

데이터 파이프라인과 AI 알고리즘의 AWS 활용



강사 : 고병화

12. AWS Redshift



AWS Redshift

Amazon Redshift 는 완전 관리형, 페타 바이트급 Data Warehouse 서비스

Amazon Redshift

클라우드 데이터 웨어하우스를 위한 최고의 가격 대비 성능

[Amazon Redshift Serverless 시작하기](#) [Amazon Redshift 전문가와 연결](#)

[Amazon Redshift 무료로 사용해 보기](#)

데이터 이동이나 데이터 변환 없이 클릭 몇 번으로 모든 데이터에서 실시간 인사이트와 예측 인사이트를 얻고 데이터 사일로를 없앨 수 있습니다.

>

추가 비용 없이 바로 사용할 수 있는 획기적인 성능으로 다른 클라우드 데이터 웨어하우스보다 최대 5배 더 뛰어난 가격 대비 성능을 얻을 수 있습니다.

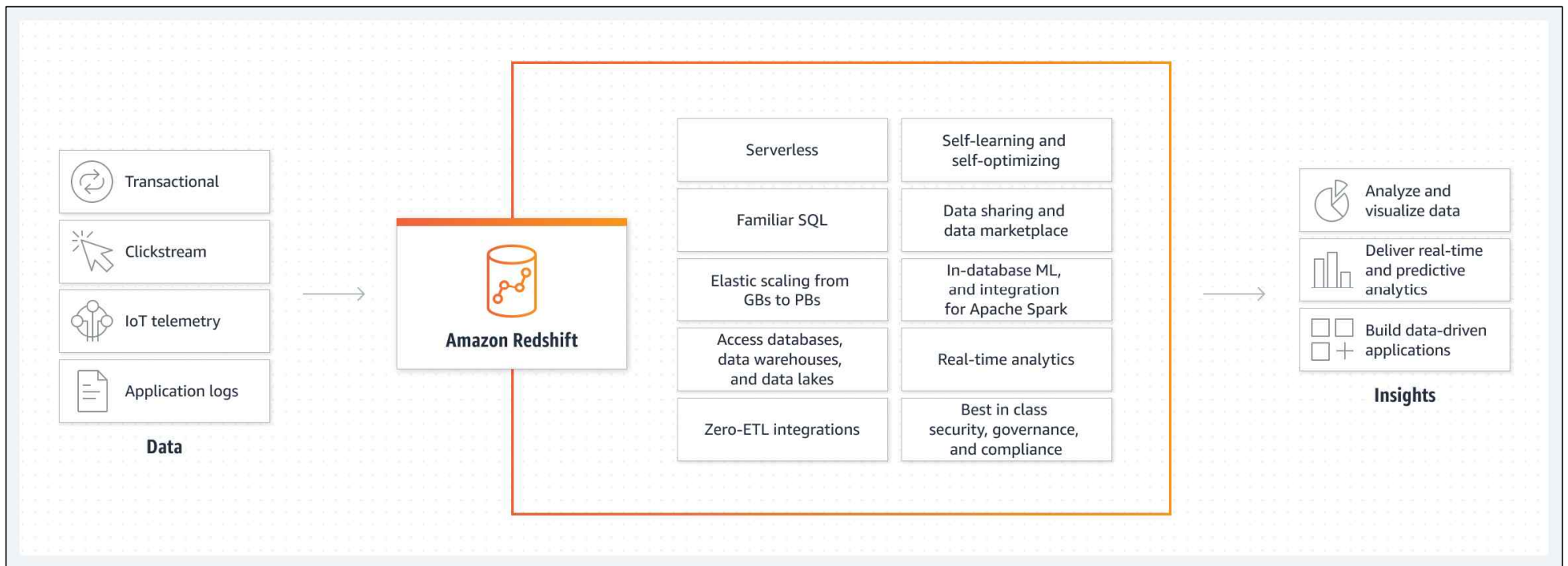
>

가장 안전하고 신뢰성이 높은 데이터 웨어하우스 서비스로 인프라를 관리해야 하는 부담 없이 몇 초 만에 인사이트를 얻을 수 있습니다.

>

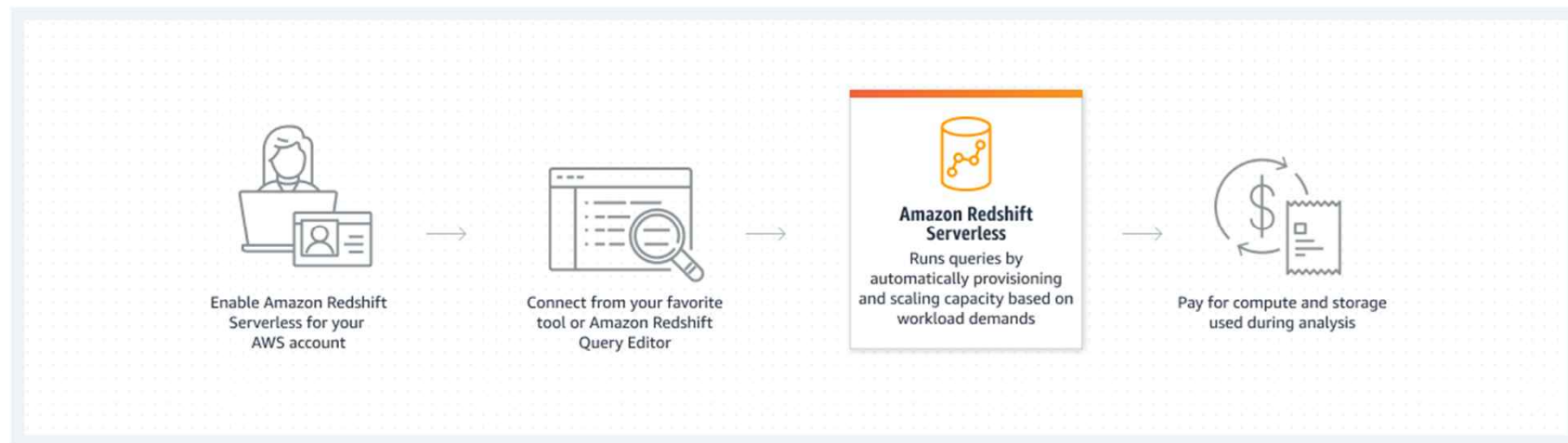
AWS Redshift

Amazon Redshift는 SQL을 사용하여 여러 Data Warehouse, 운영 데이터베이스 및 데이터 레이크에서 정형 데이터 및 반정형 데이터를 분석하고 AWS가 설계한 하드웨어 및 기계 학습을 사용해 어떤 규모에서든 최고의 가격 대비 성능을 지원한다



AWS Redshift

Amazon Redshift Serverless를 사용하면 데이터 웨어하우스 인프라를 관리하지 않고도 모든 규모의 분석 워크로드를 손쉽게 실행할 수 있다. 개발자, 데이터 사이언티스트, 분석가는 데이터베이스, 데이터 웨어하우스, 데이터 레이크에서 작업하여 보고서 및 대시보드 애플리케이션을 구축하고, 실시간 분석을 수행하고, 데이터를 공유 및 협업하며, 기계 학습 모델을 구축 및 훈련할 수 있다



AWS Redshift

AWS Redshift Serverless 쿼리



The screenshot displays the AWS Redshift Serverless console interface. At the top, the query title is "Sales per event". Below the title, there are tabs for "Run" and "Limit 100". The SQL query is as follows:

```
1 SET search_path to ticket;
2 SELECT eventname, total_price
3 FROM (SELECT eventid, total_price, ntile(1000) over(order by total_price desc) as percentile
4       FROM (SELECT eventid, sum(pricepaid) total_price
5             FROM ticket.sales
6             GROUP BY eventid)) Q, ticket.event E
7 WHERE Q.eventid = E.eventid
8 AND percentile = 1
9 ORDER BY total_price desc;
```

Below the query, there are tabs for "Result 1" and "Result 2 (9)". The "Result 2 (9)" tab is selected, showing a table with two columns: "eventname" and "total_price". The table contains 9 rows of data, sorted by total price in descending order. The data is as follows:

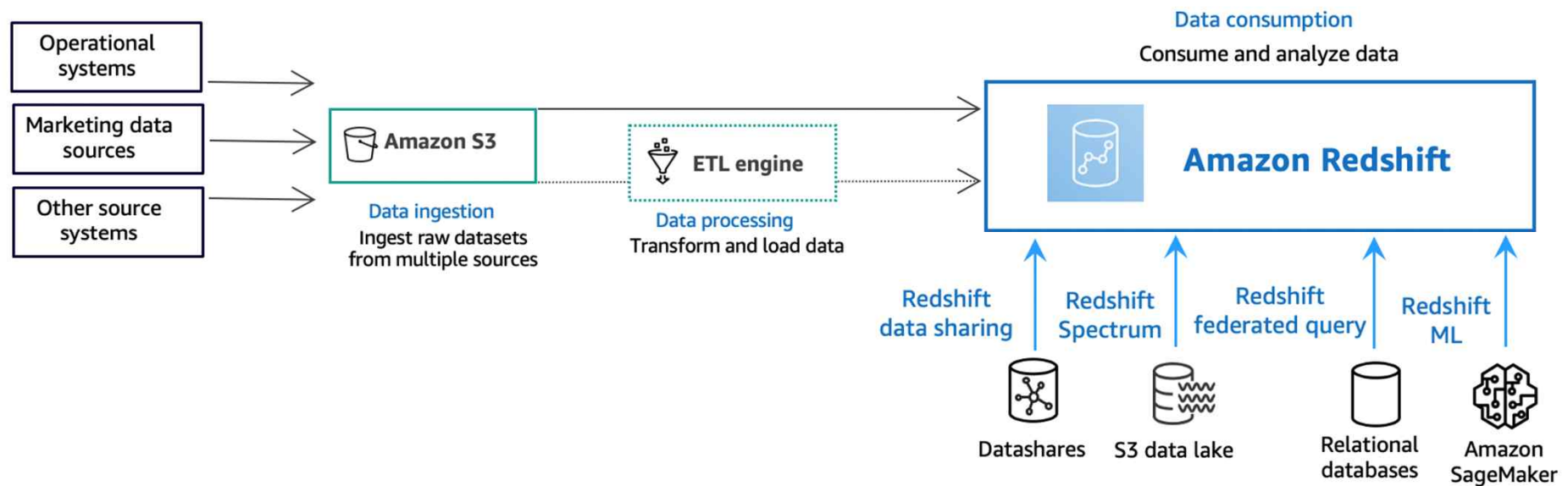
eventname	total_price
Adriana Lecouvreur	51846
Janet Jackson	51049
Phantom of the Opera	50301
The Little Mermaid	49956
Citizen Cope	49823
Sevendust	48020
Electra	47883
Mary Poppins	46780
Live	46661

At the bottom right of the console, the execution details are shown: "Elapsed time: 401 ms" and "Total rows: 9".

https://docs.aws.amazon.com/ko_kr/redshift/latest/gsg/new-user-serverless.html

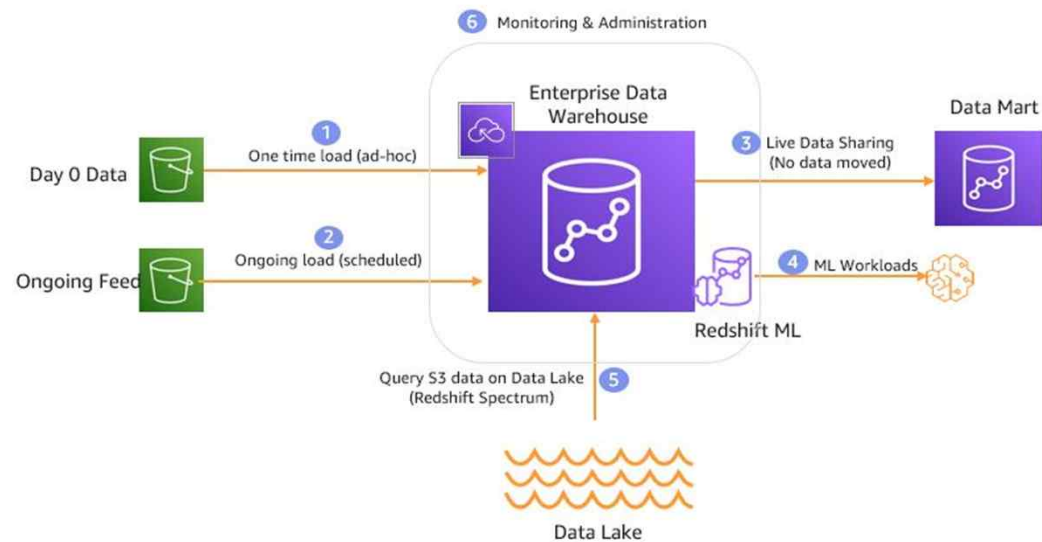
AWS Redshift

AWS Redshift ETL



AWS Redshift

- AWS Redshift Lab



Redshift Immersion Day Labs

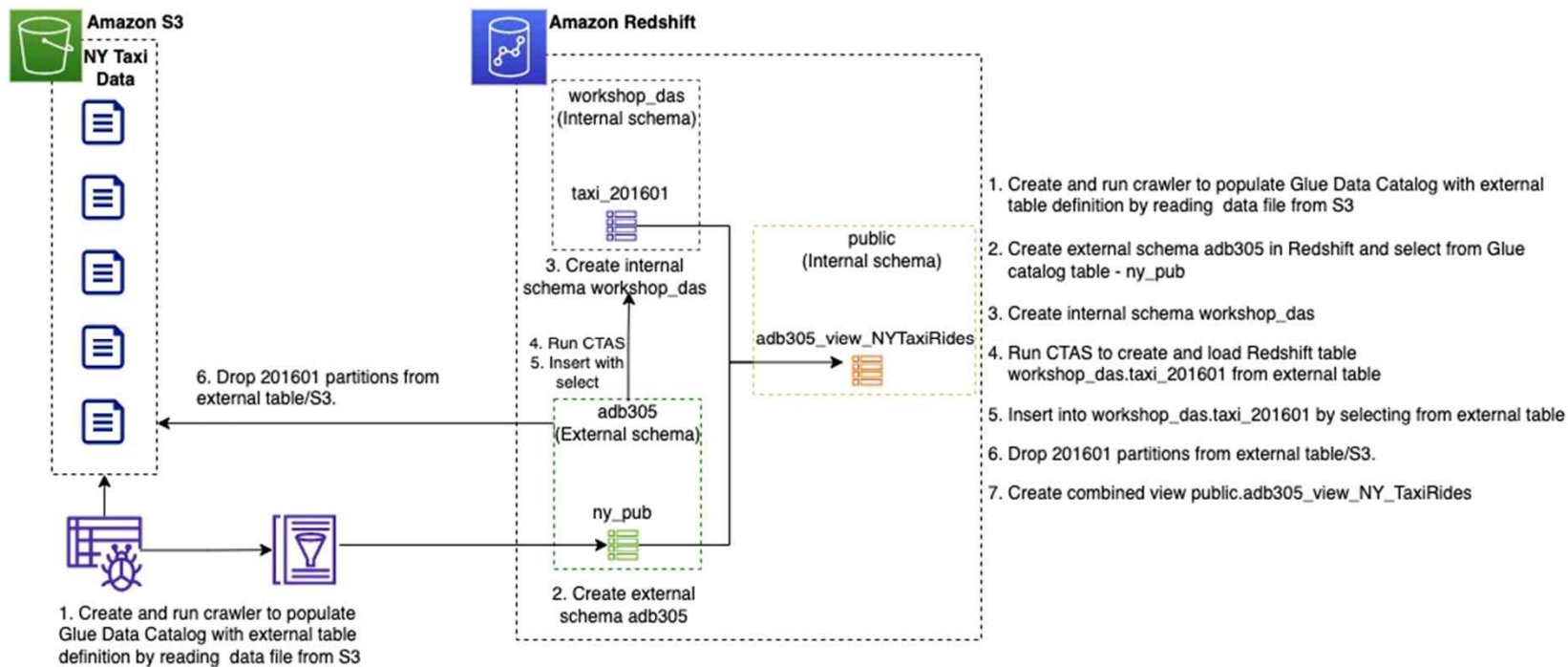
- 0 Creating Redshift Serverless Warehouse 
- 1 Table Creation & Manual Load  
- 2 Ongoing loads with stored procs 
- 3 Data Sharing with Data Mart  
- 4 Run ML workloads   
- 5 Query Data Lake (S3)  
- 6 Operations Lab 

 DBA
 Data Engineer
 BI User
 Data Scientist

<https://catalog.us-east-1.prod.workshops.aws/workshops/9f29cdba-66c0-445e-8cbb-28a092cb5ba7/en-US/lab1>

AWS Redshift

AWS Redshift 실습 사전 준비 : AWS Glue Crawler 사용



<https://catalog.us-east-1.prod.workshops.aws/workshops/9f29cdba-66c0-445e-8cbb-28a092cb5ba7/en-US/lab4>

AWS Redshift

AWS Redshift 실습

- 오레곤(us-west-2)리전으로 리전 변경하여 실습한다

- stack을 먼저 생성한다

<https://s3-us-west-2.amazonaws.com/redshift-immersionday-labs/immersionserverless.yaml>

The End