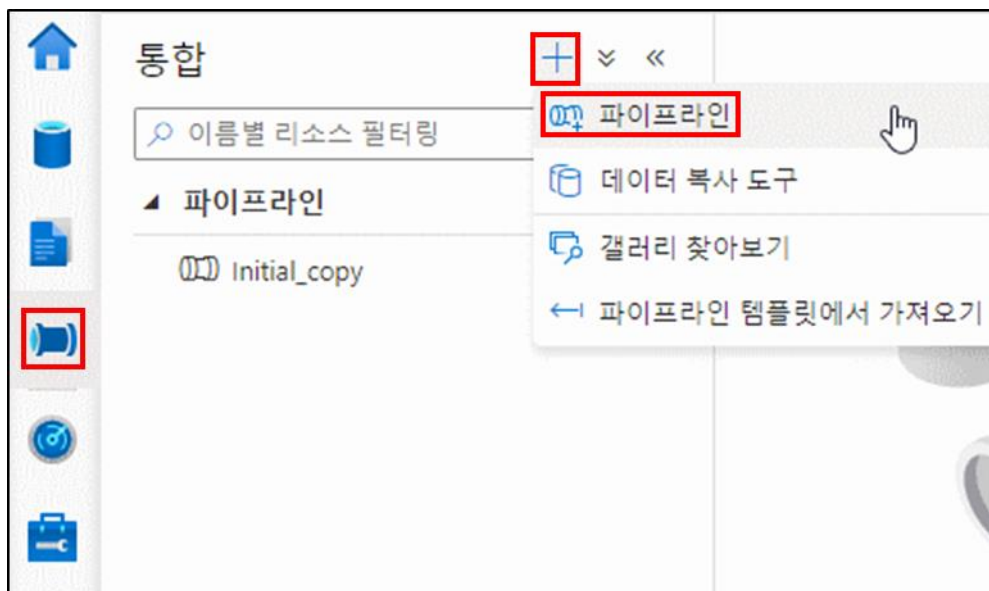


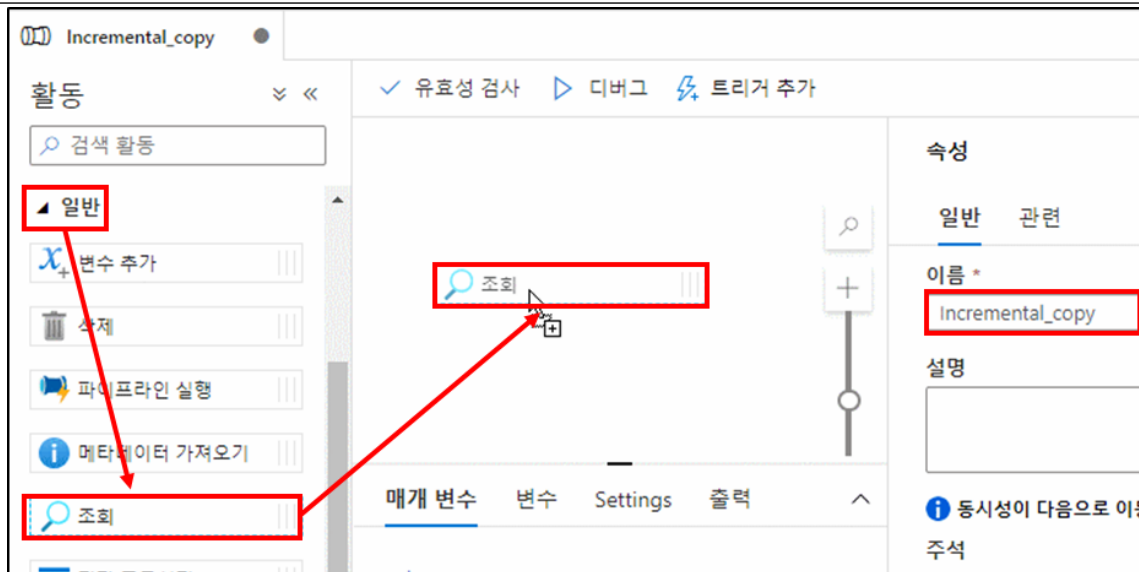
Lab 3 – Pipeline 을 이용한 원격 DB 의 증분 데이터 LOAD

Task 1 : Pipeline 을 통한 Transaction Table(OLTP Data 의 증분 데이터 마이그레이션

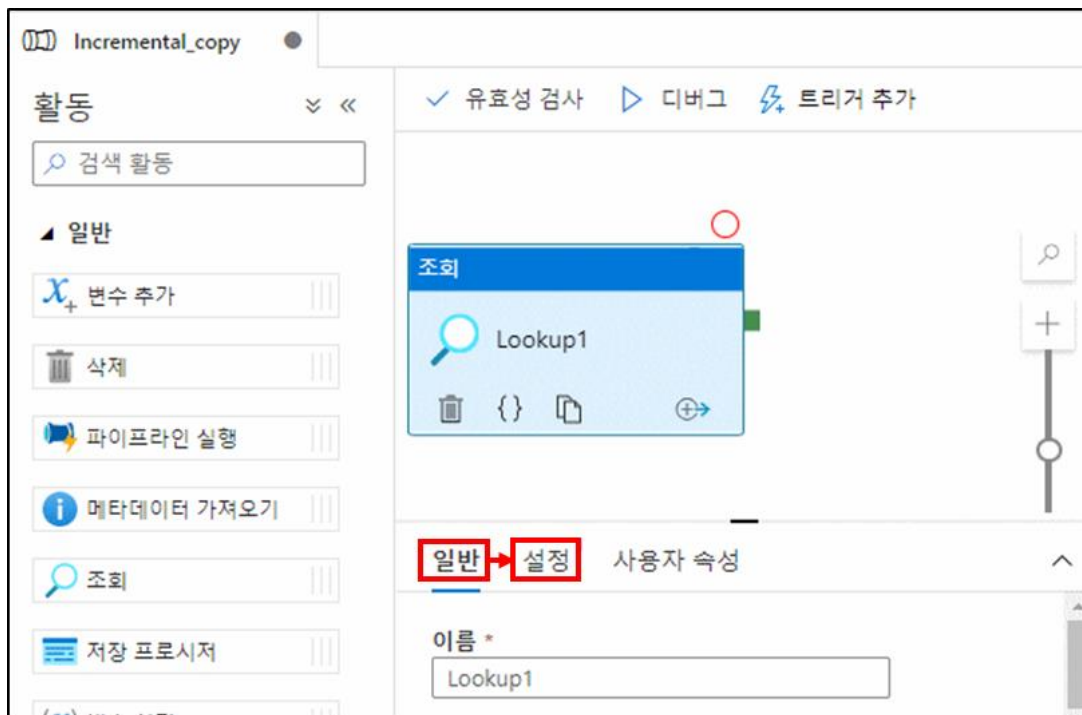
1. 파이프라인을 새로 생성하기 위해 **통합 Hub** 로 이동합니다. + 를 눌러 파이프라인을 선택합니다.



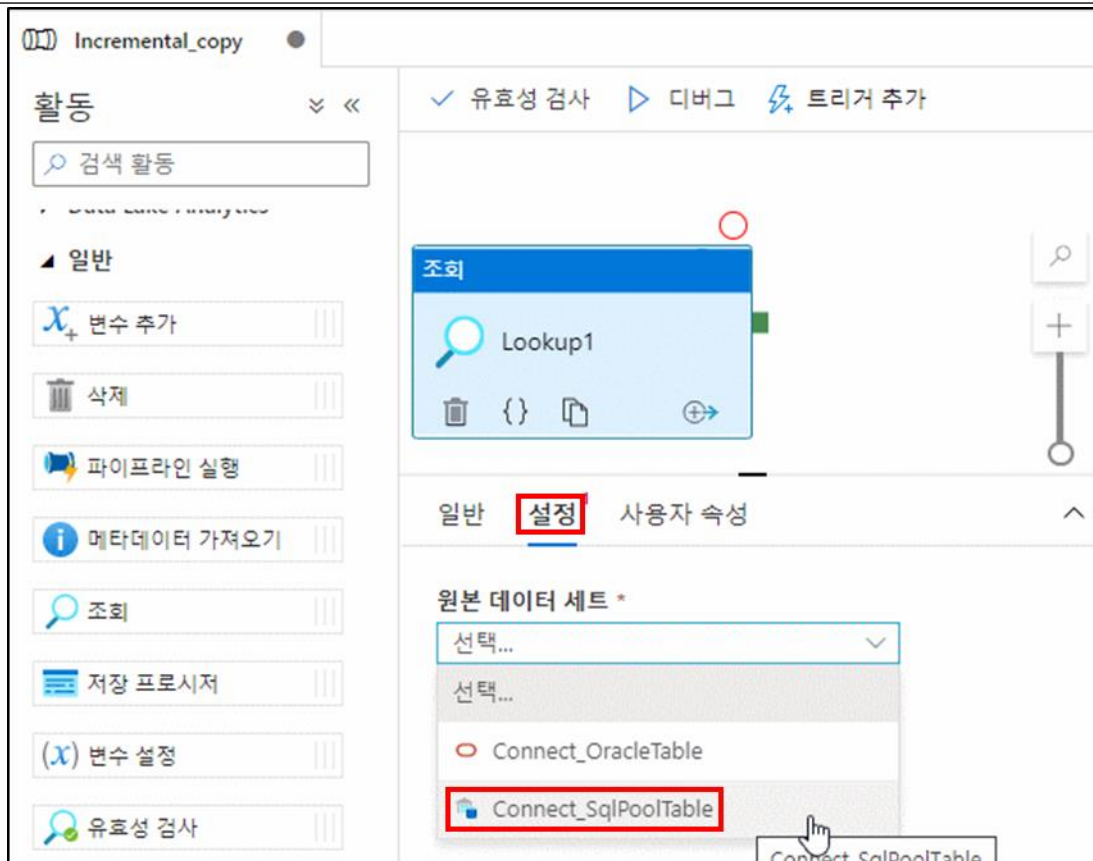
2. 파이프 라인의 이름을 **Incremental_copy** 로 기입합니다. 일반 항목 중 **조회** 박스를 Drag & Drop 하여 캔버스 중앙에 가져다 놓습니다.



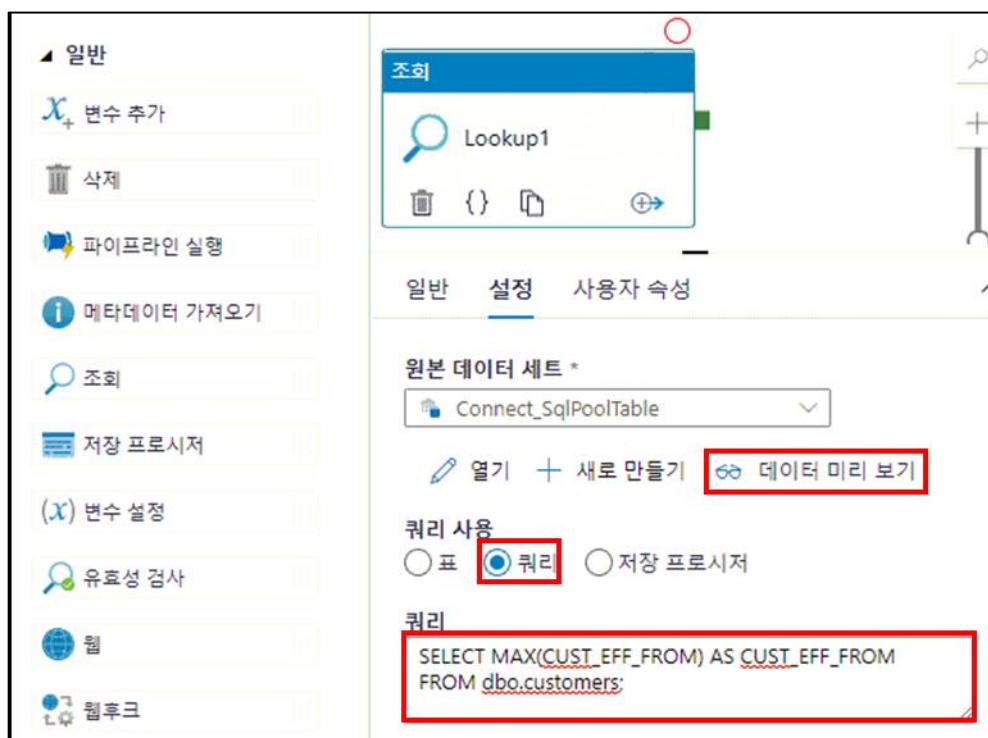
3. 조회 Activity 를 선택하여 설정화면을 **활성화**합니다. 일반 설정은 그대로 두고 **설정** 항목으로 이동합니다.



4. 설정 항목에서 원본 데이터 세트는 앞 작업에서 sqlpool 에 접근하기 위해 생성했던 **Connect_SqlPoolTable**을 선택합니다.



5. 다음으로 쿼리 사용을 선택하고 쿼리 박스 안에 **SELECT MAX(CUST_EFF_FROM) AS CUST_EFF_FROM FROM dbo.customers;** 쿼리를 입력합니다. 데이터 미리보기 버튼을 눌러 결과 값이 정상적으로 출력되는지 확인 합니다.



데이터 미리 보기

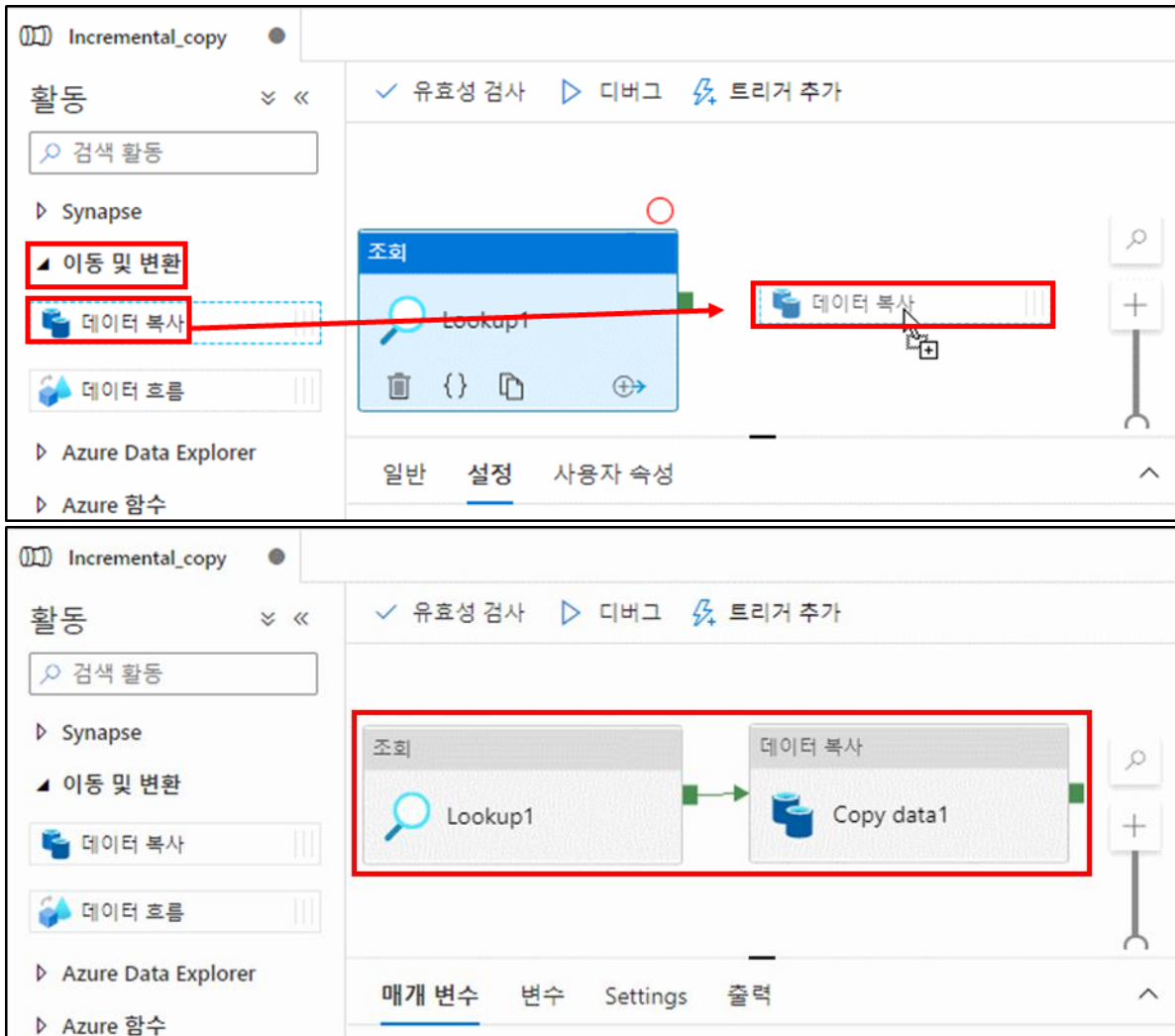
연결된 서비스: sqlpool

개체:

CUST_EFF_FROM

1998-01-01T00:00:00

- 다음으로 이동 및 변환 Activity의 **데이터 복사**를 화면 중앙에 Drag & Drop 으로 끌어 놓고 Lookup1 박스와 **연결**합니다.

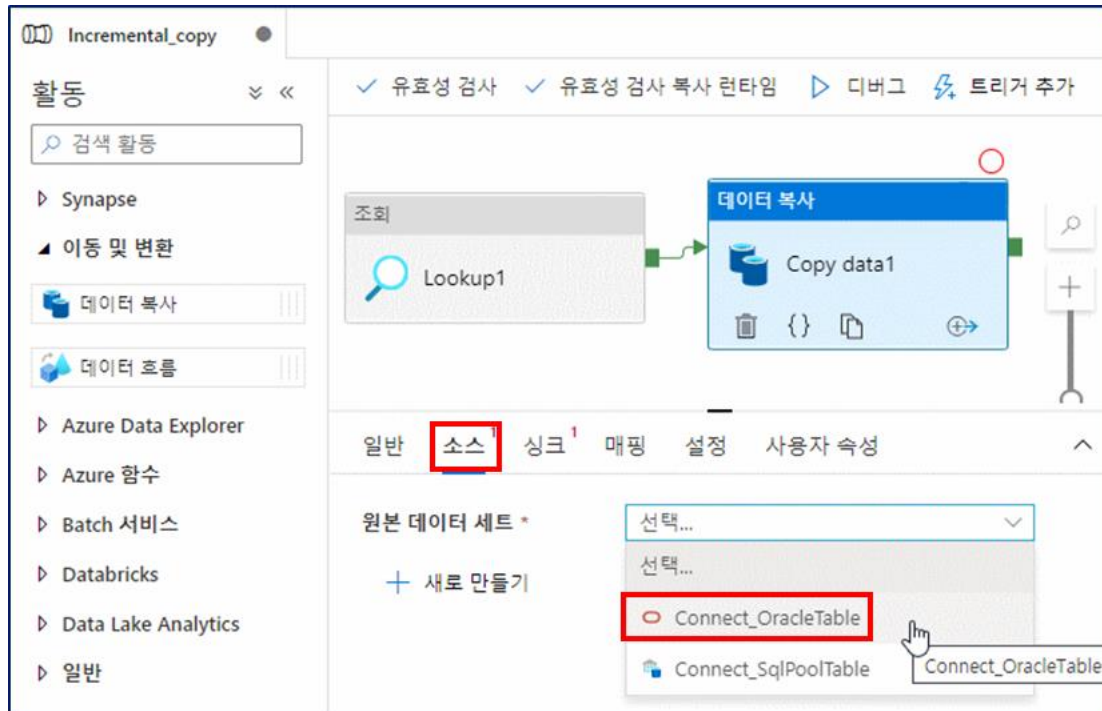


The image shows two screenshots of the 'Incremental_copy' activity editor in a workflow tool.

Top Screenshot: The '활동' (Activities) pane on the left shows a list of activities. '이동 및 변환' (Move and Transform) is expanded, and '데이터 복사' (Copy Data) is highlighted with a red box. A red arrow points from this box to the '데이터 복사' activity in the main canvas. The main canvas shows a 'Lookup1' activity in a '조회' (Query) container. The bottom tabs are '일반' (General), '설정' (Settings), and '사용자 속성' (User Properties).

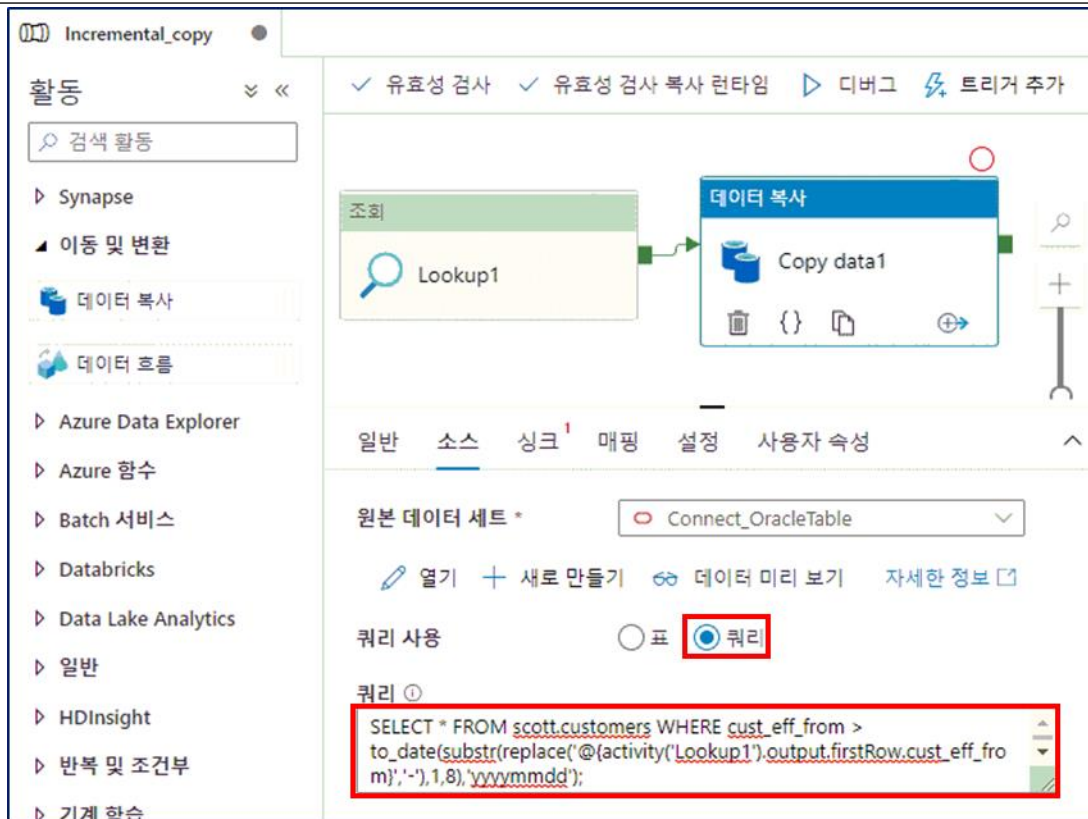
Bottom Screenshot: The '데이터 복사' activity has been added to the main canvas and is now connected to the 'Lookup1' activity by a green arrow. The activity is labeled 'Copy data1'. The bottom tabs are '매개 변수' (Parameters), '변수' (Variables), 'Settings', and '출력' (Output).

7. 데이터 복사 Activity 를 눌러 설정 항목을 **활성화**시킨 후 **소스** 항목으로 바로 이동합니다. **소스** 항목에서 원본 데이터 세트로 **Connect_OracleTable** 을 선택합니다. (전 작업에서 Oracle Server 에 접속하기 위해 만들어 사용하던 데이터세트)

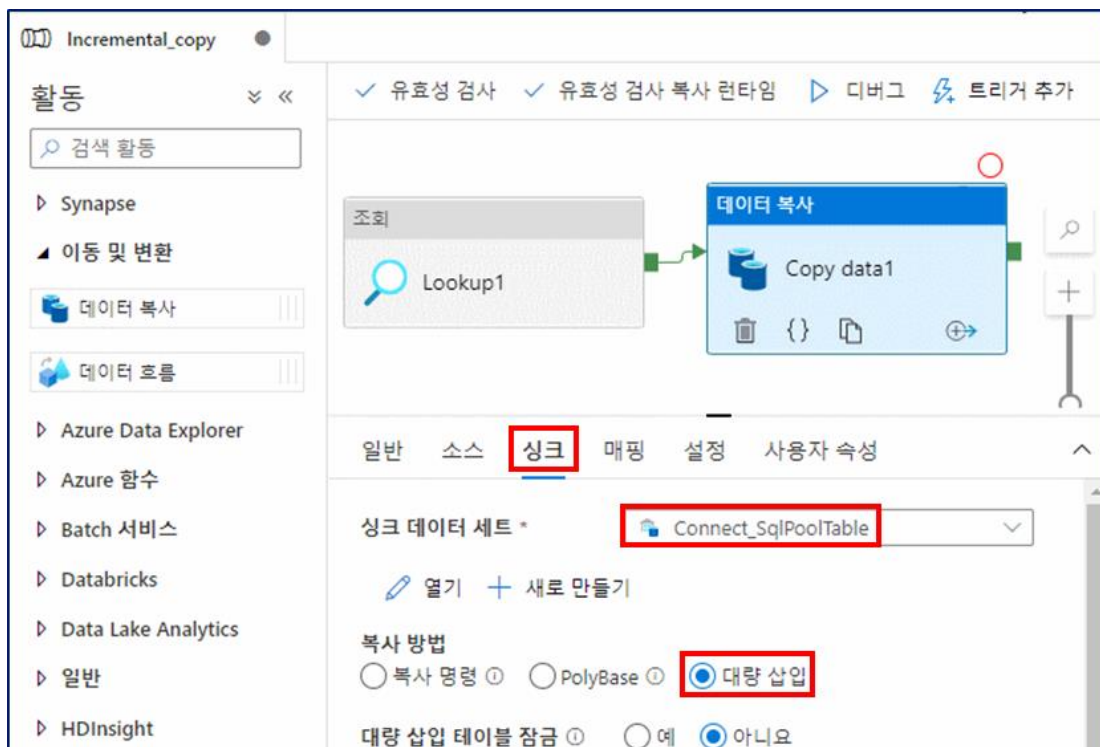


8. 쿼리 사용에서 쿼리를 선택하고 쿼리 박스에 다음의 쿼리를 기재합니다.

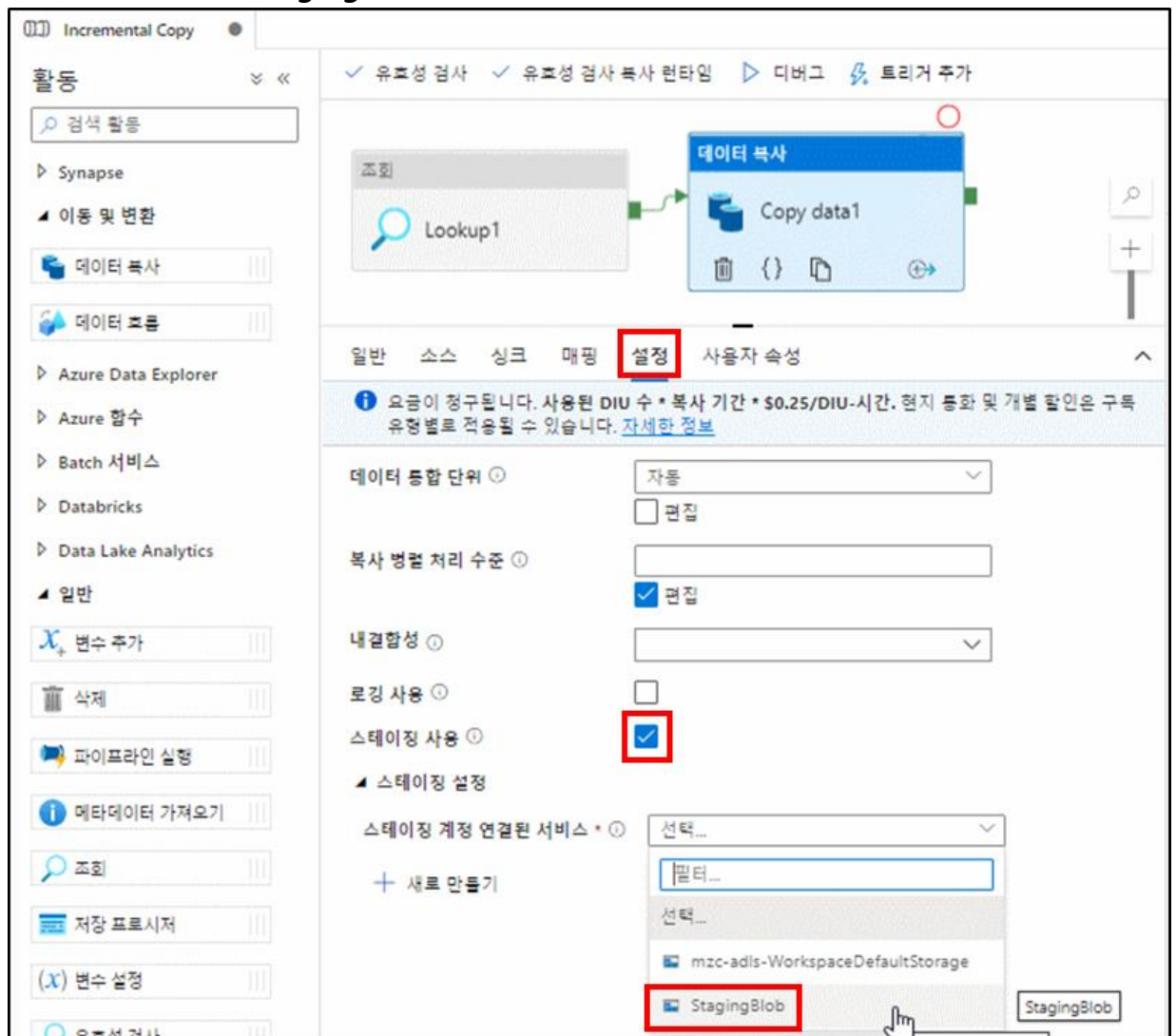
```
SELECT * FROM scott.customers WHERE cust_eff_from >
to_date(substr(replace('@{activity(''Lookup1'').output.firstRow.cust_eff_fro
m}','-'),1,8),'yyyymmdd');
```



9. 다음으로 싱크 항목으로 이동합니다. 싱크 데이터 세트 항목에서 **Connect_SqlPoolTable** 을 선택합니다. 복사 방법으로는 **대량 삽입** 을 선택 후 **설정** 항목으로 이동합니다.



10. 설정 항목에서 스테이징 사용을 선택하고 스테이징 계정 연결된 서비스에 기존에 생성했던 **StagingBlob**을 선택합니다.



11. 스토리지 경로 옵션에서 찾아보기 버튼을 눌러 **stage** 폴더를 선택합니다.

Incremental Copy

활동

검색 활동

Synapse

이동 및 변환

데이터 복사

데이터 흐름

Azure Data Explorer

Azure 함수

Batch 서비스

Databricks

Data Lake Analytics

일반

변수 추가

삭제

파이프라인 실행

메타데이터 가져오기

조회

저장 프로시저

변수 설정

유효성 검사

전

유효성 검사

유효성 검사 복사 런타임

디버그

트리거 추가

조회

Lookup1

데이터 복사

Copy data1

일반

소스

싱크

매핑

설정

사용자 속성

요금이 청구됩니다. 사용된 DIU 수 * 복사 기간 * \$0.25/DIU-시간. 현지 통화 및 개별 환인은 구독 유형별로 적용될 수 있습니다. [자세한 정보](#)

데이터 통합 단위

자동

편집

복사 병렬 처리 수준

편집

내결합성

로깅 사용

스테이징 사용

스테이징 설정

스테이징 계정 연결된 서비스

StagingBlob

연결 테스트

편집

새로 만들기

통합 런타임

AutoResolveIntegrationRuntime

편집

스토리지 경로

찾아보기

압축 사용

폴더 선택

파일 또는 폴더를 선택하세요.

루트 폴더

cosmosdemo

exportfile

models

mzcsynapse

pipeline

stage

1~6/6개 항목 표시

확인

취소

12. 설정 완료한 Pipeline 을 모두 게시합니다.

모두 게시

Incremental Copy

활동

검색 활동

Synapse

이동 및 변환

데이터 복사

데이터 흐름

Azure Data Explorer

Azure 함수

Batch 서비스

Databricks

Data Lake Analytics

일반

변수 추가

삭제

파이프라인 실행

메타데이터 가져오기

조회

저장 프로시저

변수 설정

유효성 검사

유효성 검사 복사 런타임

디버그

트리거 추가

조회

Lookup1

데이터 복사

Copy data1

일반

소스

싱크

매핑

설정

사용자 속성

요금

이 청구됩니다. 사용된 DIU 수 * 복사 시간 * \$0.25/DIU-시간. 현지 통화 및 개별 할인은 구독 적용될 수 있습니다. [자세한 정보](#)

데이터 통합 단위

자동

편집

복사 병렬 처리 수준

편집

내결합성

로깅 사용

스테이징 사용

스테이징 설정

스테이징 계정 연결된 서비스

StagingBlob

연결 테스트

편집

새로 만들기

통합 런타임

AutoResolveIntegrationRuntime

편집

스토리지 경로

stage

찾아보기

모두 게시

보류 중인 모든 변경 내용을 라이브 환경에 게시합니다.

보류 중인 변경 내용 (1)

이름

변경

파이프라인

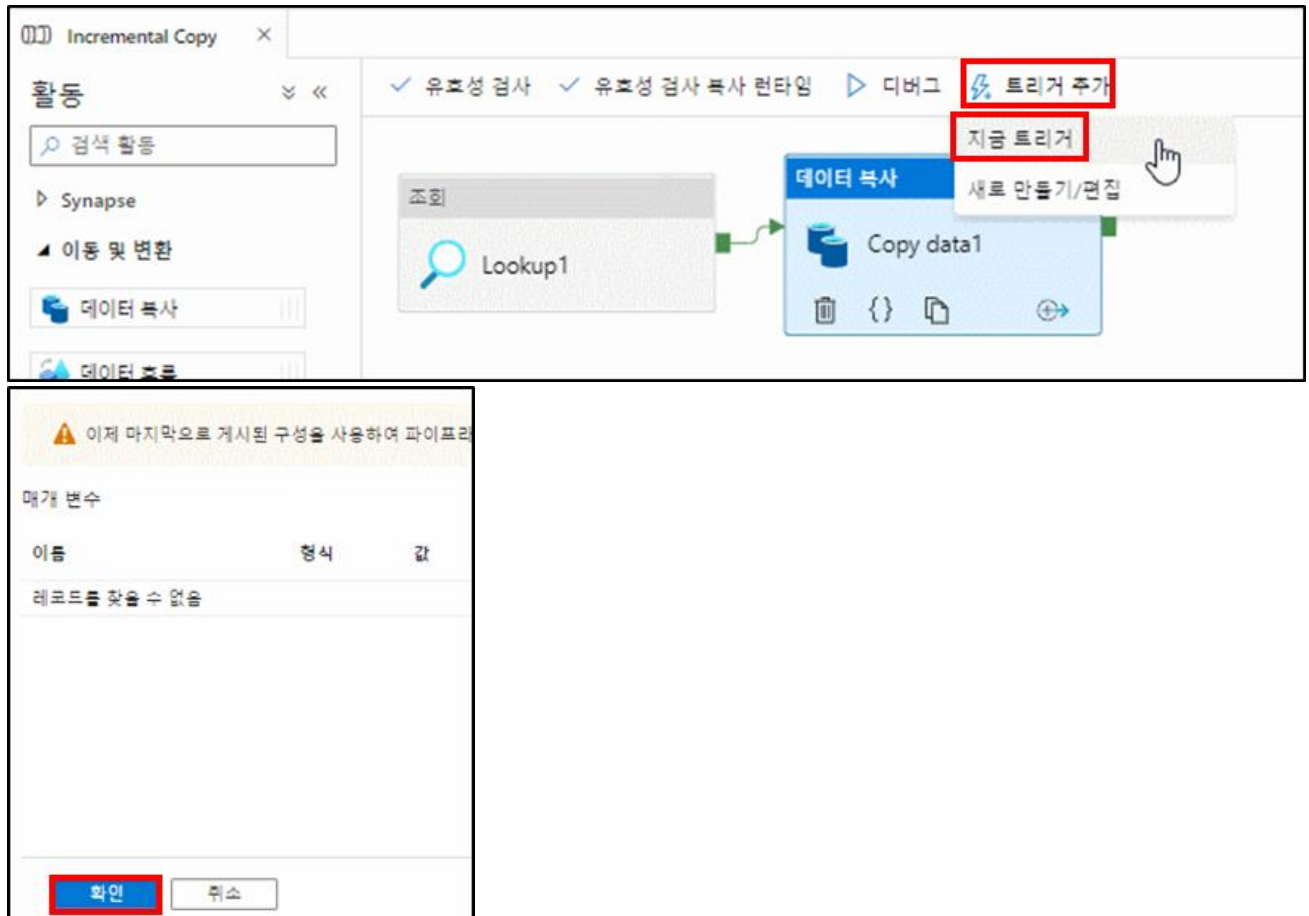
Incremental Copy

(신규)

게시

취소

13. Pipeline 이 게시 완료되면, **지금 트리거**를 실행하도록 합니다.



14. Pipeline 이 수행되는 과정을 모니터링 합니다. 모니터 **Hub** 탭 à **파이프라인 실행** à **Incremental Copy** 를 순서대로 선택합니다.




15. Incremental Copy 파이프라인 내부에서 Lookup1, Copy data1 의 진행 상태를 확인할 수 있습니다. 좀 더 상세한 내용을 확인하기 위해 Copy data1 의 안경모양 이미지(자세히 버튼)를 클릭합니다.

Incremental Copy

목록 Gantt

다시 실행 작업에서 다시 실행 Copy data1 실패한 작업에서 다시 실행 새로 고침 파이프라인 업데이트




활동 실행
파이프라인 실행 ID e72e934f-8229-4090-8e01-83b318b21665

모든 상태 ▾

1~2/2개 항목 표시

작업 이름	작업 유형	실행 시작 ↑↓	기간	상태	오류
Copy data1	데이터 복사	9/26/21, 오후 1:44:00	00:00:18	✓ 성공	
Lookup1	조회	9/26/21, 오후 1:43:56	00:00:04	✓ 성공	

16. 파이프라인의 실행 단계와 결과를 좀 더 상세히 확인할 수 있습니다.



원본 데이터: 72.363 KB
원본 행: 130
최대 연결: 1

쓰 데이터: 67.285 KB
기록된 행: 130
최대 연결: 2

복사 기간: 00:00:14
처리량: 5.169 KB/초

Oracle → Azure Blob Storage

시작 시간: 9/26/21, 오후 1:44:01
사용된 DIU: 4
사용된 병렬 복사본: 1

기간: 00:00:07

자세히	작업 기간	총 기간
✓ 큐		00:00:05
✓ 이전	<div> 첫 번째 바이트까지의 시간: 00:00:00 원본에서 읽기: 00:00:00 싱크에 쓰기: 00:00:00 </div>	00:00:01

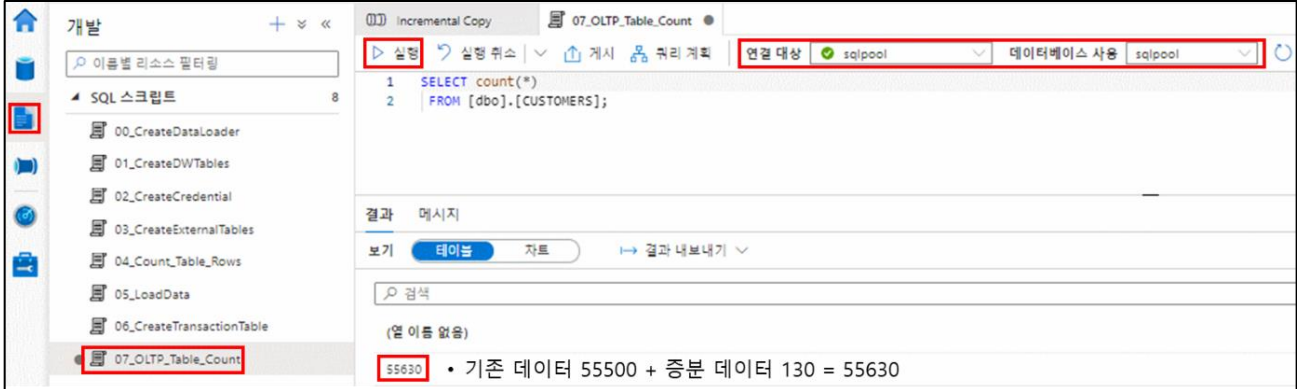
Azure Blob Storage → Azure Synapse Analytics

시작 시간: 9/26/21, 오후 1:44:08
사용된 DIU: 4
사용된 병렬 복사본: 1

기간: 00:00:07

자세히	작업 기간	총 기간

17. 마지막으로 증분 데이터가 정상적으로 들어왔는지 확인합니다. **개발 Hub** 탭으로 돌아가 스크립트 **07_OLTP_Table_Count** 를 실행하여 총 건수를 확인합니다. 증분 데이터는 Oracle 서버에 30 초마다 랜덤 데이터로 10 행이 생성되게 설정하였습니다.



The screenshot displays the MEGAZONE CLOUD development hub interface. On the left sidebar, the '개발' (Development) tab is active, and the 'SQL 스크립트' (SQL Scripts) section is expanded. The script '07_OLTP_Table_Count' is selected and highlighted. The main panel shows the script's execution details. At the top, the 'Incremental Copy' tab is active, and the script '07_OLTP_Table_Count' is selected. The '연결 대상' (Connection Target) is set to 'sqlpool', and the '데이터베이스 사용' (Database Used) is also 'sqlpool'. The script content is as follows:

```
1 SELECT count(*)
2 FROM [dbo].[CUSTOMERS];
```

Below the script, the '결과' (Results) tab is selected, showing the execution results. The '보기' (View) button is set to '테이블' (Table). The results show a single row with the value '55630'. A message below the results states: '(열 이름 없음)' (No column name) and '55630 • 기존 데이터 55500 + 증분 데이터 130 = 55630'.