```

데이터 흐름 이름 Campaign_Analytics_Data

1  source(output(
2      Column_1 as string,
3      Column_2 as string,
4      Column_3 as string,
5      Column_4 as string,
6      Column_5 as string,
7      Column_6 as string,
8      Column_7 as string,
9      Column_8 as string
10 ),
11     allowSchemaDrift: true,
12     validateSchema: false,
13     ignoreNoFilesFound: false,
14     skipLines: 1) ~> campaignAnalyticsCsv
  
```

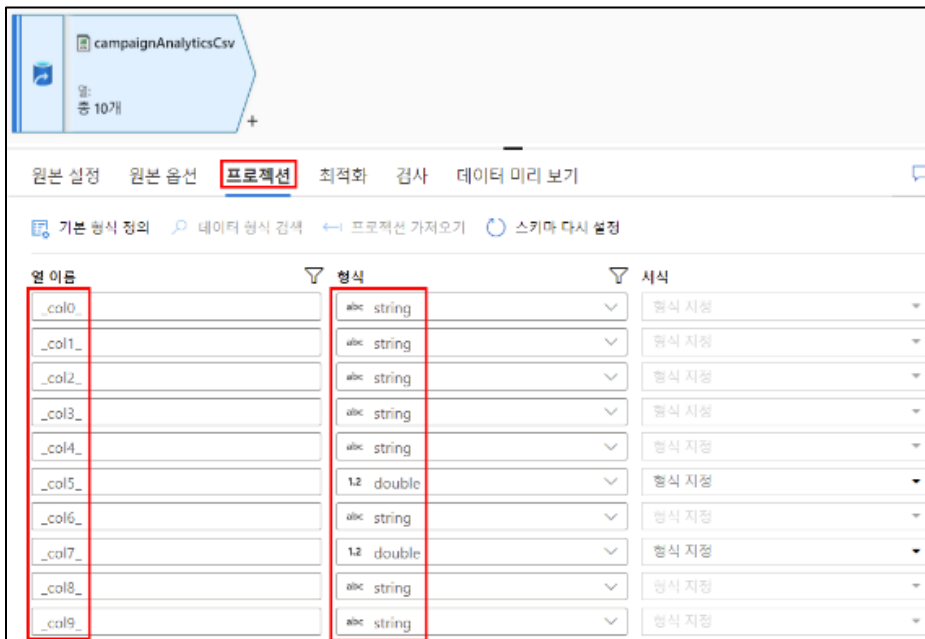
[변경 후]

```

데이터 흐름 이름 Campaign_Analytics_Data

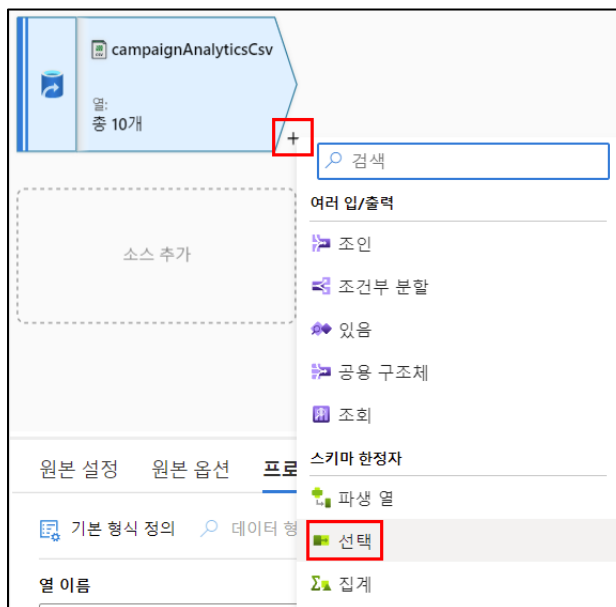
1  source(output(
2      {_col0_} as string,
3      {_col1_} as string,
4      {_col2_} as string,
5      {_col3_} as string,
6      {_col4_} as string,
7      {_col5_} as string,
8      {_col6_} as string,
9      {_col7_} as string,
10     {_col8_} as string,
11     {_col9_} as string
12 ),
13     allowSchemaDrift: true,
14     validateSchema: false,
15     ignoreNoFilesFound: false,
16     skipLines: 1) ~> campaignAnalyticsCsv
  
```

14. "프로젝션" 탭으로 이동하여 아래와 같이 나타나는지 확인합니다.



원 이름	형식	서식
col0	string	형식 지정
col1	string	형식 지정
col2	string	형식 지정
col3	string	형식 지정
col4	string	형식 지정
col5	double	형식 지정
col6	string	형식 지정
col7	double	형식 지정
col8	string	형식 지정
col9	string	형식 지정

15. 이제 아래 +를 클릭하여 선택(Select)을 선택합니다.



16. 출력 스트림 이름에 **mapCampaignAnalytics** 를 입력합니다.

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

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

설정 선택 최적화 검사 데이터 미리 보기

출력 스트림 이름 * 자세한 정보 [?](#)

들어오는 스트림 *

옵션

☒ 중복 입력 열 건너뛰기 ☐

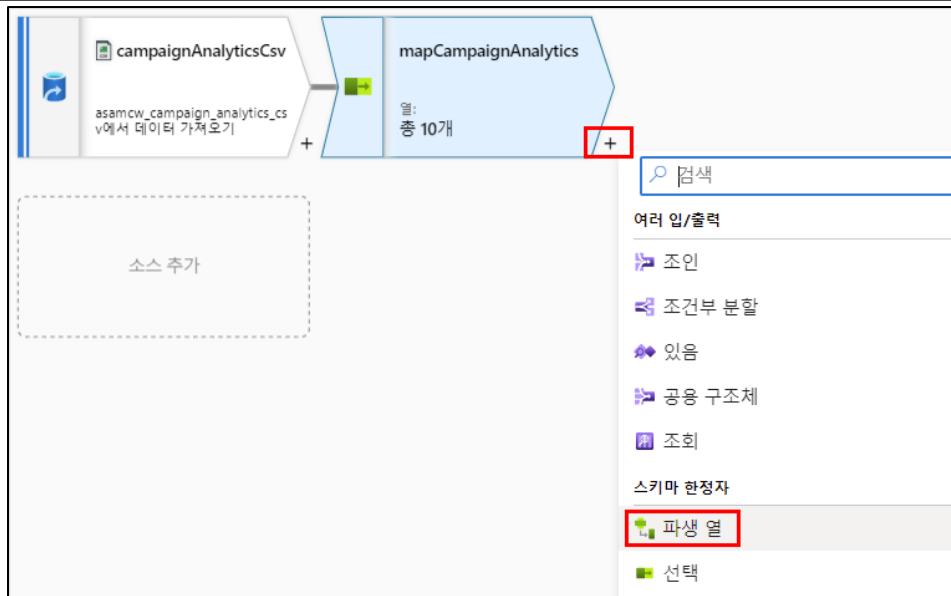
☒ 중복 출력 열 건너뛰기 ☐

입력 열 *

☐ 자동 매핑 ☐ 다시 설정 [+](#) 매핑 추가 [🗑](#) 삭제

<input type="checkbox"/> campaignAnalyticsCsv의 열	다음으로 이름 지정
<input type="checkbox"/> abc_col0_	<input type="text" value="Region"/>
<input type="checkbox"/> abc_col1_	<input type="text" value="Country"/>
<input type="checkbox"/> abc_col2_	<input type="text" value="ProductCategory"/>
<input type="checkbox"/> abc_col3_	<input type="text" value="CampaignName"/>
<input type="checkbox"/> abc_col4_	<input type="text" value="RevenuePart1"/>
<input type="checkbox"/> t2_col5_	<input type="text" value="Revenue"/>
<input type="checkbox"/> abc_col6_	<input type="text" value="RevenueTargetPart1"/>
<input type="checkbox"/> t2_col7_	<input type="text" value="RevenueTarget"/>
<input type="checkbox"/> abc_col8_	<input type="text" value="City"/>
<input type="checkbox"/> abc_col9_	<input type="text" value="State"/>

18. 다시 +를 클릭하여 파생 열(Derived Column)을 선택합니다.

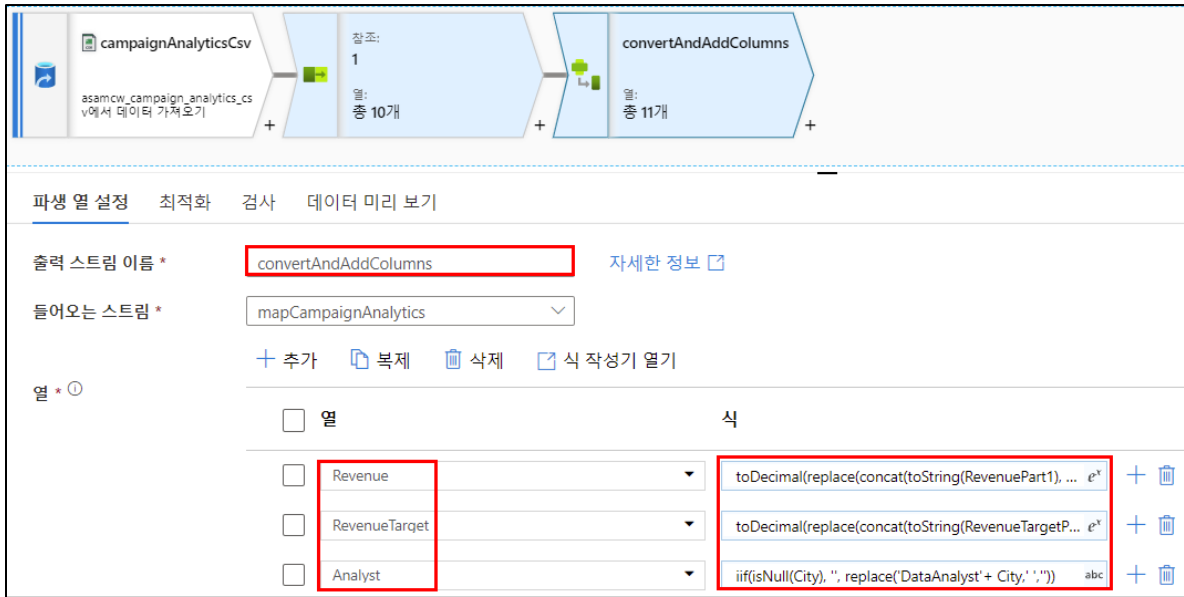


19. 이름을 **convertAndAddColumns**로 입력합니다.

20.아래를 참고하여 입력합니다.

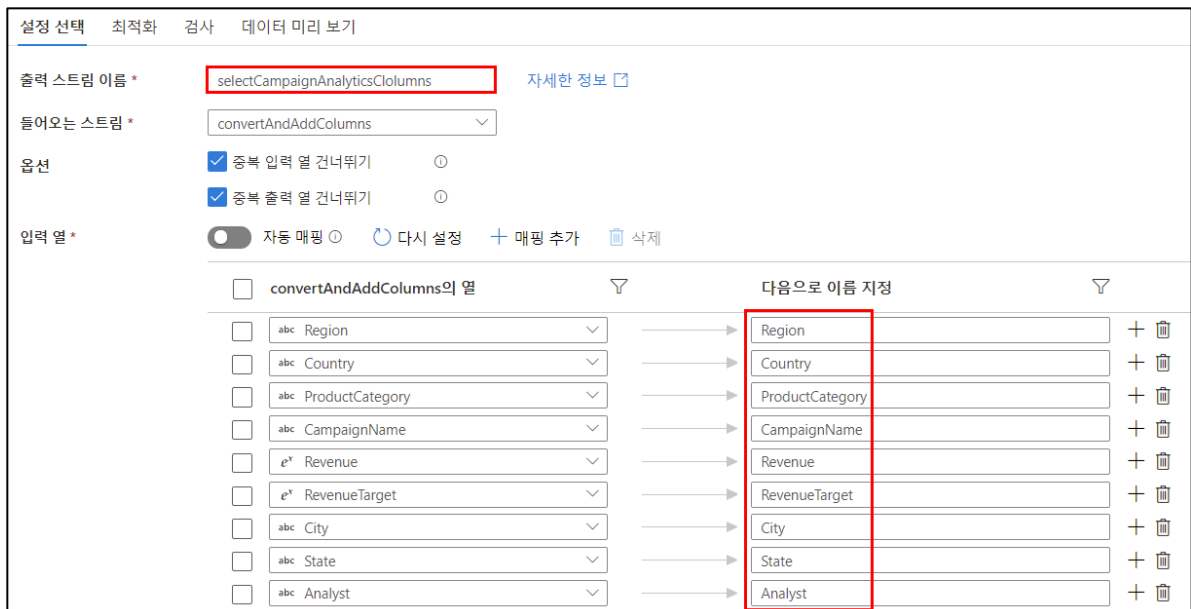
Column	Expression	Description
Revenue	<code>toDecimal(replace(concat(toString(RevenuePart1), toString(Revenue)), '\\', ''), 10, 2, '\$###,###.##')</code>	RevenuePart1과 Revenue필드를 연결하고 \ 문자가 잘못된 경우 데이터를 10진수 형식으로 변환합니다.
RevenueTarget	<code>toDecimal(replace(concat(toString(RevenueTargetPart1), toString(RevenueTarget)), '\\', ''), 10, 2, '\$###,###.##')</code>	RevenueTargetPart1과 RevenueTarget필드를 연결하고 허용되지 않은 \ 문자를 변환한 다음, 10진수 형식으로 변환합니다.
Analyst	<code>iif(isNull(City), '', replace('DataAnalyst'+ City, ', '''))</code>	City필드가 NULL이면 Analyst필드에 빈 문자열을 할당하고, 그렇지 않다면 DataAnalyst를 City값에 연결하여 모든 공백구간을 제거합니다.

(복사해서 사용하실 경우, 1줄로 만든 후 붙여넣기 하시면 됩니다.)

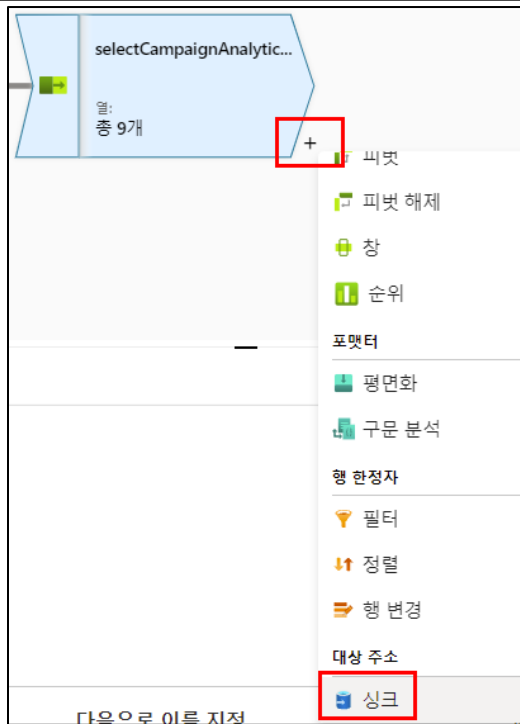


21. 다시 +를 클릭하고 선택(Select)을 선택합니다.

22. 이름에 **selectCampaignAnalyticsColumns**를 입력하고 **RevenuePart1**과 **RevenueTargetPart1**를 지웁니다.



23. 마지막으로 +를 클릭하고 싱크(Sink)를 선택합니다.



24. 이름에 **campaignAnalyticsAsa**를 입력하고 "데이터 세트"를 선택할 때 "+ 새로 만들기" 를 눌러 새로 생성합니다. 새 통합 데이터 세트에는 "Azure Data Lake Storage Gen2"를 선택하고, 형식 선택에는 "Parquet"를 선택합니다.

25. 이름에 **asamcw_campaign_analytics_parquet** 를 입력하고 아래와 같이 속성 설정 값을 넣어줍니다.

속성 설정

이름

asamcw_campaign_analytics_parquet

연결된 서비스 *

youngjoong-synapse-WorkspaceDefaultStorage

통합 런타임을 통해 연결 * ①

AutoResolveIntegrationRuntime

파일 경로

pipeline / convert / 파일

스키마 가져오기

☒ 연결/저장소에서
 ☐ 샘플 파일에서
 ☐ 없음

> 고급

26. 위와 같이 된 것을 확인 후 “설정” 탭으로 가서 **파일 이름 옵션**(File Name Option)에서 **단일 파일로 출력**(Output to one single file)을 선택합니다.

27. 파일 명에는 **campaign_analytics.parquet** 를 입력합니다. **단일 파티션 설정** (Set Single Partition)도 클릭합니다.

싱크

설정

매핑

최적화

검사

데이터 미리 보기

⚠ '단일 파일로 출력'을 사용하려면 '단일 파티션'이 선택된 파티션 형식이어야 합니다. 칠 수 있으며 더 작은 데이터 세트에만 사용해야 합니다.

단일 파티션 설정

폴더 지우기

☐

파일 이름 옵션 *

단일 파일로 출력

단일 파일로 출력 * ①

campaign_analytics.parquet

동적 콘텐츠 추가 [Alt+Shift+D]

Umask ①

소유자

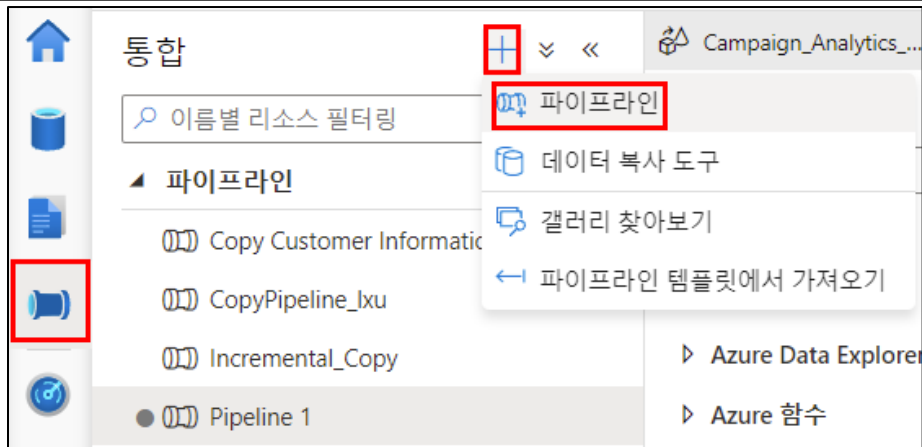
☐ R

☐ W

☐ X

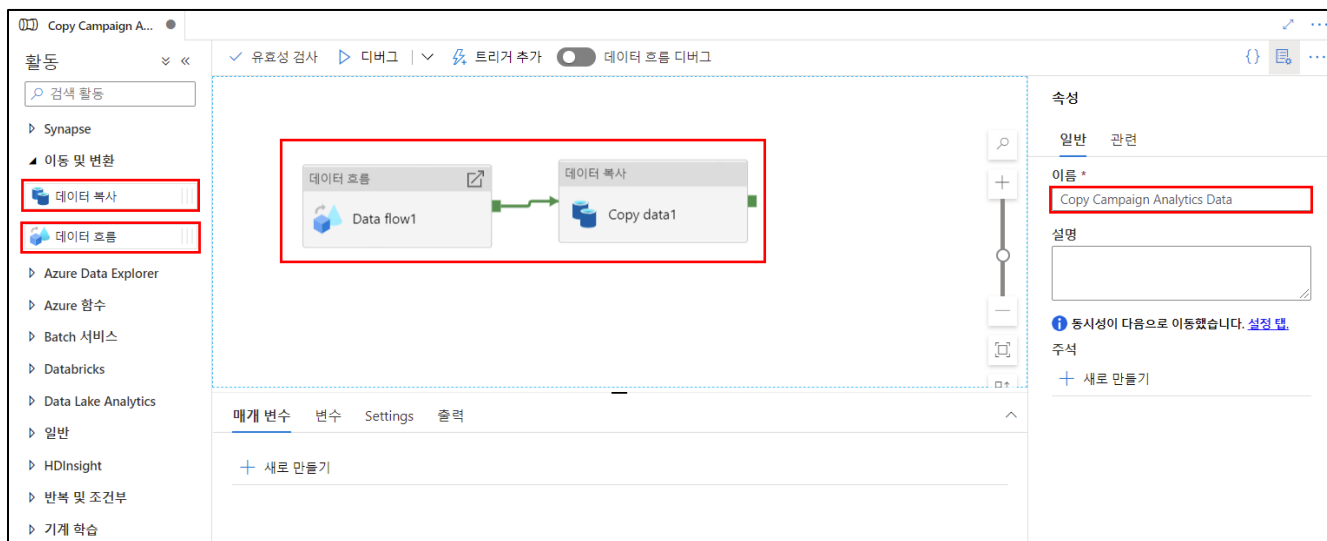
28. Publish All 하여 저장합니다.

29. 통합(Integrate) 탭에서 +를 클릭 후 파이프라인(Pipeline)을 생성합니다.



30. 속성의 이름을 **Copy Campaign Analytics Data** 로 입력합니다.

31. 이동 및 변환의 "데이터 흐름"을 Drag & Drop 하고, "데이터 복사"를 Drag & Drop 합니다. 아래와 같이 따라합니다. 데이터 흐름을 데이터 복사로 이어줍니다.



32. 데이터 흐름을 선택하고 설정 탭에 Campaign_Analytics_data를 선택합니다.

일반

설정

매개 변수¹

사용자 속성

데이터 흐름 *

Campaign_Analytics_Data

열기

+ 새로 만들기

실행 위치(Azure IR) * ①

AutoResolveIntegrationRuntime

컴퓨팅 형식 * ①

기본(범용)

코어 수 * ①

4(+ 4개 드라이버 코어)

로깅 수준 * ①

☒ 자세한 정보 표시
 ☐ 기본
 ☐ 없음

▶ 싱크 속성

▶ 스테이징 ①

33. 데이터 복사를 선택하고, 소스(Source) 탭으로 이동하여, **asamcw_campaignanalytics_parquet**를 원본 소스 데이터 세트(Source Data Set)로 선택합니다. 파일 경로 유형에 **와일드 파일 경로(Wildcard File Path)**를 선택 후 **pipeline / convert / *.parquet** 를 입력합니다.

활동

유효성 검사

유효성 검사 복사 런타임

디버그

트리거 추가

데이터 흐름 디버깅

데이터 흐름

Campaign_Analytics_d ata

데이터 복사

Copy data1

일반

소스

싱크

매핑

설정

사용자 속성

원본 데이터 세트 *

asamcw_campaignanalytics_parquet

열기

+ 새로 만들기

데이터 미리 보기

자세한 정보

파일 경로 유형

☐ 데이터 세트의 파일 경로
 ☒ 와일드카드 파일 경로
 ☐ 파일 목록

와일드카드 경로

pipeline / convert / *.parquet

마지막으로 수정한 날짜 별 필터링

시작 시간(UTC)

종료 시간(UTC)

재귀적으로

☒

파티션 검색 사용

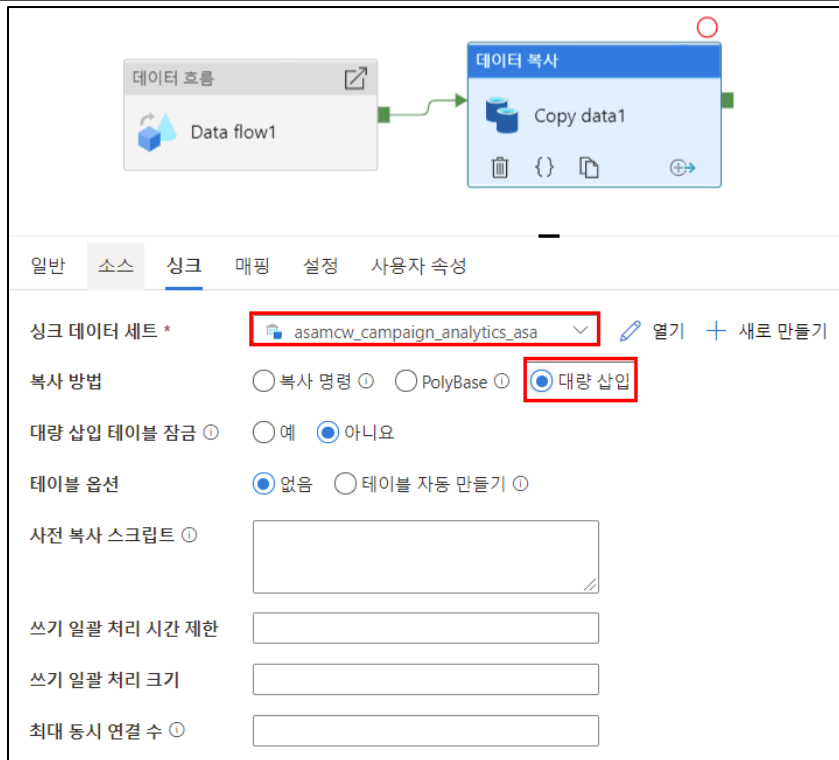
☐

최대 동시 연결 수

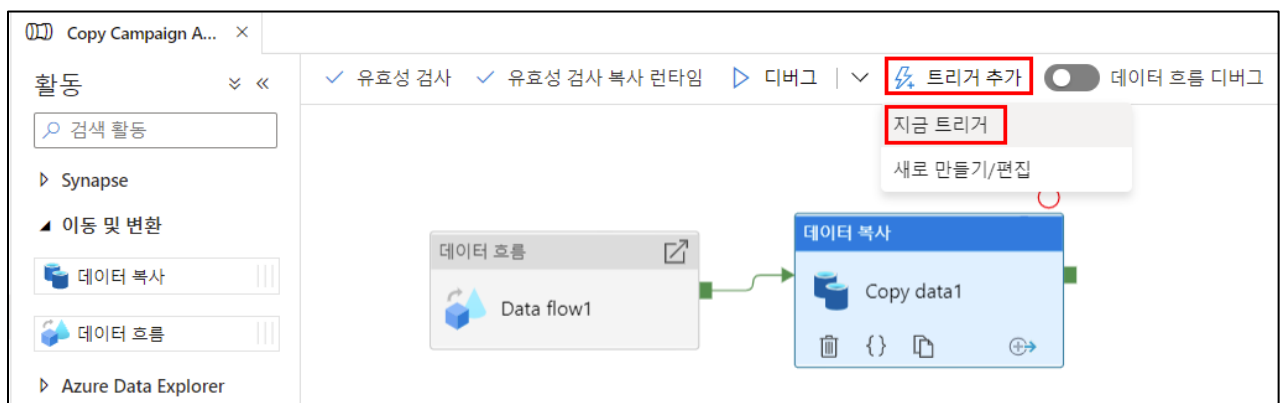
추가 열

+ 새로 만들기

34. 싱크(Sink) 탭으로 이동해서 싱크 데이터 세트 **asamcw_campaign_analytics_asa**를 선택하고 대량 삽입(Bulk Insert)을 선택합니다.



35. 모두 게시(Publish All) 하여 저장합니다. 트리거 추가 버튼의 지금 트리거(Trigger Now)를 선택하여 실행합니다.



36. 모니터(Monitoring) 탭에서 확인합니다.

모든 파이프라인 실행 > Copy Campaign Analytics Data - 활동 실행

Copy Campaign Analytics Data

목록 Gantt

다시 실행 작업에서 다시 실행 실패한 작업에서 다시 실행 새로 고침 파이프라인 업데이트

데이터 흐름 Data flow1 데이터 복사 Copy data1

활동 실행

파이프라인 실행 ID de0817a9-3157-4883-b4e5-85f623143bb5

모든 상태

1~2/2개 항목 표시

작업 이름	작업 유형	실행 시작	기간	상태	오류	로그	통합 런타임
Copy data1	데이터 복사	10/12/21, 오전 10:08:59	00:00:12	성공			DefaultIntegrationRuntime (Korea Central)
Data flow1	데이터 흐름	10/12/21, 오전 10:05:52	00:03:07	성공			DefaultIntegrationRuntime (Korea Central)

37. 작업이 완료되면 Count_Campaign_Analytics_Table 스크립트를 실행하여 결과를 확인합니다.

개발

이름별 리소스 필터링

SQL 스크립트 20

- 00_CreateDataLoader
- 01_CreateDWTables
- 02_CreateCredential
- 03_CreateExternalTables
- 04_LoadData
- 05_CreateTransactionTable
- 06_OLTP_Table_Count
- 10_Create_Campaign_Analytics_Table
- 11_Count_Campaign_Analytics_Table

실행 실행 취소 게시 쿼리 계획 연결 대상 sqlpool 데이터베이스 사용 sqlpool

1 select count(Region) from dbo.WMICampaignAnalytics;

결과 메시지

보기 테이블 차트 결과 내보내기

검색

(열 이름 없음)

554