

# 서버리스 세미나

**AWS** 환경에서 **serverless**란?

# 서버리스(Serverless)

- 개발자가 서버를 관리할 필요가 없는 클라우드 네이티브 개발 모델
- 장점
  - 애플리케이션 구동할 때만 비용 지출
  - 일반 작업들(파일시스템 관리, 용량 관리)을 클라우드 서비스로 사용할 수 있음

클라우드 네이티브 애플리케이션 : 애플리케이션 배포 위험은 낮추고 속도와 유연성 및 품질이 높은 클라우드 컴퓨팅 모델을 활용하도록 설계된 애플리케이션

# 기존 환경

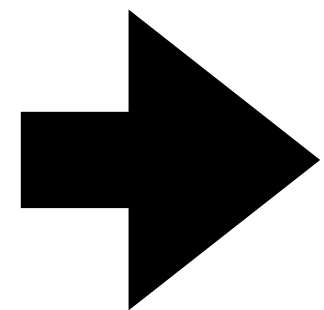
- 새 서비스 => 인프라 배포 필요
- 주기적 버전 관리 필요
- 용량 확장 계획 상시 대비
- 취약점 발견 => 패치를 통한 주기적 관리 필요

패치(patching) : 수정/개선을 위해 [컴퓨터 프로그램](#)이나 데이터를 업데이트하는 소프트웨어

전지전능 or 업무분담

# AWS에서의 Serverless

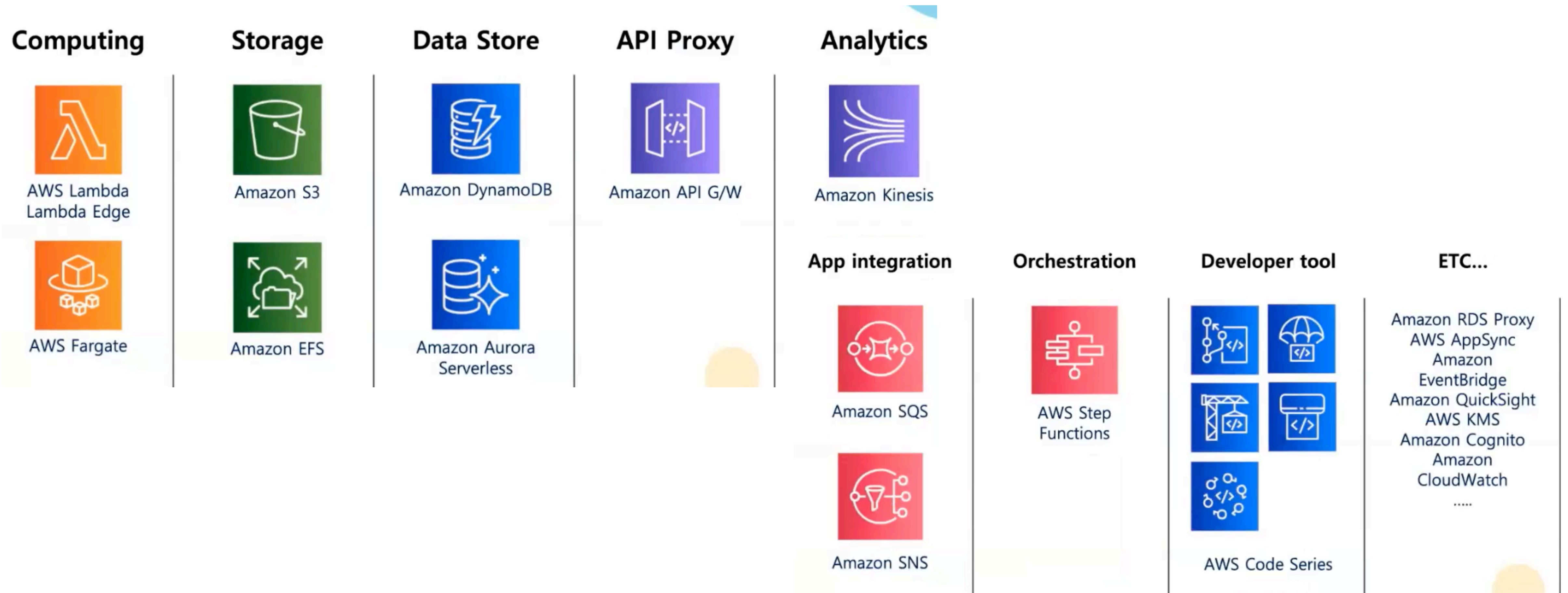
- 서비스 관리자가 많은 역할 X
- 클라우드 업체가 대신 관리



개발자들의 빠른 개발&배포 가능

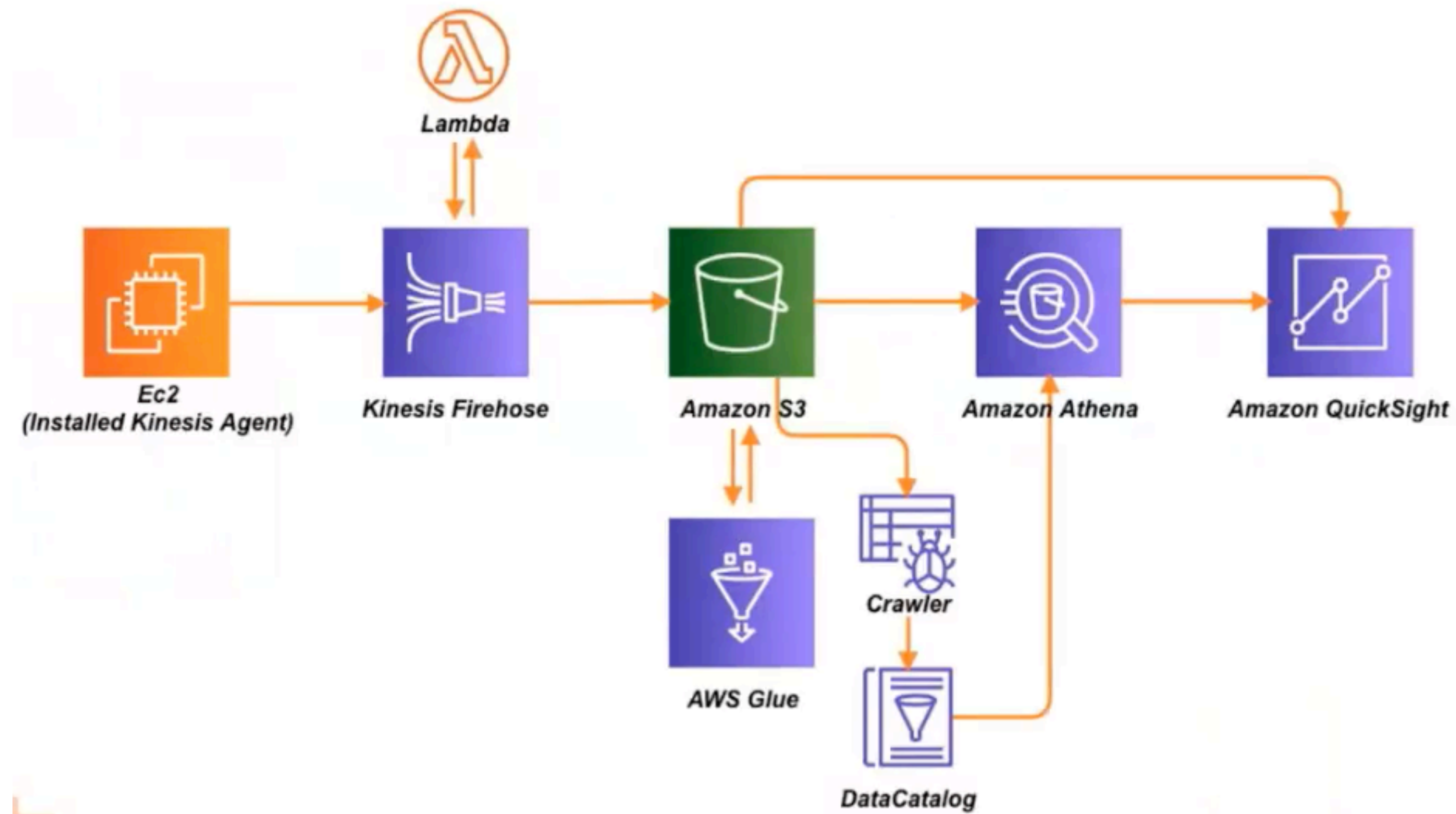
# AWS에서의 Serverless

다양한 서비스 제공



# AWS에서의 Serverless

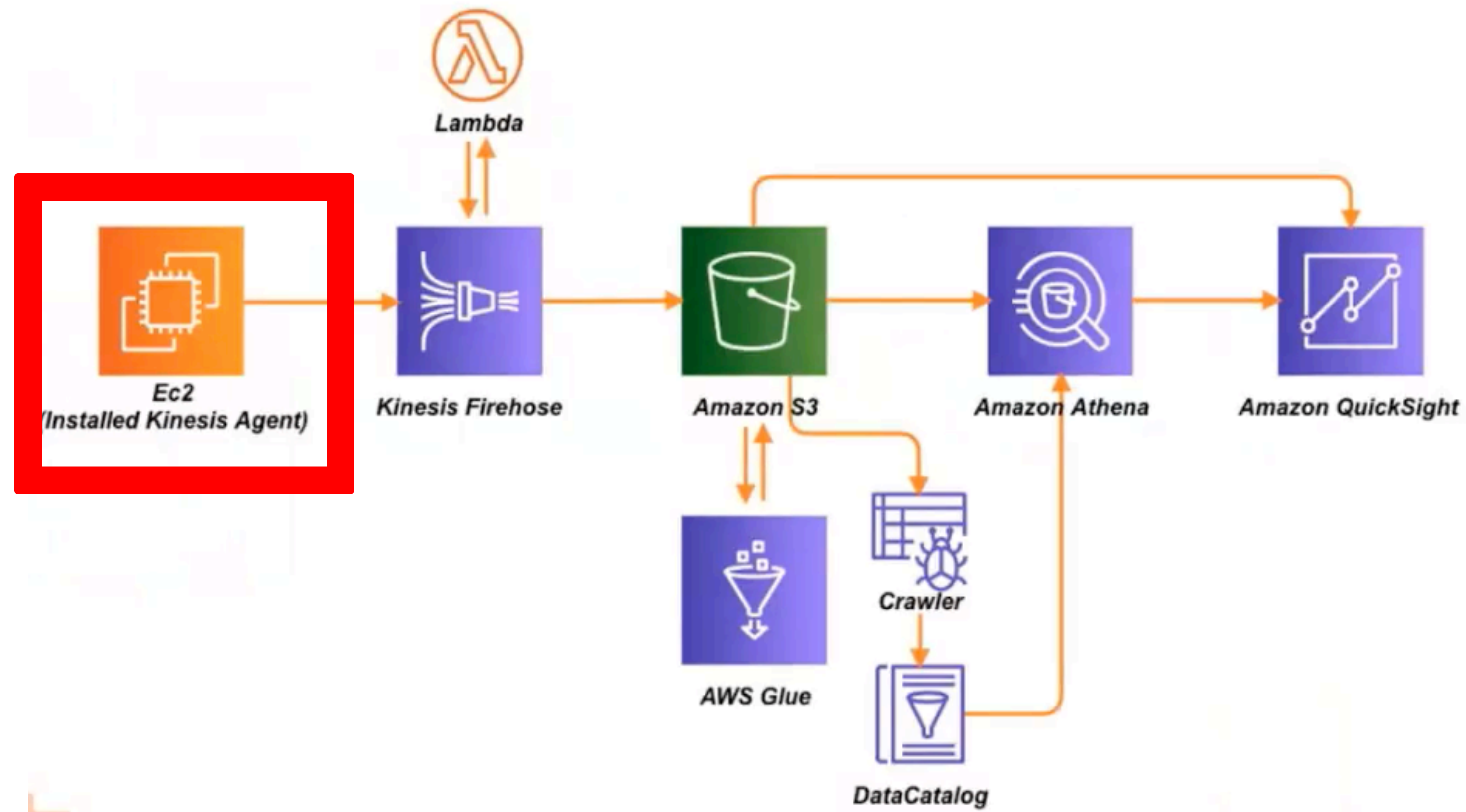
실시간 데이터 분석 플랫폼 구조





# AWS에서의 Serverless

## 실시간 데이터 분석 플랫폼 구조

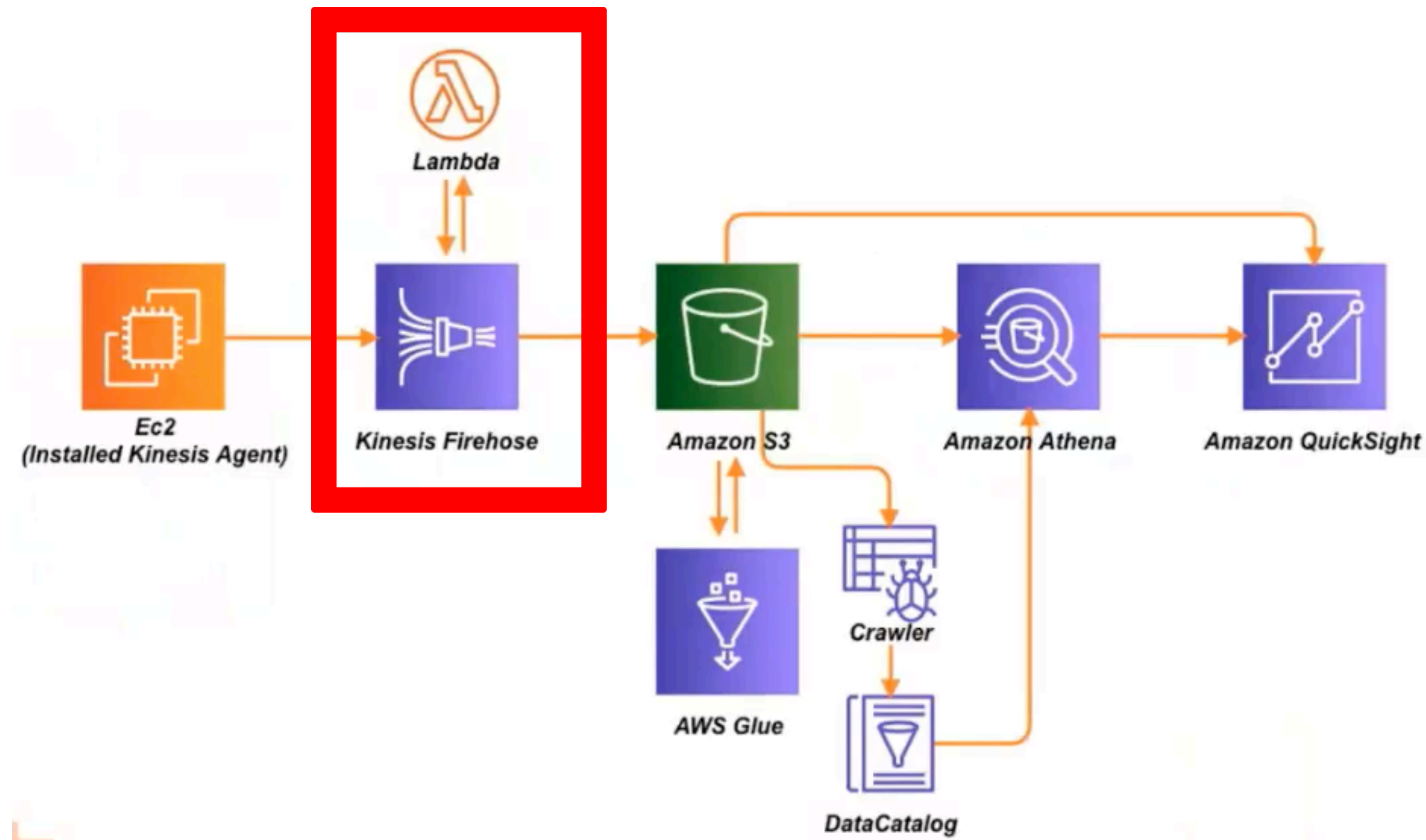


- 로그들을 실시간으로 firehose로 전달 by kinesis agent



# AWS에서의 Serverless

## 실시간 데이터 분석 플랫폼 구조

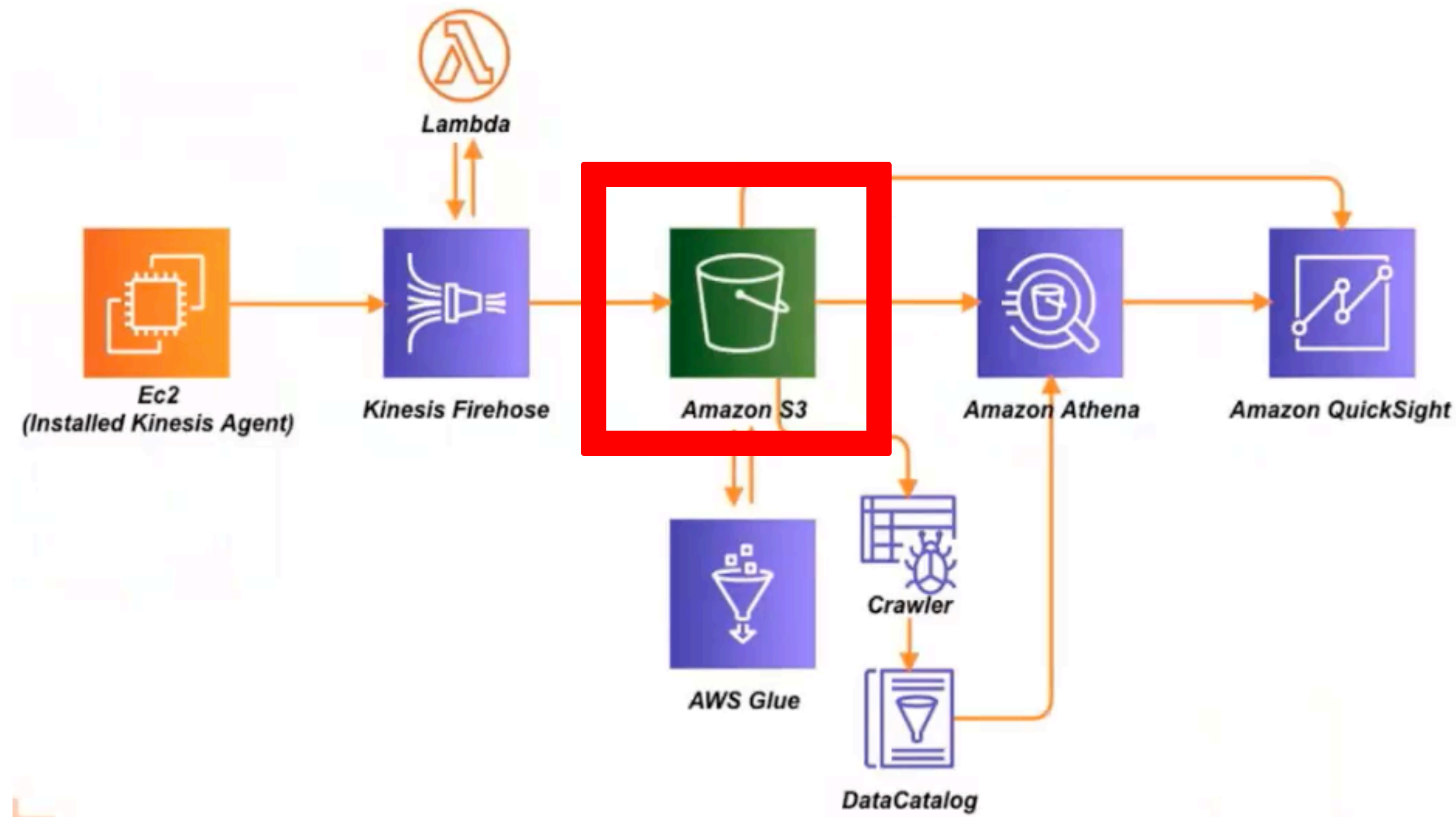


- Lamda를 통해서 데이터를 간단하게 transform함

Lamda : 스트림 데이터에 대해 형식 변환, 전환, 강화 또는 필터링이 필요한 경우  
AWS Lambda 함수를 사용하여 데이터를 사전 처리

# AWS에서의 Serverless

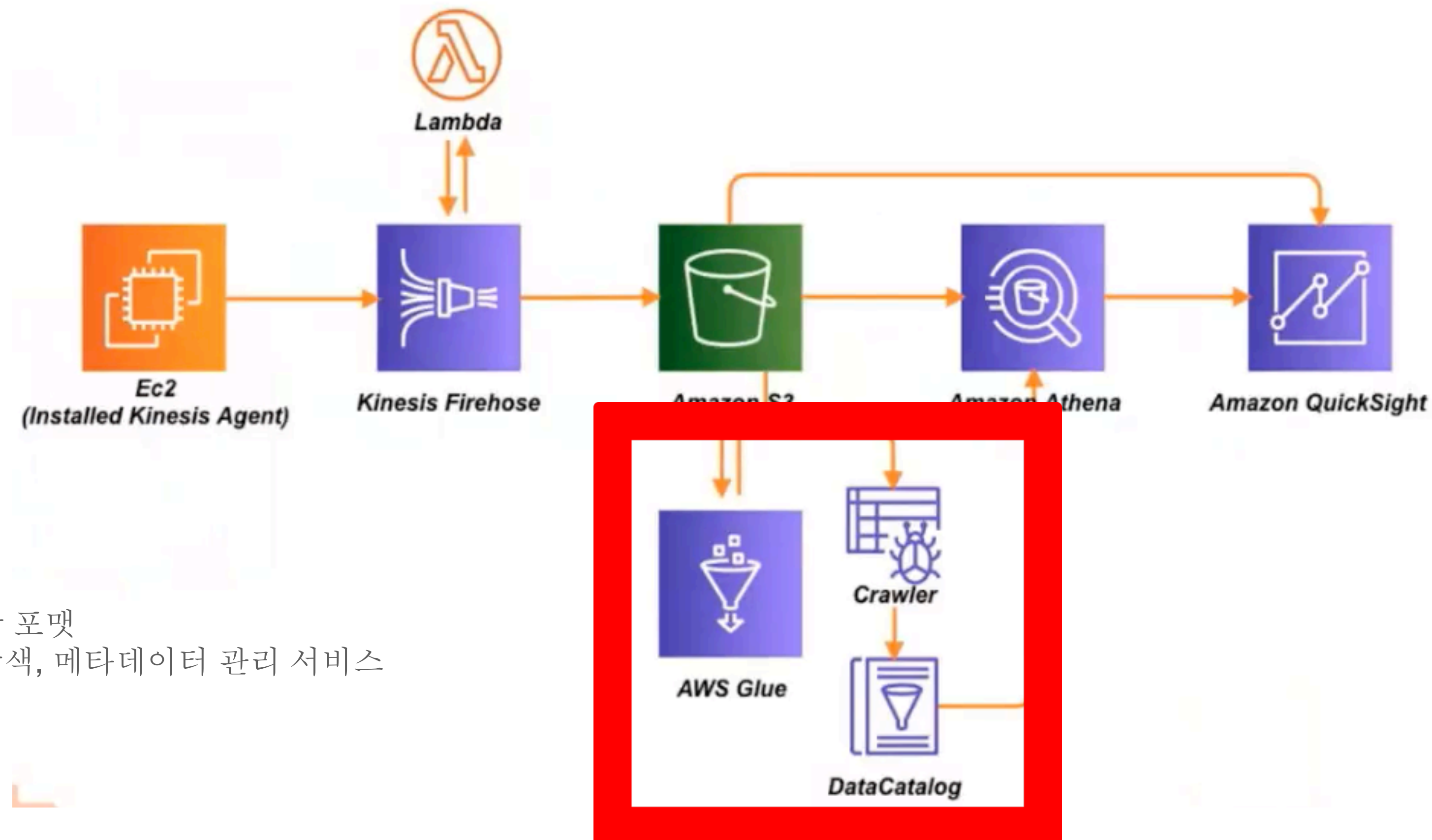
## 실시간 데이터 분석 플랫폼 구조



- 수집된 데이터는 s3라는 storage data에 저장

# AWS에서의 Serverless

## 실시간 데이터 분석 플랫폼 구조



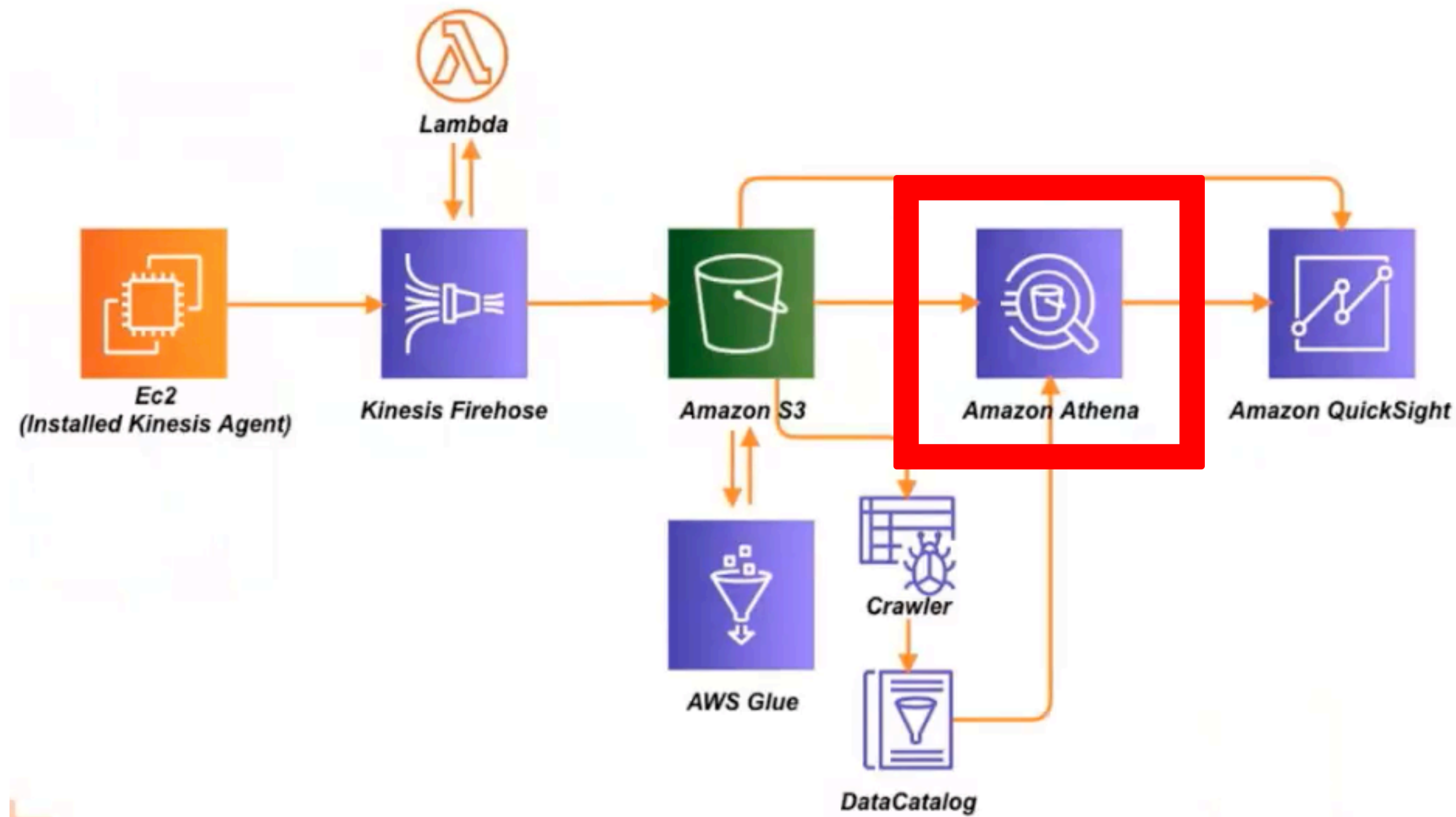
parquet(파케이) : columnar 저장 포맷

Data Catalog : 관리형 데이터 탐색, 메타데이터 관리 서비스

- 저장된 데이터에 대해서 parquet형태로 스키마를 저장하고 Catalog를 보여주는 서비스

# AWS에서의 Serverless

## 실시간 데이터 분석 플랫폼 구조

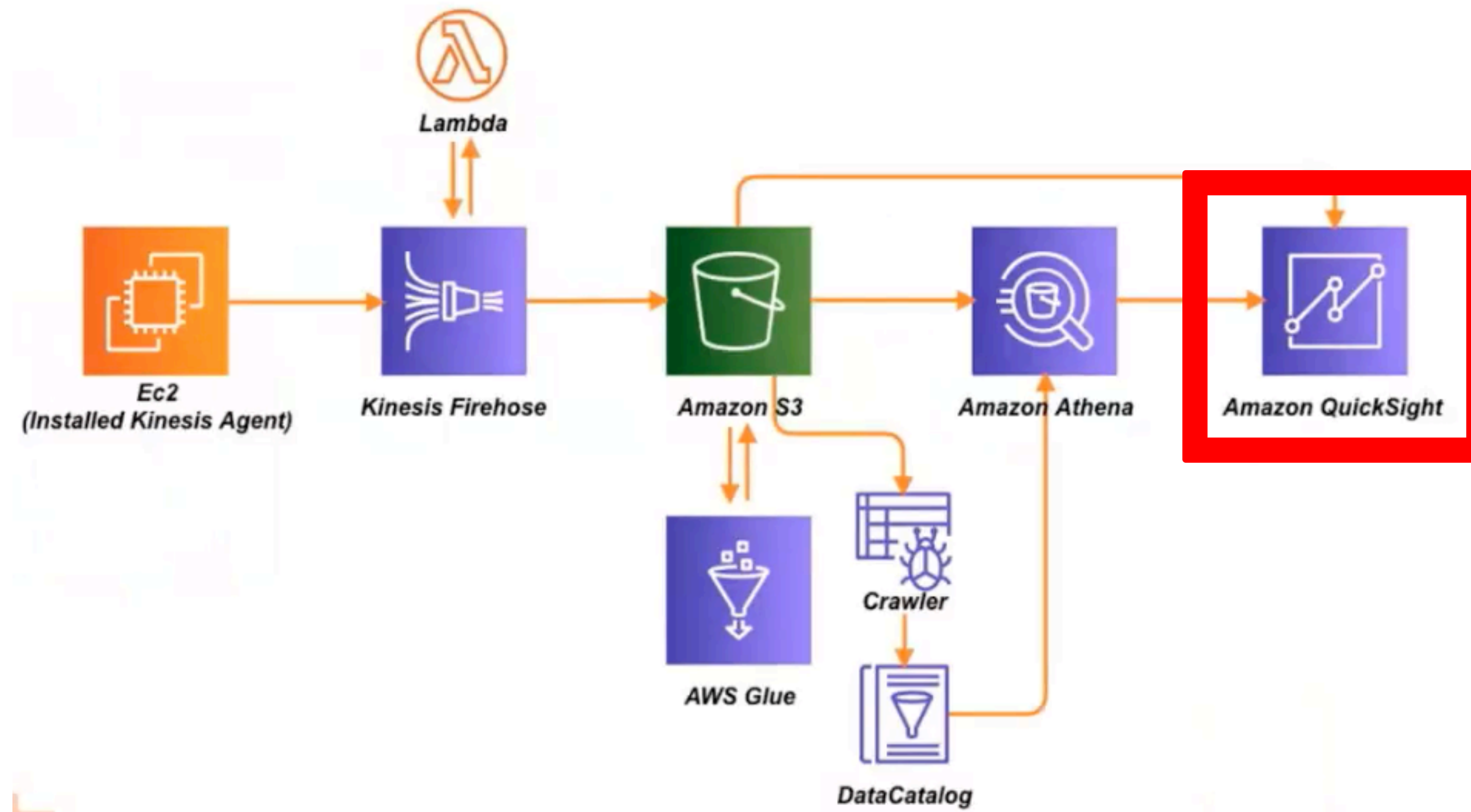


- 데이터들을 SQL로 조회하는 서비스
- 스캔한 데이터의 트랩 용량(TB) 비용당 비용이 발생함



# AWS에서의 Serverless

## 실시간 데이터 분석 플랫폼 구조



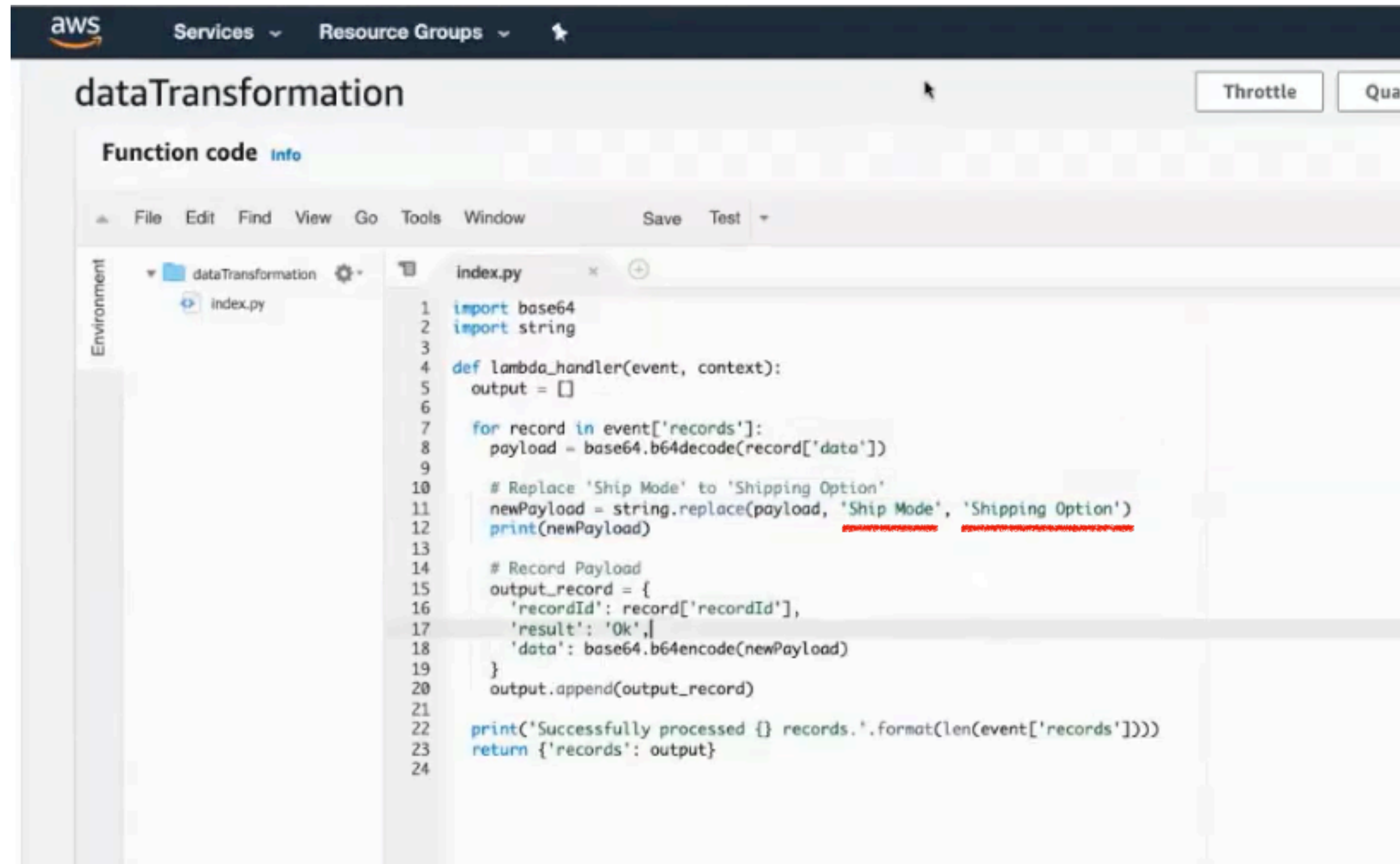
- Business intelligence tool -> 데이터 시각화

Business intelligence : 데이터를 통합/분석해 의사결정을 돕는 프로세스

# AWS에서의 Serverless

## 실시간 데이터 분석 플랫폼

- 람다로 데이터 변환



The screenshot displays the AWS Lambda console interface for a function named 'dataTransformation'. The 'Function code' tab is selected, showing a Python script in the 'index.py' file. The script is a Lambda handler that processes a list of records, decoding the payload, replacing 'Ship Mode' with 'Shipping Option', and encoding the result back to base64. The console header shows 'Services' and 'Resource Groups' dropdowns. The left sidebar shows the 'Environment' section with the 'dataTransformation' folder and 'index.py' file. The top right has 'Throttle' and 'Qual' buttons. The code editor has a menu bar with 'File', 'Edit', 'Find', 'View', 'Go', 'Tools', 'Window', 'Save', and 'Test'.

```
1 import base64
2 import string
3
4 def lambda_handler(event, context):
5     output = []
6
7     for record in event['records']:
8         payload = base64.b64decode(record['data'])
9
10        # Replace 'Ship Mode' to 'Shipping Option'
11        newPayload = string.replace(payload, 'Ship Mode', 'Shipping Option')
12        print(newPayload)
13
14        # Record Payload
15        output_record = {
16            'recordId': record['recordId'],
17            'result': 'Ok',
18            'data': base64.b64encode(newPayload)
19        }
20        output.append(output_record)
21
22    print('Successfully processed {} records.'.format(len(event['records'])))
23    return {'records': output}
24
```

# AWS에서의 Serverless

## 실시간 데이터 분석 플랫폼

- firehose에서 partitioning

The image displays four sequential screenshots of the AWS S3 console, illustrating the partitioning structure of a bucket named 'realtime-analytics'.

- Year Partitioning:** The first screenshot shows the 'realtime-analytics' bucket with a search bar and 'Upload' and 'Create folder' buttons. A folder named '2020' is visible in the list.
- Month Partitioning:** The second screenshot shows the '2020' folder selected, displaying a list of sub-folders including '09'.
- Day Partitioning:** The third screenshot shows the '09' folder selected, displaying a list of sub-folders including '01', '14', and '17'.
- Time Partitioning:** The fourth screenshot shows the '01' folder selected, displaying a list of files including '05', '06', and '07'.

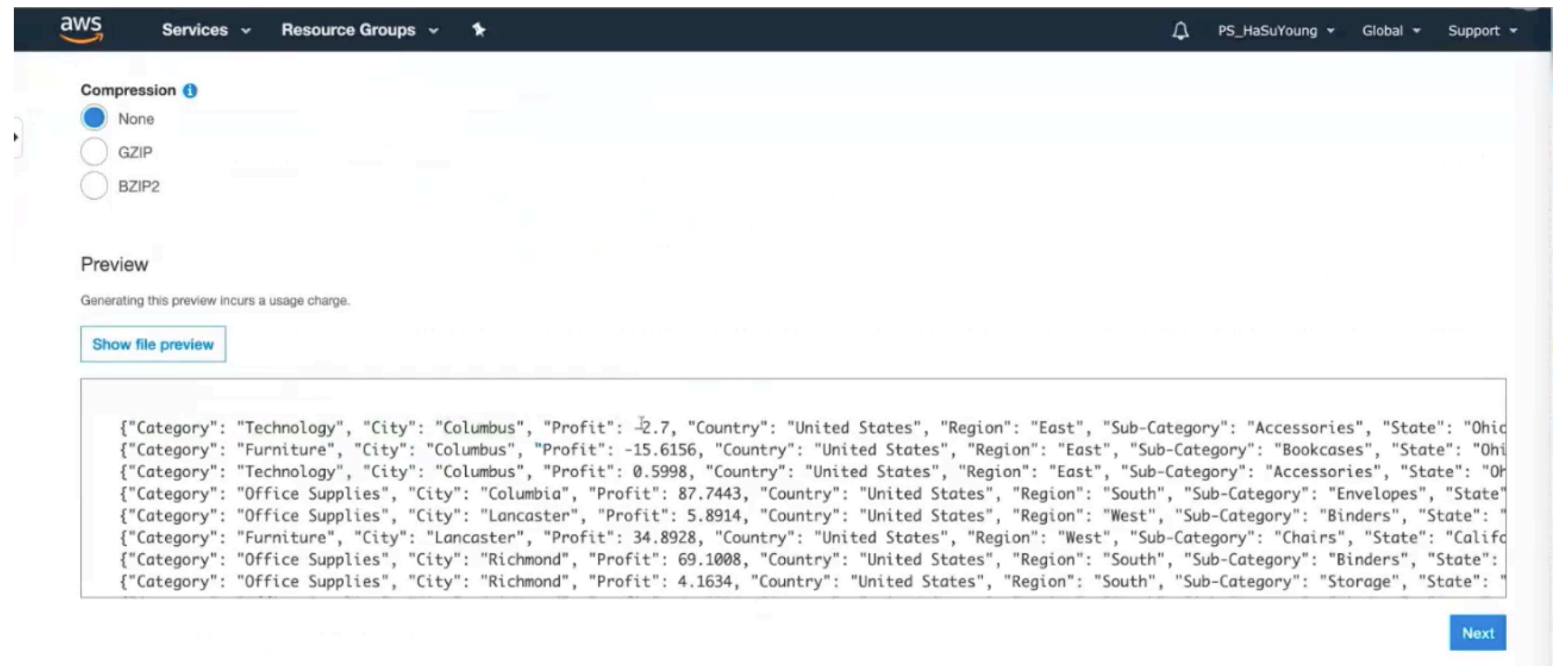
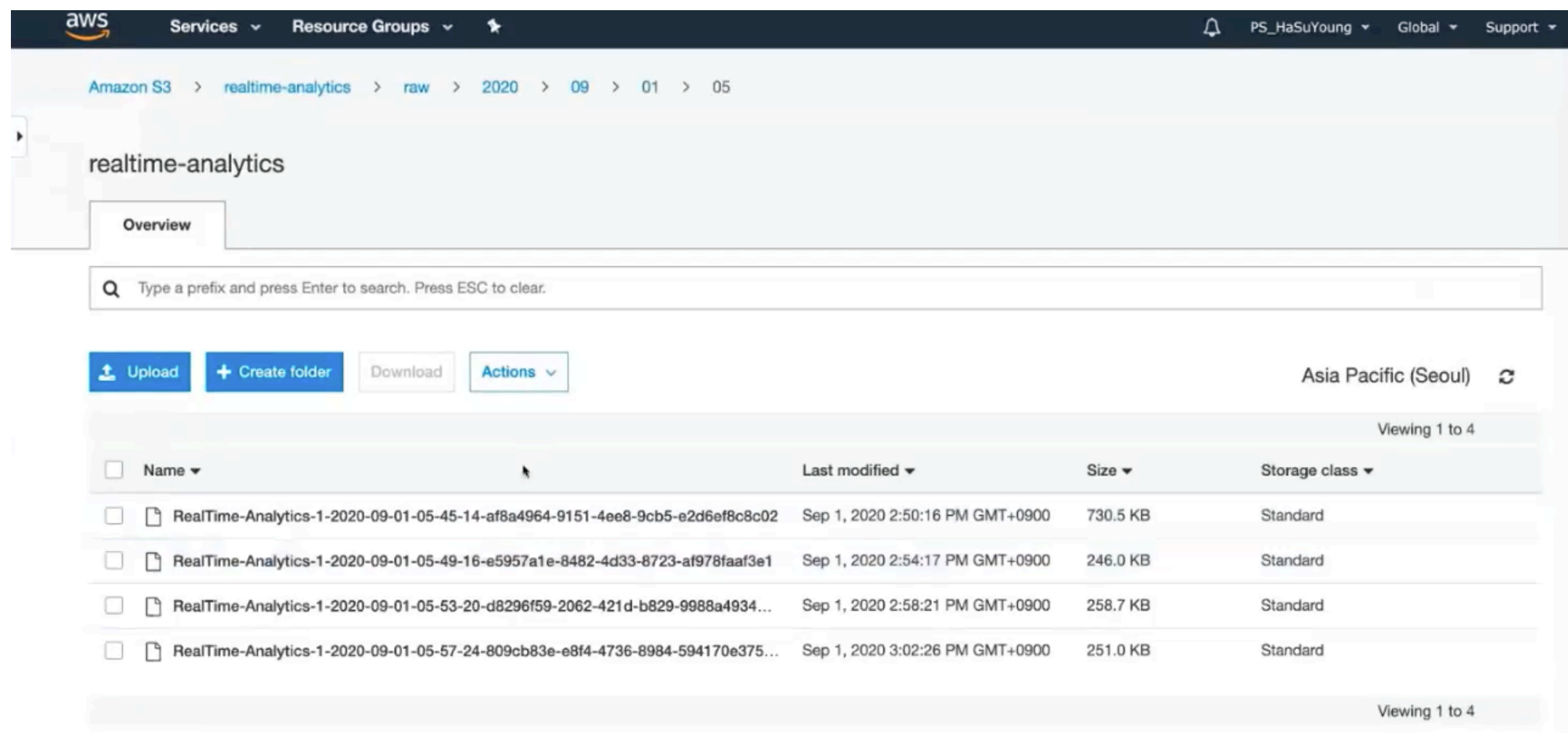
Below the screenshots, the labels '연' (Year), '월' (Month), '일' (Day), and '시간' (Time) are positioned, corresponding to the partitioning levels shown in the screenshots.



# AWS에서의 Serverless

## 실시간 데이터 분석 플랫폼

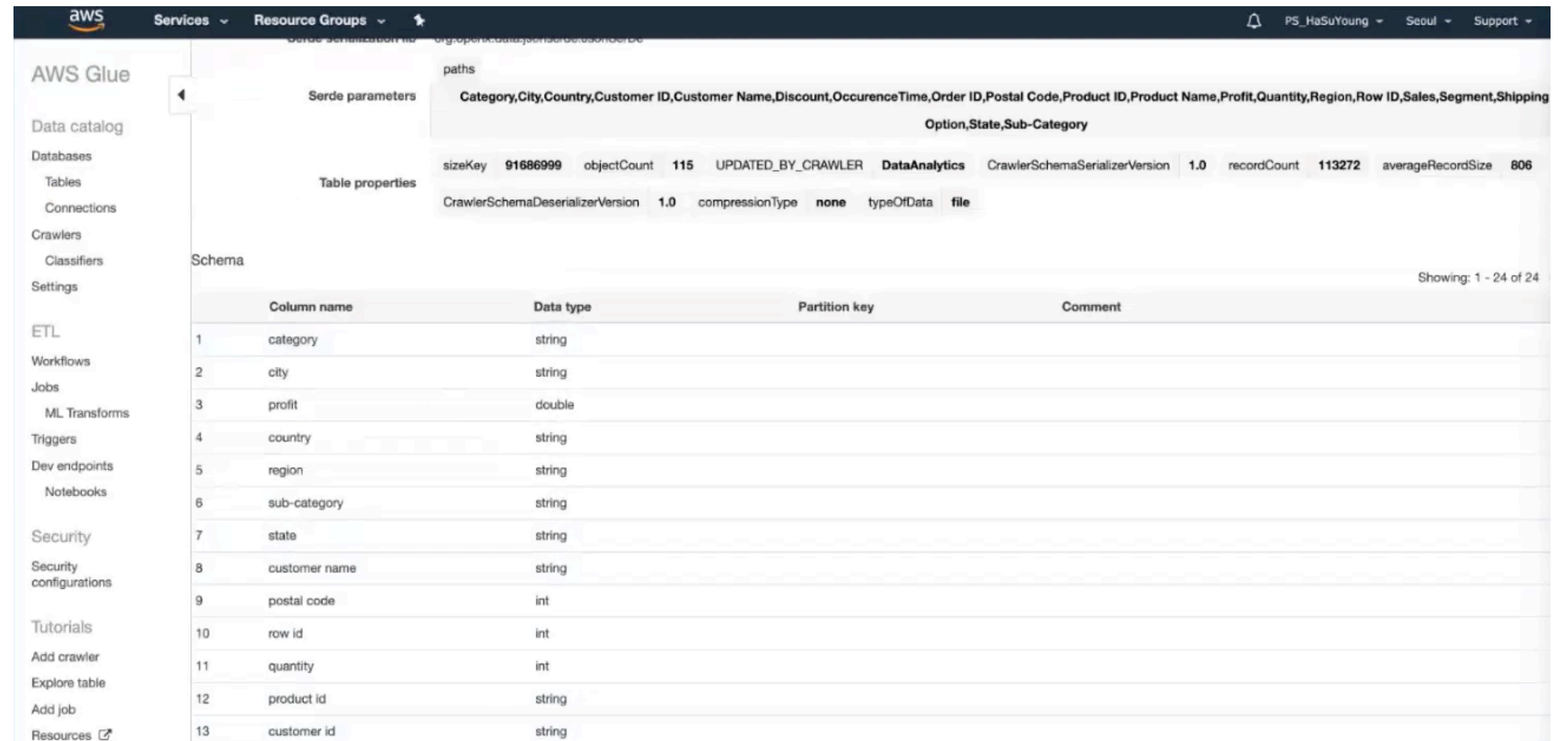
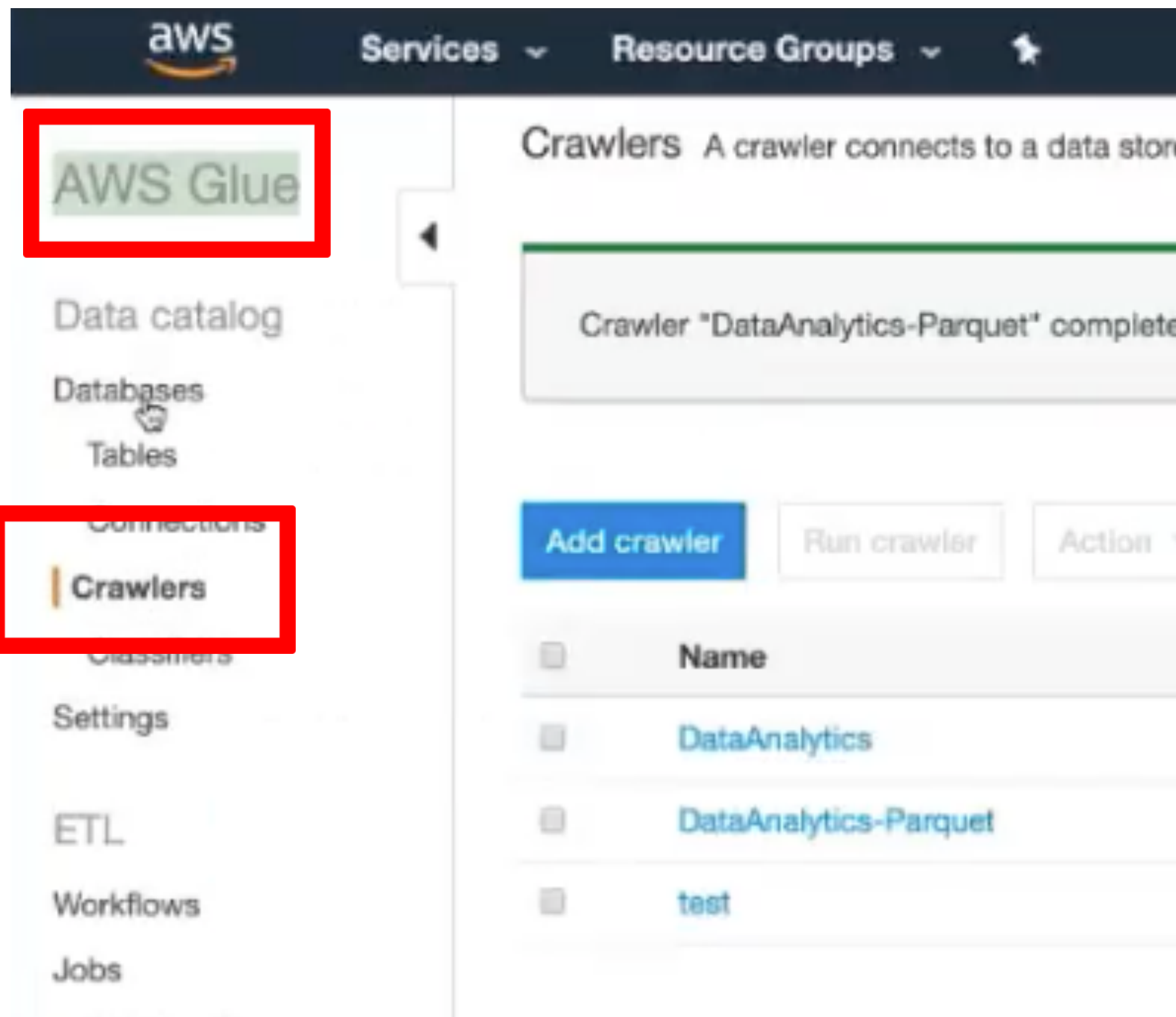
- 데이터가 csv파일 형태로 저장



# AWS에서의 Serverless

## 실시간 데이터 분석 플랫폼

- Glue로 ETL 작업(추출, 변환, 로드)



# AWS에서의 Serverless

## 실시간 데이터 분석 플랫폼

- Athena로 저장된 데이터를 SQL로 쿼리

The screenshot displays the AWS Athena Query Editor interface. On the left, the 'Data source' is set to 'AwsDataCatalog' and the 'Database' is 'dataanalytics'. Below this, a list of tables is shown, including 'raw (Partitioned)'. The main area contains a SQL query editor with the following query:

```
1 SELECT * FROM 'dataanalytics-parquet'. "parquet"
2
```

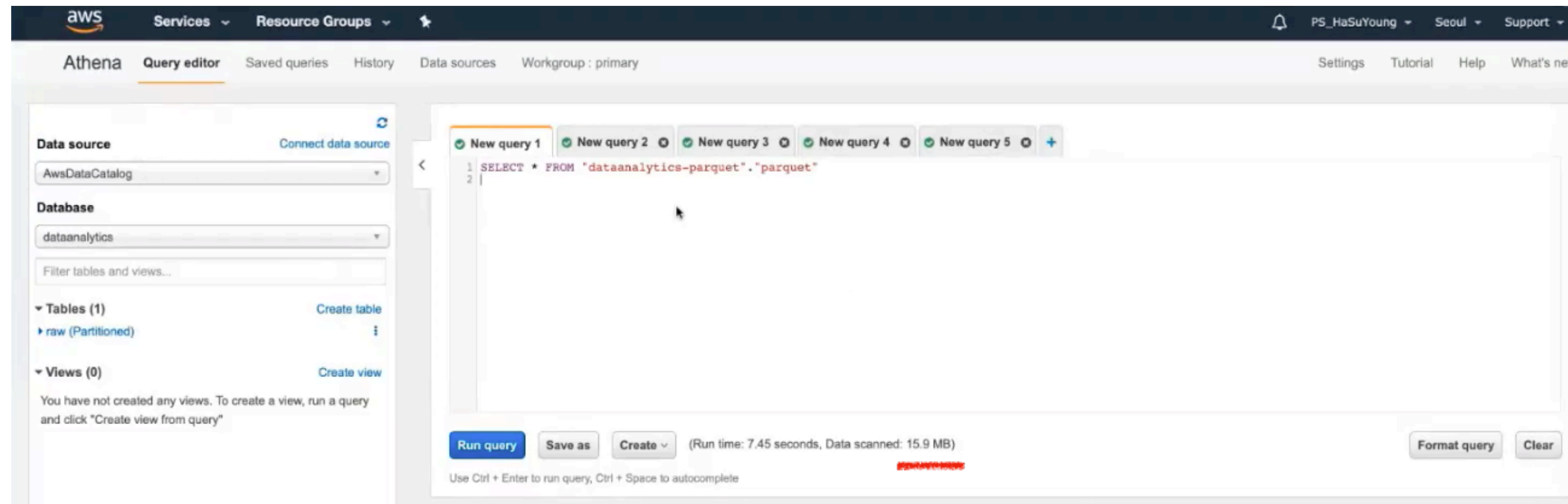
Below the query editor, there are buttons for 'Run query', 'Save as', and 'Create'. The 'Run query' button is highlighted. Below these buttons, the execution status is shown: '(Run time: 7.45 seconds, Data scanned: 15.9 MB)'. At the bottom, the 'Results' section displays a table with 13 columns: category, city, profit, country, region, sub-category, state, customer\_name, postal\_code, row\_id, quantity, product\_id, and customer\_id. The table contains 8 rows of data.

	category	city	profit	country	region	sub-category	state	customer_name	postal_code	row_id	quantity	product_id	customer_id
1	Office Supplies	Las Vegas	25.5798	United States	West	Binders	Nevada	Victor Preis	89115	395	3	OFF-BI-10004492	VP-21730
2	Office Supplies	Warwick	9.4924	United States	East	Storage	Rhode Island	Saphhira Shifley	2886	396	2	OFF-ST-10000798	SS-20140
3	Office Supplies	Warwick	6.2208	United States	East	Paper	Rhode Island	Saphhira Shifley	2886	397	2	OFF-PA-10002552	SS-20140
4	Office Supplies	Jackson	21.036	United States	Central	Supplies	Michigan	Harold Pawlan	49201	398	4	OFF-SU-10002573	HP-14815
5	Office Supplies	Houston	3.5952	United States	Central	Storage	Texas	Anna Gayman	77036	399	3	OFF-ST-10001580	AG-10675
6	Furniture	Houston	-317.1528	United States	Central	Bookcases	Texas	Anna Gayman	77036	400	4	FUR-BO-10004834	AG-10675
7	Office Supplies	Houston	-32.784	United States	Central	Storage	Texas	Anna Gayman	77036	401	4	OFF-ST-10000934	AG-10675
8	Technology	Houston	0.7198	United States	Central	Accessories	Texas	Anna Gayman	77036	402	2	TEC-AC-10000158	AG-10675

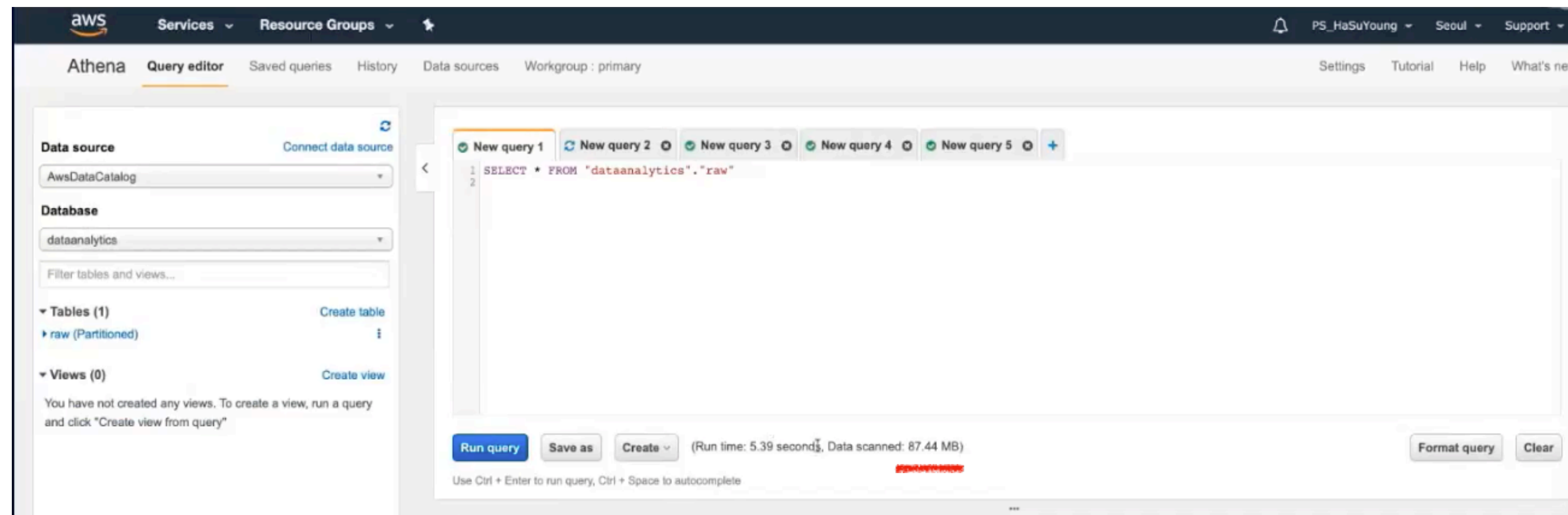


# AWS에서의 Serverless

## 실시간 데이터 분석 플랫폼



parquet 데이터



raw 데이터

감사합니다.