[4/1] AWS Builders Korea Program [기본 과정]

시간	제목	발표내용
9:00 – 12:00	**기본과정 :** **워크로드에 적합한 데이터베이스 선택하기**	워크로드에 적합한 데이터베이스 선택하기 현실에서 발생하는 데이터와 요구되는 비즈니스 형태는 다양합니다. AWS는 이를 위해 다수의 데이터베이스 서비스를 제공하고 있습니다. 이번 세션에서는 여러 상황별 선택의 기준을 제시해 적합한 데이터베이스를 선택하고 기본적 구성을 소개해 드립니다. 일부 데이터베이스에 대해 모니터링, 성능 및 이관 방법을 전달해 드립니다. 발표자: 박준용, AWS Solutions Architect
12:00 – 13:30	점심시간	
13:30 – 17:00	**기본과정 :** **DB보다 먼 빅데이터보다 가까운 Amazon Redshift 알아보기**	DB보다 먼 빅데이터보다 가까운 Amazon Redshift 알아보기 우리가 익숙하게 사용했던 SQL문을 빅데이터 분석에 활용합니다. Amazon Redshift 기반 DW 시스템을 구축하고, S3 데이터레이크를 활용한 빅데이터 플랫폼으로 확장할 수 있습니다. Redshift Spectrum을 활용하여 S3에 저장된 대규모 데이터를 DW 데이터와 통합하여 분석해 봅니다. 발표자: 우창식, AWS Solutions Architect

DB보다 먼 빅데이터보다 가까운 Amazon Redshift 알아보기

시간	내용
13:30 - 14:00	데이터 트렌드 / Amazon Redshift 아키텍처
14:00 - 14:50	실습 I 1. 클러스터 생성 및 연결 2. 데이터 로딩 3. 테이블 디자인 및 쿼리 튜닝
14:50 - 15:00	휴식시간 I
15:00 - 15:30	Amazon Redshift 확장성 및 현대화 데이터 아키텍처
15:30 - 16:20	실습 II 4. Spectrum을 활용한 현대화 5. Spectrum 쿼리 튜닝
16:20 - 16:30	휴식시간 II
16:30 - 17:00	Amazon Redshift 신규 기능 및 정리

AWS Builders Korea Program 200

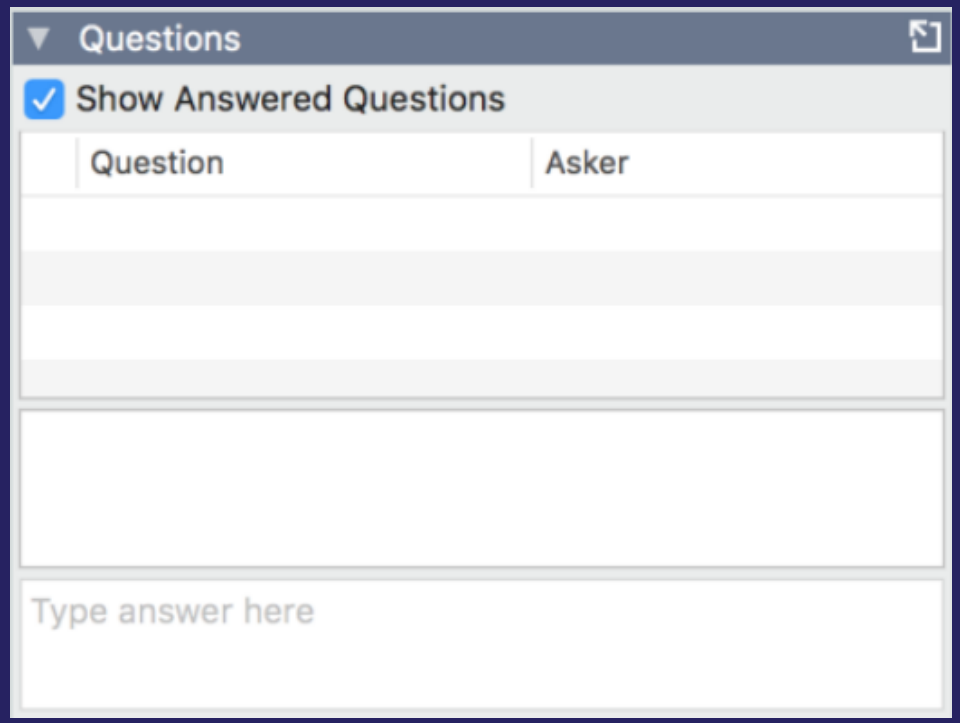
DB보다 먼 빅데이터보다 가까운 Redshift 이해하기

우창식

Partner Solutions Architect

강연 중 질문하는 방법

- AWS Builders Go to Webinar “Questions” 창에 자신이 질문한 내역이 표시됩니다. 기본적으로 모든 질문은 공개로 답변 됩니다만 본인만 답변을 받고 싶으면 (비공개)라고 하고 질문해 주시면 됩니다.



고지 사항 (Disclaimer)

본 컨텐츠는 고객의 편의를 위해 AWS 서비스 설명을 위해 온라인 세미나용으로 별도로 제작, 제공된 것입니다. 만약 AWS 사이트와 컨텐츠 상에서 차이나 불일치가 있을 경우, AWS 사이트(aws.amazon.com)가 우선합니다. 또한 AWS 사이트 상 에서 한글 번역문과 영어 원문에 차이나 불일치가 있을 경우(번역의 지체로 인한 경우 등 포함), 영어 원문이 우선합니다.

AWS는 본 컨텐츠에 포함되거나 컨텐츠를 통하여 고객에게 제공된 일체의 정보, 콘텐츠, 자료, 제품(소프트웨어 포함) 또는 서비스를 이용함으로 인하여 발생하는 여하한 종류의 손해에 대하여 어떠한 책임도 지지 아니하며, 이는 직접 손해, 간접 손해, 부수적 손해, 징벌적 손해 및 결과적 손해를 포함하되 이에 한정되지 아니합니다.

실습 시작 전 준비 사항

AWS 계정으로 시작

1. 실습 전 계정을 꼭 신청해주세요 : https://portal.aws.amazon.com/billing/signup#/start

2. AWS 계정이 없으신 경우, 행사 참여 전에 미리 AWS 계정 생성 가이드를 확인하시고 AWS 계정을 생성해 주시길 바랍니다.

   - *AWS 계정 생성 가이드: https://bit.ly/3wvn2Wj

3. 웨비나 종료 후 설문조사에 참여해주신 분들께는 실습 비용 지원을 위한 AWS 크레딧(1인당 $50 크레딧)을 추가로 지원드립니다. 해당 AWS 크레딧은 등록하신 이메일 계정으로 4월 중 발송 드릴 예정입니다.

4. 검증된 호환성을 위하여 실습 시 사용할 웹 브라우저는 Mozilla Firefox 또는 Google Chrome Browser로 진행 부탁드립니다.

Topics

- 데이터 트렌드
- Redshift 아키텍처
- Redshift 확장성
- Redshift 현대화 데이터 아키텍처
- 신규 기능

Bradley Todd
Liberty Mutual, Technology Architect

Redshift allows us to quickly spin up clusters and provide our data scientists with a fast and easy method to access data and generate insights

데이터 트렌드

Data is a strategic asset for every organization

"The world's most valuable resource is no longer oil, but data."*

*Copyright: The Economist, 2017, David Parkins

More data is created every *hour* today than in an entire year just 20 years ago

*Csource: IDC Report

데이터분석 제약사항

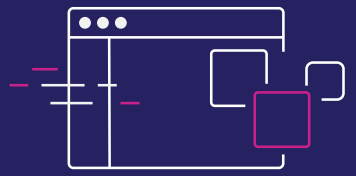
다양성

원천 데이터/종류의 다양함

다각도 분석 필요

데이터 크기 및 속도

성능



느린 성능

관리의 어려움

확장의 어려움

비용

예상치 못한 비용 증가

도구의 고착화

보안 이슈

Amazon Redshift

다양한 데이터 분석

AWS DW, Data Lake, DB를
통합하여 분석 가능

Modern Data Architecture

안정적인 성능 보장

일반적인 DW 대비
최대 3배 성능 보장

self-tuning & 10x with AQUA

비용 절감

필요한 만큼만 On-demand
RI로 선택 시 최대 75% 절감

50% less expensive

주요 고객 사례

AWS 고객은 Amazon Redshift를 활용하여 매일 엑사 단위의 데이터를 처리합니다.

NTT DOCOMO

Moved >10 PB of data from on-premises to cloud

WARNER BROS.

Performance, scale, cost-effeciency

Yelp

Enabling a data-driven organization with concurrency scaling

Jack in the Box

Improved ops by moving off of on-premises DW

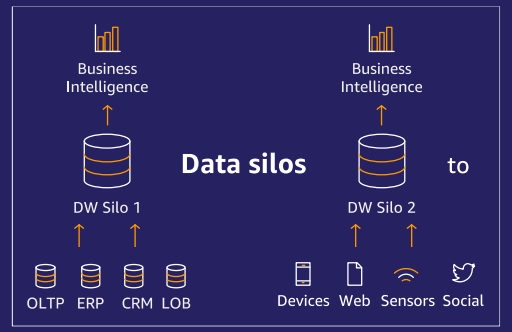
Pfizer

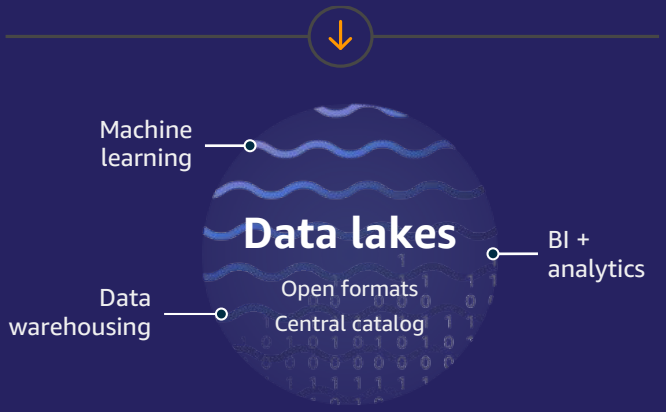
Provide scientists with near real-time analysis

기존 DW 아키텍처 제약사항

- 기존 온프레미스 기반 DW Dark Data 유발
- DW Silo로 인한 통합 분석 및 인사이트 도출 문제점 발생

확장성	• 필요시 쉽게 확장 불가 • HW 변경, 업그레이드 시 장기간 작업시간 필요
비용	• 높은 플랫폼 관리 비용 • Cold/Warm 데이터 보관에 따른 공간 낭비
사용성 제약	• 데이터 포맷 제약 • 데이터 Silo 발생 • 별도의 수집 및 변환(ETL) 아키텍처 구축 필요 • 사용자 기준 제약 발생
고전적인 구조	• 규격화된 사이즈 일반화로 인한 제약





AWS Analytics 포트폴리오

Data, visualization, engagement, & machine learning

NEW Data Exchange | QuickSight | Pinpoint | SageMaker | Comprehend | Lex | Polly | Rekognition | Translate

Analytics

+ many more

Redshift | EMR (Spark & Hadoop) | AWS Glue (Spark & Python) | Athena | Elasticsearch Service | Kinesis Data Analytics

Data lake infrastructure & management

S3/Glacier | Lake Formation | AWS Glue

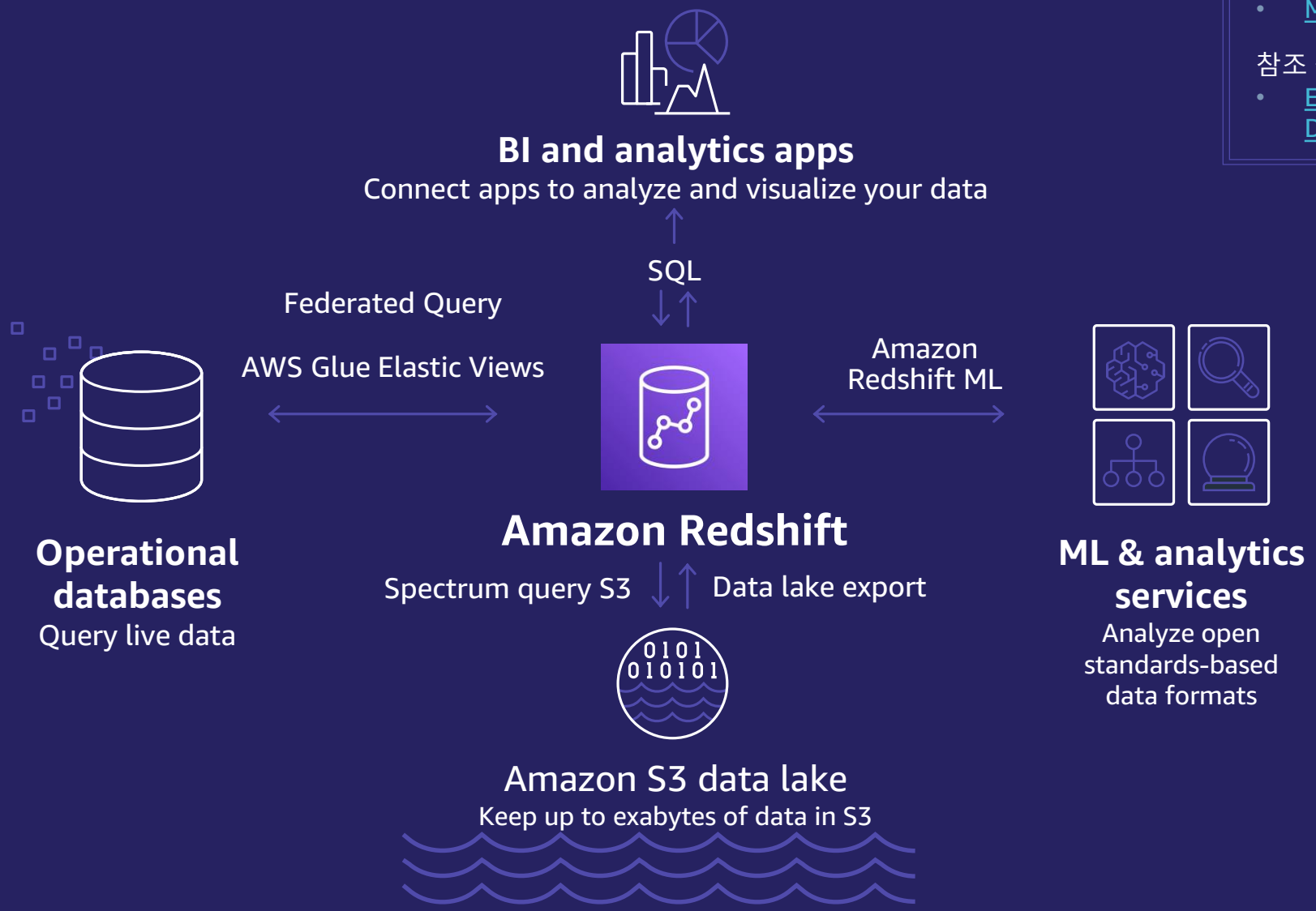
Data movement

Database Migration Service | Snowball | Snowmobile | Kinesis Data Streams | Kinesis Data Firehose | Managed Streaming for Apache Kafka

Redshift 현대적 데이터 아키텍처

참조 문서
- Modern Data Architecture

참조 블로그
- ETL and ELT design patterns for Modern Data Architecture



Redshift 아키텍처

Anil Chalasani

Gainsight, VP Product Operations

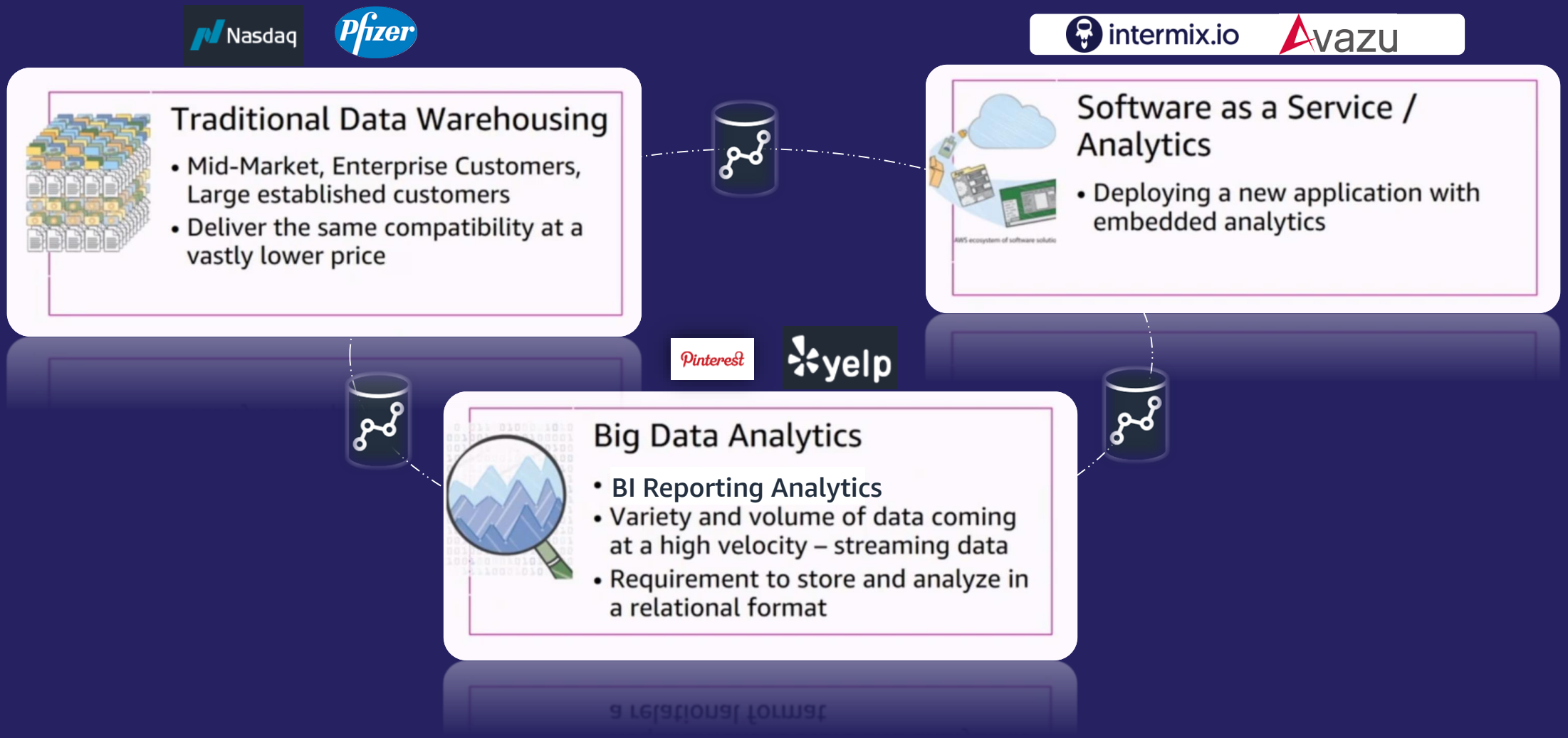
Using Redshift's DC2 node, we generate reports 35 percent faster. This enables our customers to spend more time curating and visualizing their data in Gainsight to take advantage of opportunities to drive customer success

Amazon Redshift 란?

단일 클러스터 DW 뿐만 아니라 데이터 Silo를 제거하여 폭넓은 데이터 활용 가능

완전 관리형	자동화된 운영 및 워크로드 관리 지원 클라우드 기반 비용최적화 DW
빠른 처리 성능	머신러닝 기반 최적화 기능
높은 안정성	SLA 99.9%
보안성	End-to-end 암호화 / SSO Compliance 제공
확장성	GB 부터 EB까지 확장 가능 Auto scaling, compute/storage scaling
높은 신뢰성	다양한 활용사례를 통해 입증 공식 기관을 통한 신뢰성 인정
RDS, ML, Data Lake 호환성	데이터레이크를 활용한 통합 분석 ML 워크로드 확장 가능
범용 쿼리/데이터 포멧 지원	다양한 형태의 오픈 포맷 지원 Parquet, CSV, ORC, Avro, JSON …

일반적인 활용 사례

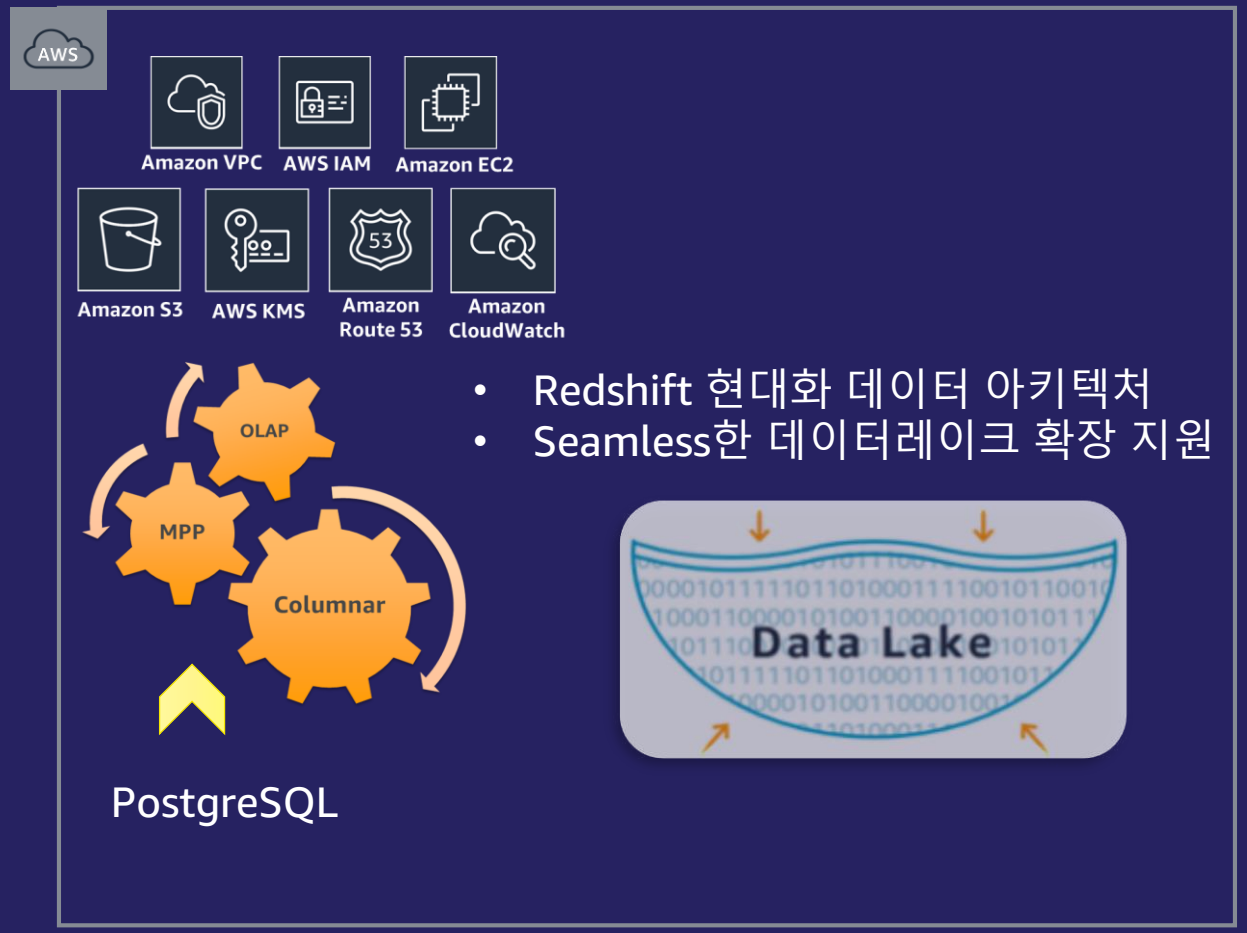


Redshift 기본원리

클라우드에 최적화된 MPP 기반 DW, Columnar 기반 OLAP DB

참조 문서

- 데이터 웨어하우스 시스템 아키텍처



- Redshift 현대화 데이터 아키텍처
- Seamless한 데이터레이크 확장 지원

Amazon Redshift

성능, 확장성, 비용효율성, 보안, 안정성을 갖추고 있으며 데이터레이크 확장을 제공

클라우드 DW 아키텍처 표준

Redshift 클러스터 아키텍처

Massively parallel, shared nothing architecture

Leader 노드

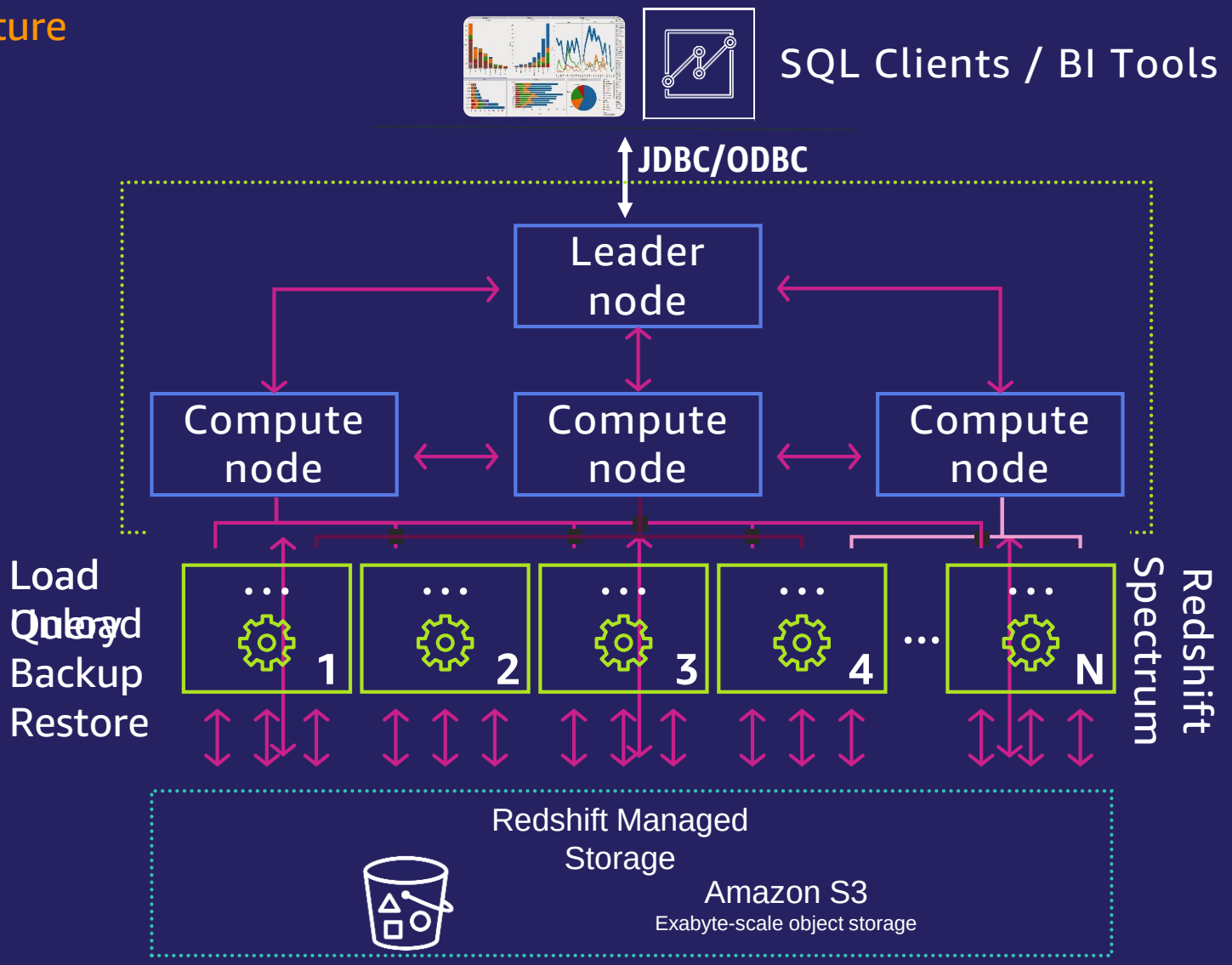
- SQL 엔드포인트
- 메타데이터 관리
- 병렬처리 구성 및 ML 최적화
- 2개 이상 노드 구성 시 비용 제외

Compute 노드

- 열기반 스토리지
- 병렬적으로 쿼리 수행
- S3 : Load, unload, backup, restore

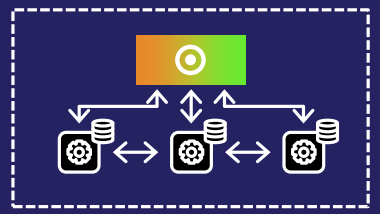
Amazon Redshift Spectrum 노드

- 데이터레이크을 활용한 쿼리 수행



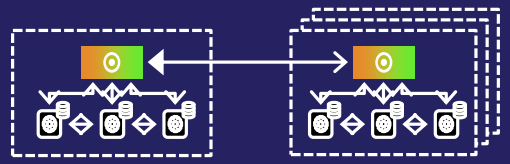
Redshift 아키텍처 발전과정

Redshift Spectrum for data lake analytics

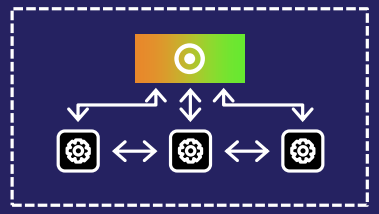


Spectrum

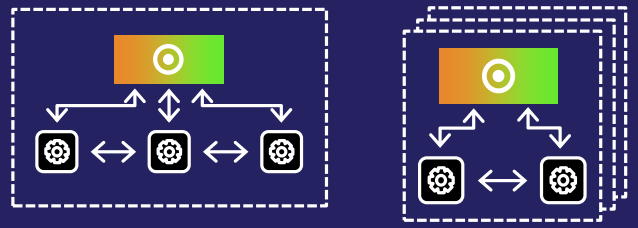
Concurrency Scaling for bursty workloads



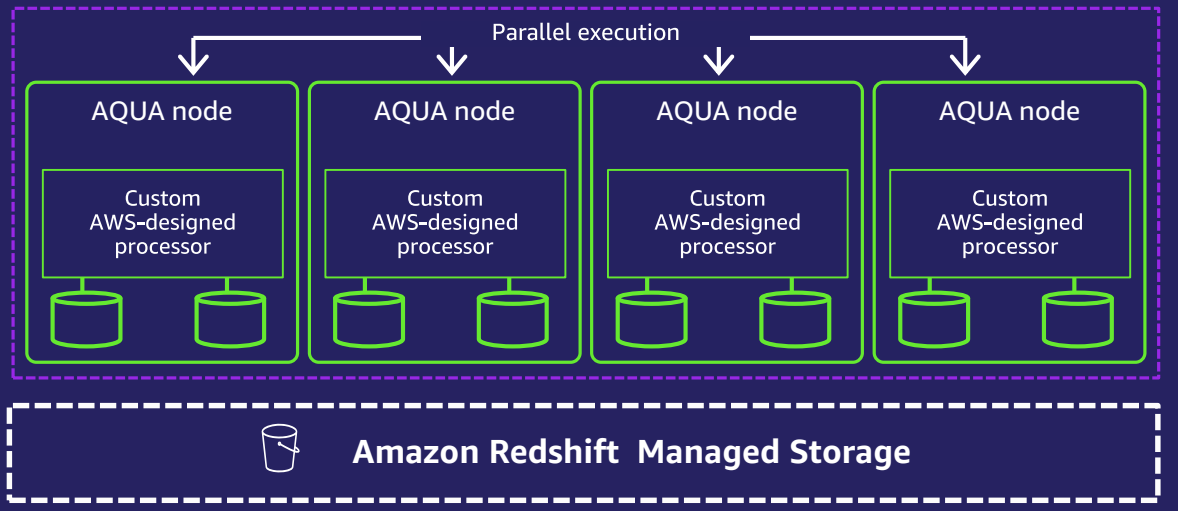
RA3 with independent compute and storage scaling



Data sharing across clusters



Query acceleration with computational cache

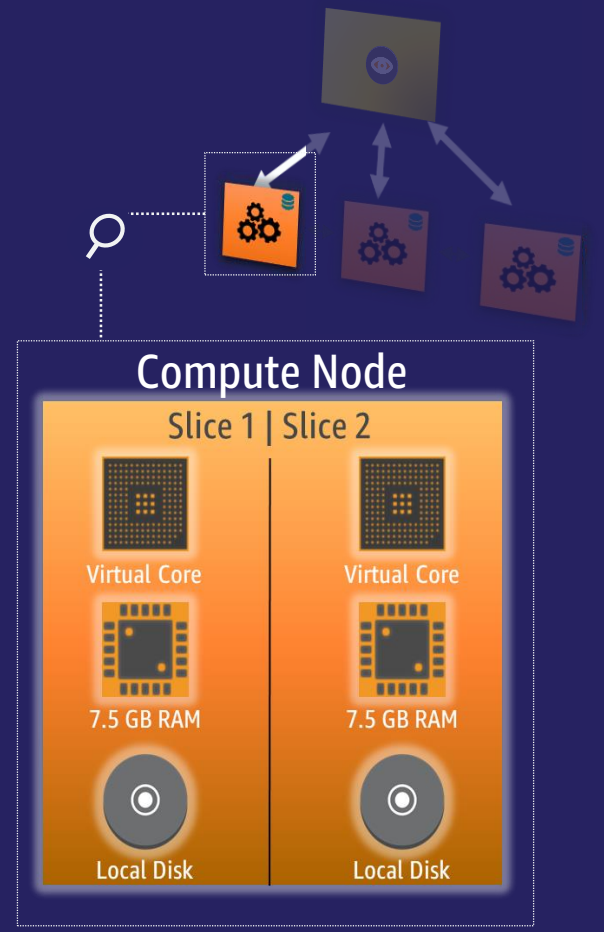


Compute 노드 | Slice

- Compute 노드는 다수의 Slice로 구성
- Slice는 가상적인 컴퓨팅 엔진 단위
- 각 Slice는 독립적인 메모리 및 디스크 공간 할당받아 Leader가 Compute 노드에 요청한 워크로드를 수행
- Leader 노드는 Slice 단위로 데이터를 분산하고 쿼리 및 DB 작업을 배정
- 병렬처리를 위한 SMP(Symmetric Multiprocessing) 메커니즘

참조 문서

- 데이터 웨어하우스 시스템 아키텍처



Redshift 인스턴스 타입

참조 문서

- Working with clusters

Amazon Redshift RA3

- Solid-state disks + Amazon S3
- Amazon Redshift Managed Storage (RMS)

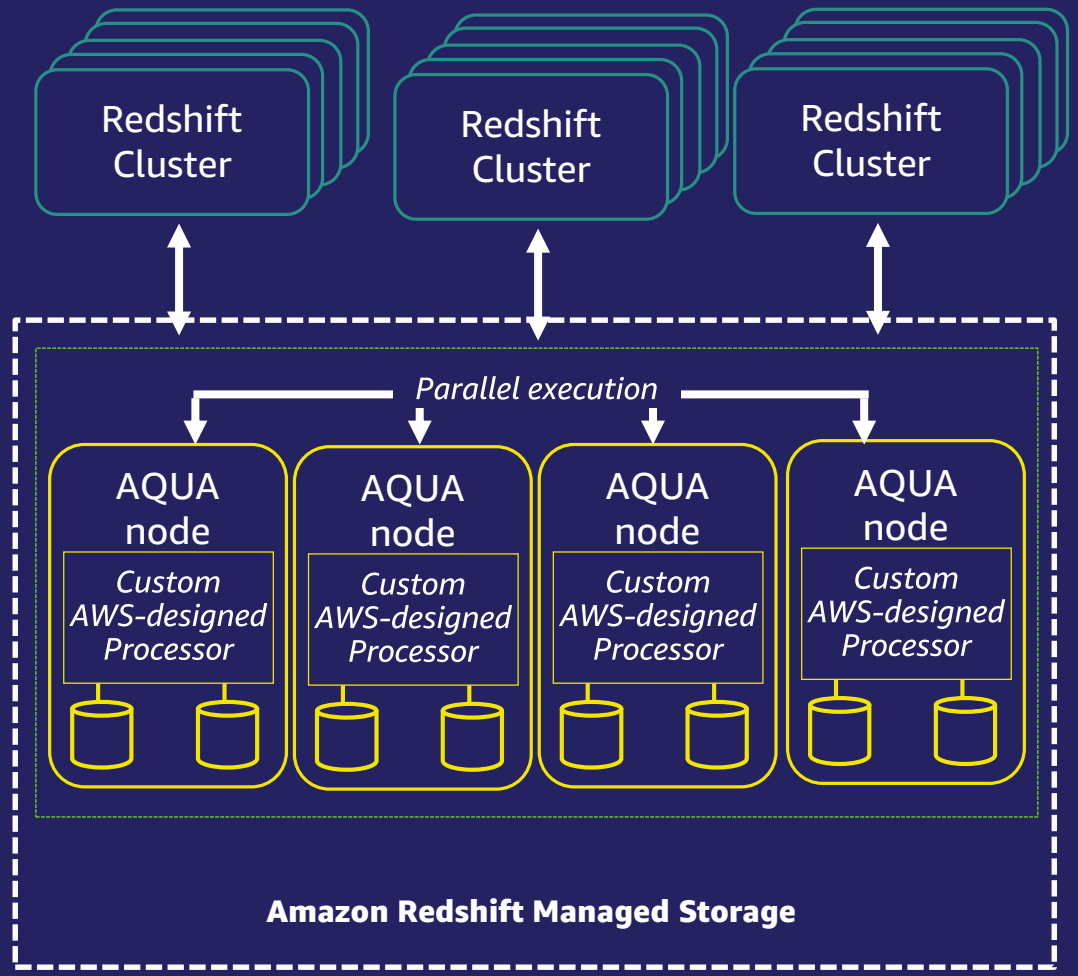
Dense compute - DC2

- Solid-state disks

Redshift 클러스터는 최대 128개의 노드 지원

	Instance type	Disk type	Size	Memory	# CPUs	# Slices
RA3 (New)	RA3 xlplus	RMS	Scales to 32 TB	32 GIB	4	2
	RA3 4xlarge	RMS	Scales to 128 TB	96 GIB	12	4
	RA3 16xlarge	RMS	Scales to 128 TB	384 GIB	48	16
Compute Optimized	DC2 large	SSD	160 GB	16 GIB	2	2
	DC2 8xlarge	SSD	2.56 TB	244 GIB	32	16

AQUA (Advanced Query Accelerator)



참조 블로그

- AQUA Blog

하드웨어 가속화 캐시를 이용한 분산 프로세스

AQUA를 활용하여 추가금없이 타 클라우드 DW대비 10배 빠른 성능 제공

AQUA 노드는 자체 제작된 AWS 분석 프로세서를 탑제하여 기존 CPU 대비 최적의 압축, 암호화, 집계 성능 제공

Redshift RA3 활용하여 변경없이 사용 가능

Redshift 보안

End-to-end data encryption

IAM integration & integration with SAML IdP's for Federation (SSO)

Amazon VPC for network isolation

Database security model (users, groups, privileges)

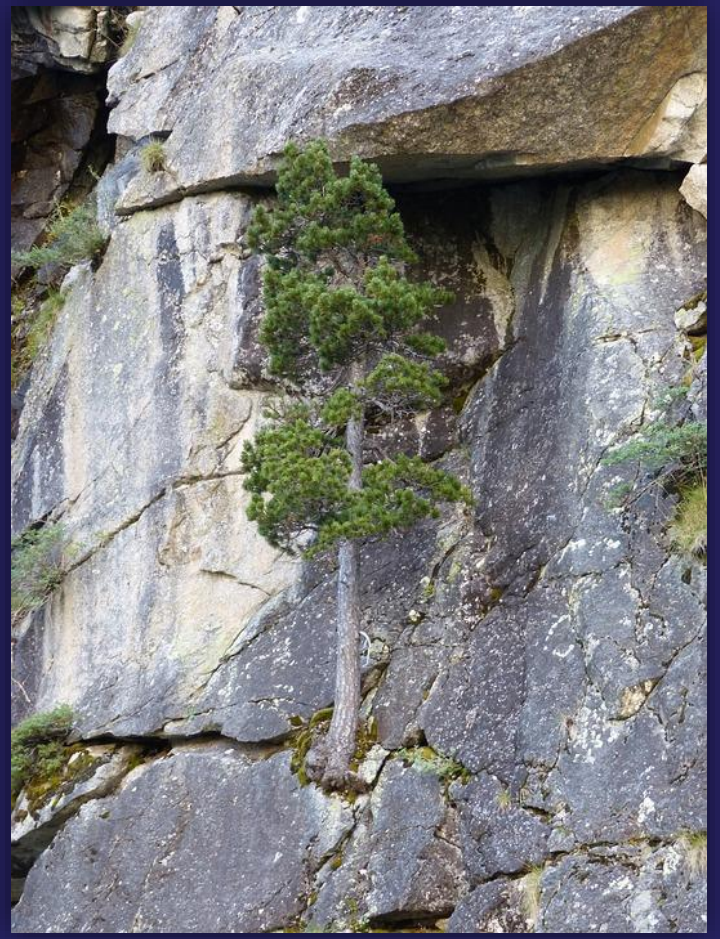
Audit logging and notifications

Certifications that include SOC 1/2/3, PCI-DSS, FedRAMP, & HIPAA

복원성

- **99.9% SLA**
- 2nd 노드로 데이터 자동 복제 기능
- 디스크 및 노드 장애에 대한 자동 감지 및 복구
- 자동 데이터 백업
- 백업 데이터를 다른 리전으로 복제

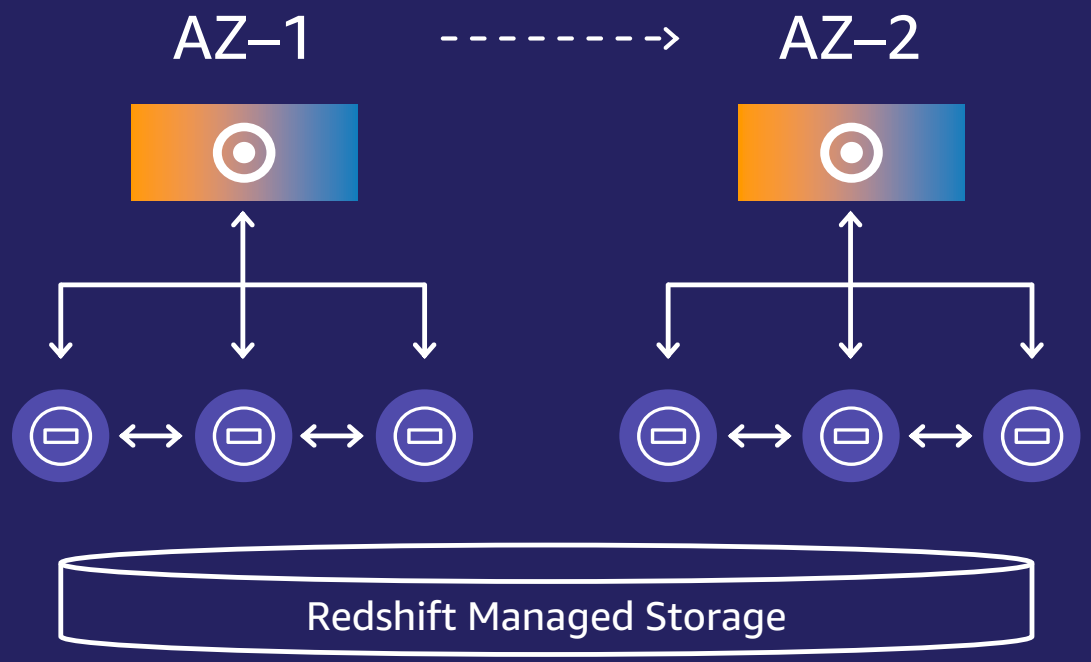
New RA3 활용 시 다른 가용영역으로 이전 가능(AZ Level Failure)



Cross-AZ 클러스터 복구

클러스터 장애 시 이중화된 가용영역(AZ)으로 서비스 이전

- 데이터 손실없이 복구 가능 (RP = Zero)
- 스냅샵을 통한 복구가 필요하지 않음
- On-demand failover
- 필요 시 다른 AZ에 클러스터 구성 가능
- RA3 인스턴스 패밀리 지원



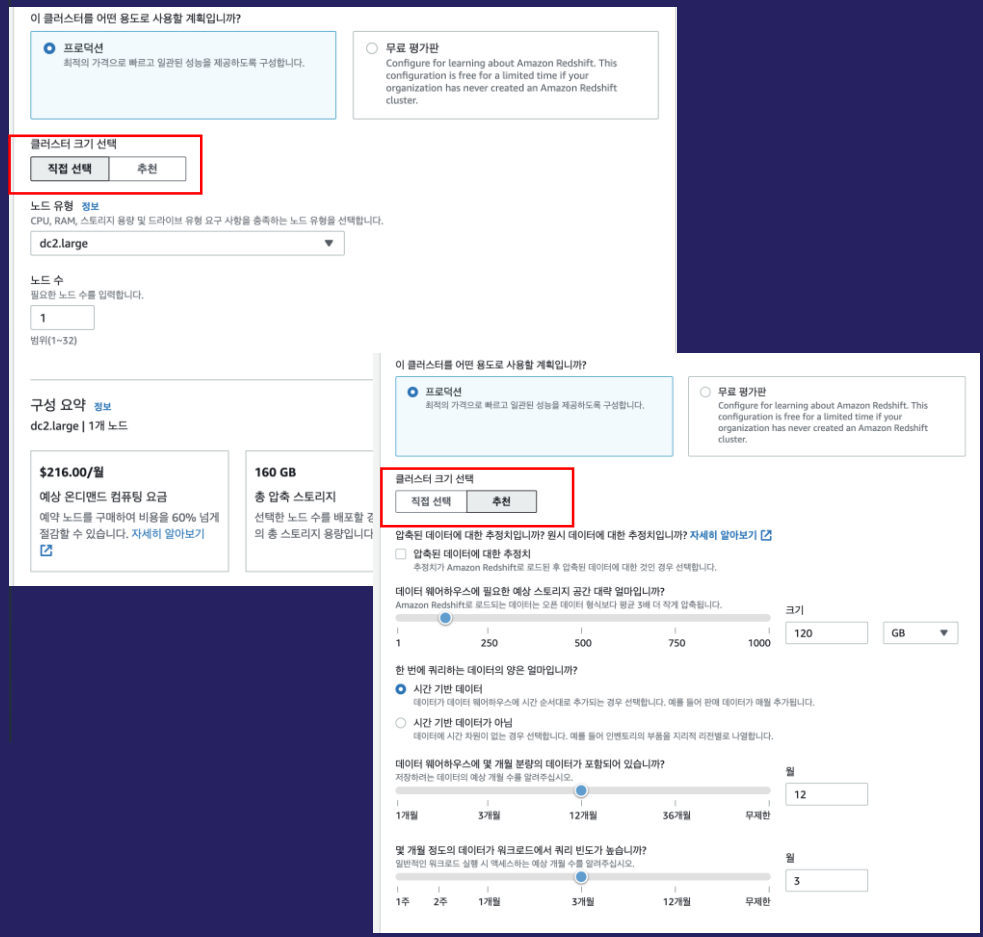
Redshift 클러스터 생성

VPC와 IAM role이 이미 구성되어 있다면, 클릭 몇 번만으로 클러스터 구축이 가능합니다.

1. AWS 콘솔에서 Redshift 선택

2. 데이터셋의 크기에 따라 인스턴스 종료,
   노드 수, IAM role 등을 지정

3. “클러스터 생성” 클릭



참조 문서

- 샘플 Amazon Redshift 클러스터 생성

Redshift 확장성

> Martin Brambley
> Sirocco Systems, Director
>
> We saw an immediate 30 percent improvement in end-to-end ETL loading using the new DC2 node from Redshift. This is fantastic news for our clients as data volumes and demand for analytics continue to grow rapidly

자동 **WLM(WorkLoad Management)**

Adaptive concurrency
- 워크로드 수행 시간을 기반으로 동시성 레벨 적응

Smarter preemption
- 높은 우선순위 쿼리가 낮은 순위의 쿼리를 선점할 수 있도록 제한
- 남은 수행시간을 고려하여 선점 조건 판단

Improved ML prediction model
- 성능 저하없이 시스템의 자원을 최대한 활용하여 더 많은 쿼리를 수행

Turbo-boost mode / SQA(Short Query Acceleration)
- 쿼리가 감지되었을때 대기 쿼리가 많은 자원을 필요로 하지 않을 경우

Concurrency Scaling
- 중요 워크로드를 SLA 충족시키기 위하여 우선순위 및 비용 기반의 동시성 확장

QMR(Query Monitoring Rule)기반
다양한 메트릭 데이터를 활용하여
최적화 지원

Eg. Query CPU Time, Query
Execution Time, Query Queue Time
…

Amazon Redshift 자동 WLM은
CLI나 SQL 명령어를 통하여
비즈니스 변화에 맞게 적용 가능

동시성 확장

예상치 못한 요청 증가에도 유연하게 동시성 확장 가능

대기상태 쿼리가 Compute 자원이 필요할 경우

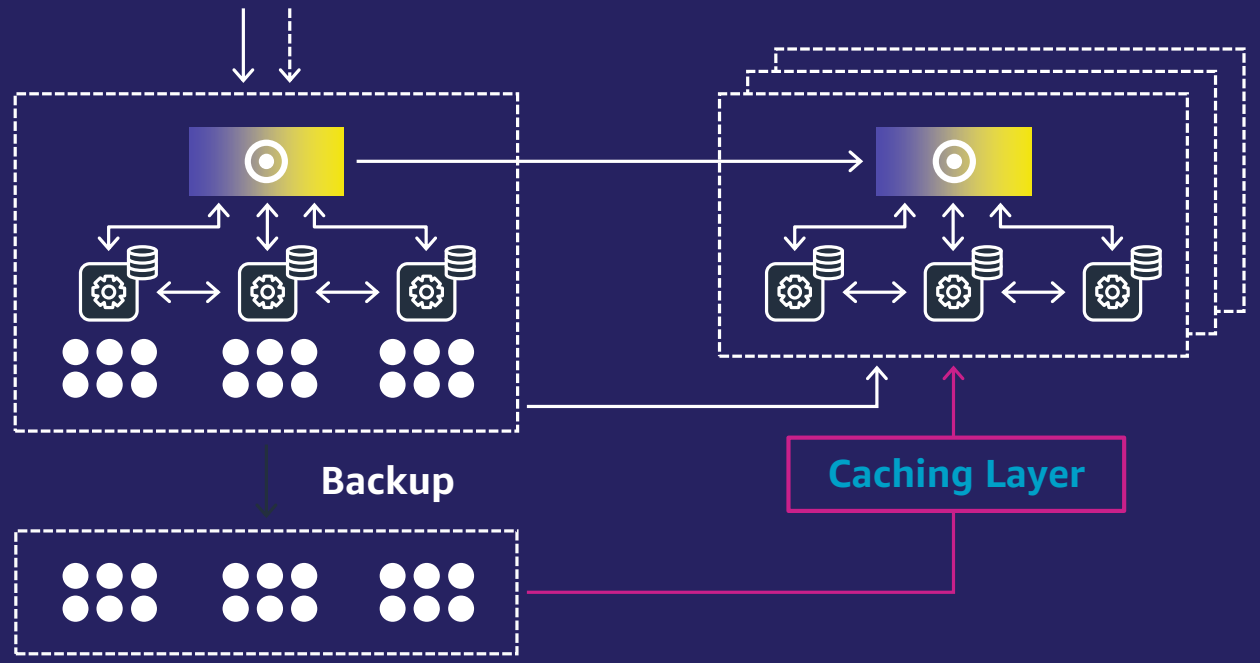
단시간(몇 초 이내)에 다중 클러스터로 확장

논리적으로 제한없이 동시 사용자 지원(SLA)

추가로 사용된 클러스터 자원은 초당 과금

매일 1시간 무료 제공

쓰기 워크로드도 동시성 확장 가능(일부 리전)



참조 문서

- Concurrency scaling

관련 블로그

- Manage and control cost for concurrency scaling

Redshift 클러스터 크기 조정

탄력적 크기 조정

- 운영중인 클러스터 노드 추가 및 삭제
- Scale Out 확장 지원
- 몇 분 이내의 빠른 전환
- 백그라운드 Slice 재분배

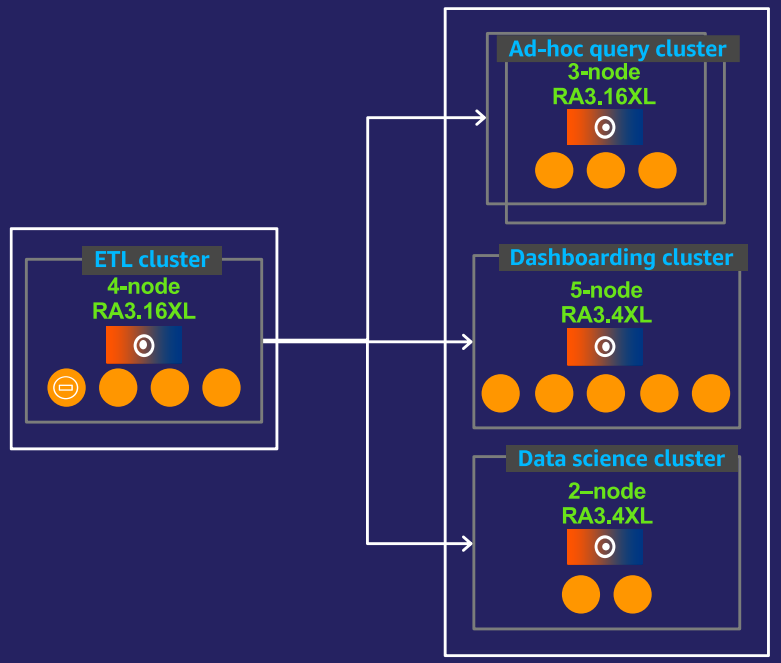
클래식 크기 조정

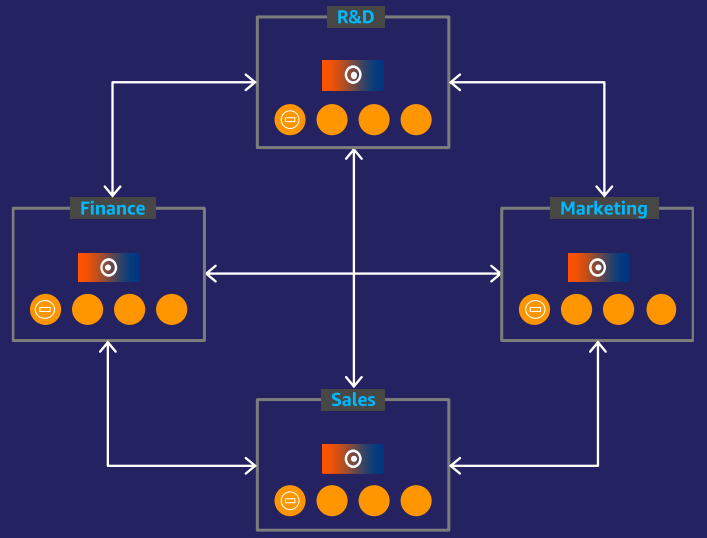
- 신규 클러스터 생성 후 원본 클러스터에서 데이터 복사
- 원본 클러스터의 데이터 크기에 따른 조정 시간 필요
- 신규 클러스터 기반으로 Slice 재조정

Redshift 콘솔에서 수행

데이터 공유

Redshift 클러스터 간에 안전하고 쉬운 공유 가능





- 데이터의 이동 및 복사없이 활용 가능
- 일관성있는 라이브 데이터 활용
- 내부/외부 조직에 안전한 협업 구조 생성

- 공유 데이터를 접근하는 워크로드를 분리 가능
- 단위 그룹이 협업 및 데이터 공유 시 활용
- 다른 AWS Analytics 서비스와 연계 가능

yelp

"Data sharing feature seamlessly allows multiple Amazon Redshift clusters to query data located in our RA3 clusters and their managed storage. This eliminates our concerns with delays in making data available for our teams, reduces the amount of data duplication and associated backfill headache. We now can concentrate even more of our time making use of our data in Amazon Redshift and enable better collaboration instead of data orchestration."

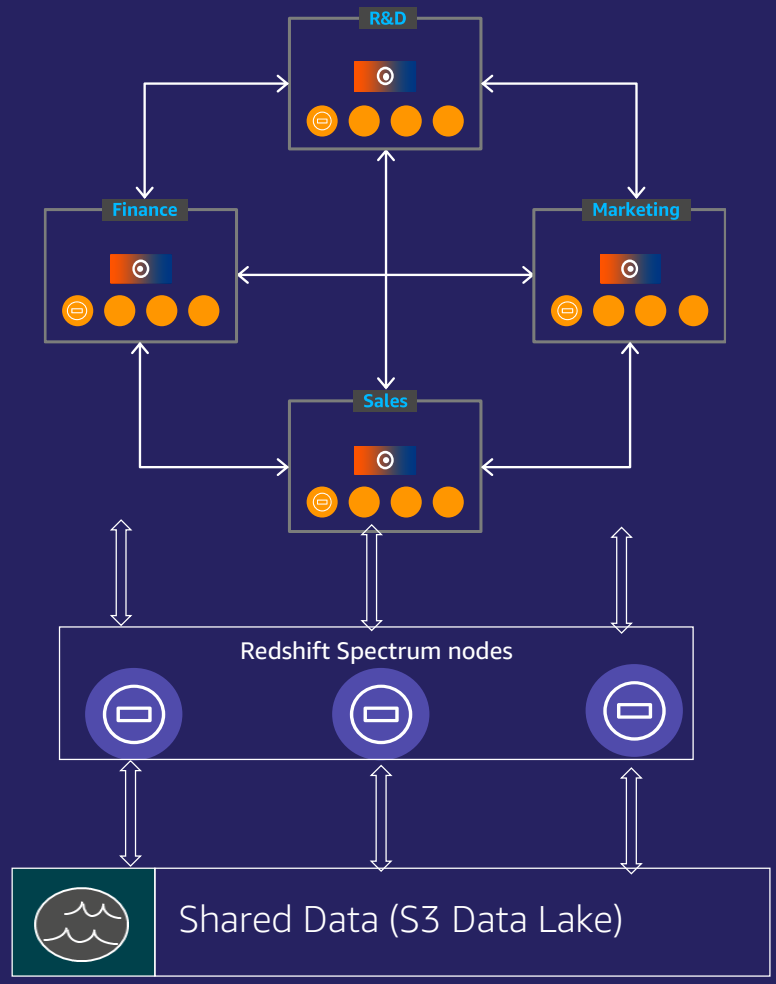
Steven Moy, Yelp

aws

S3 데이터 레이크를 활용한 다중 클러스터 접근

Advanced

- 목적에 따라 비용, 성능, SLA을 구분하여 다중 클러스터 구성
- 데이터 특성에 따른 클러스터 분리
- 사용자 그룹에 따른 비용 모델 적용
- 데이터 레이크의 데이터를 다중 클러스터가 공유해서 사용



Redshift 현대화 데이터 아키텍처

> **Hyung-Joon Kim**
> **BrandVerity, Principle Software Engineer**
>
> **We use Redshift Spectrum for interactive online queries...we can analyze far more data for our customers and deliver results much faster**

Redshift 현대적 데이터 아키텍처

현대적 데이터 아키텍처 특징

AI / ML 내재화

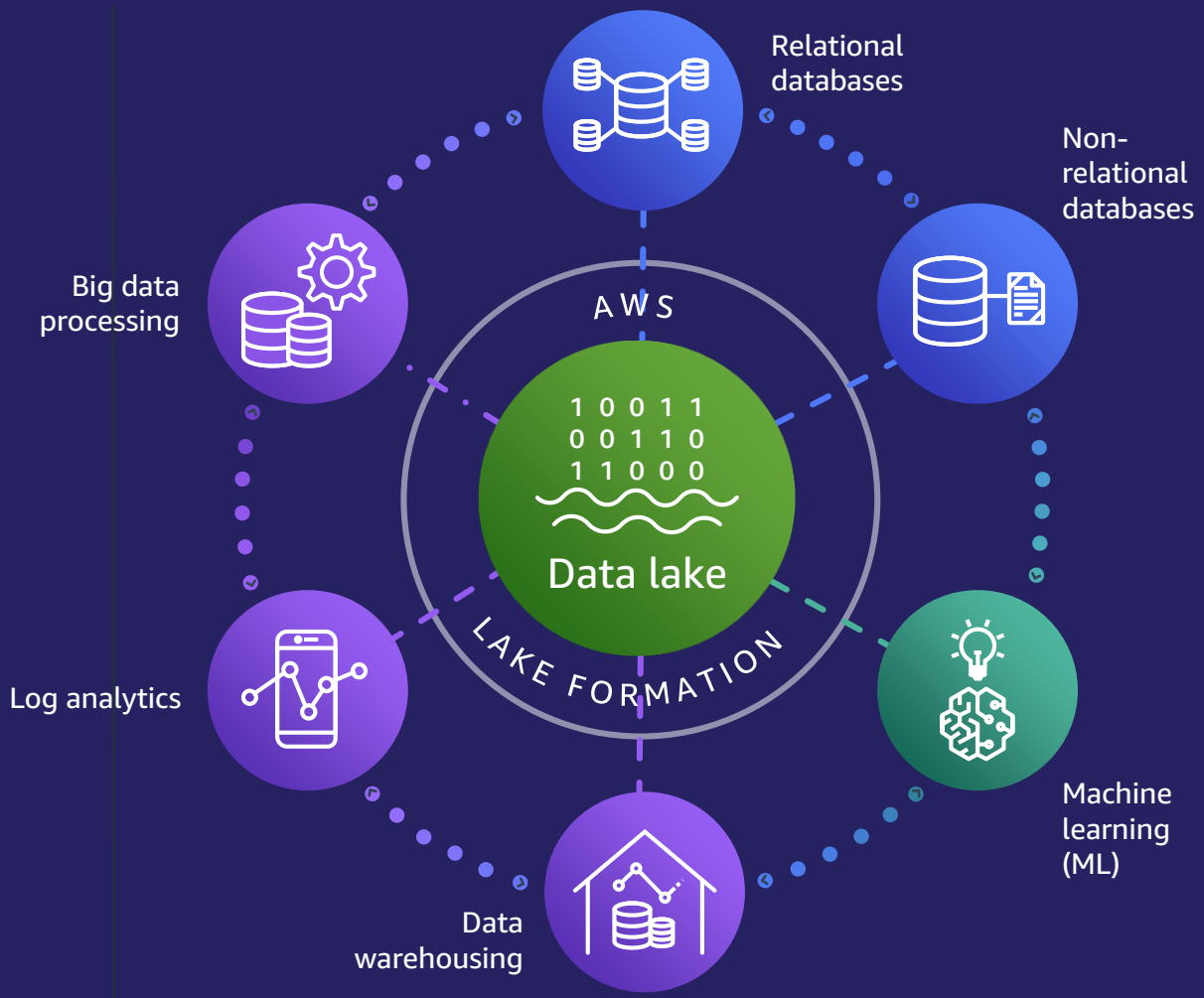
목적에 부합하는 DB 및 분석 서비스

통합된 데이터 접근, 보안, 통제

확장 가능한 데이터 레이크

저비용 고효율 아키텍처 구성

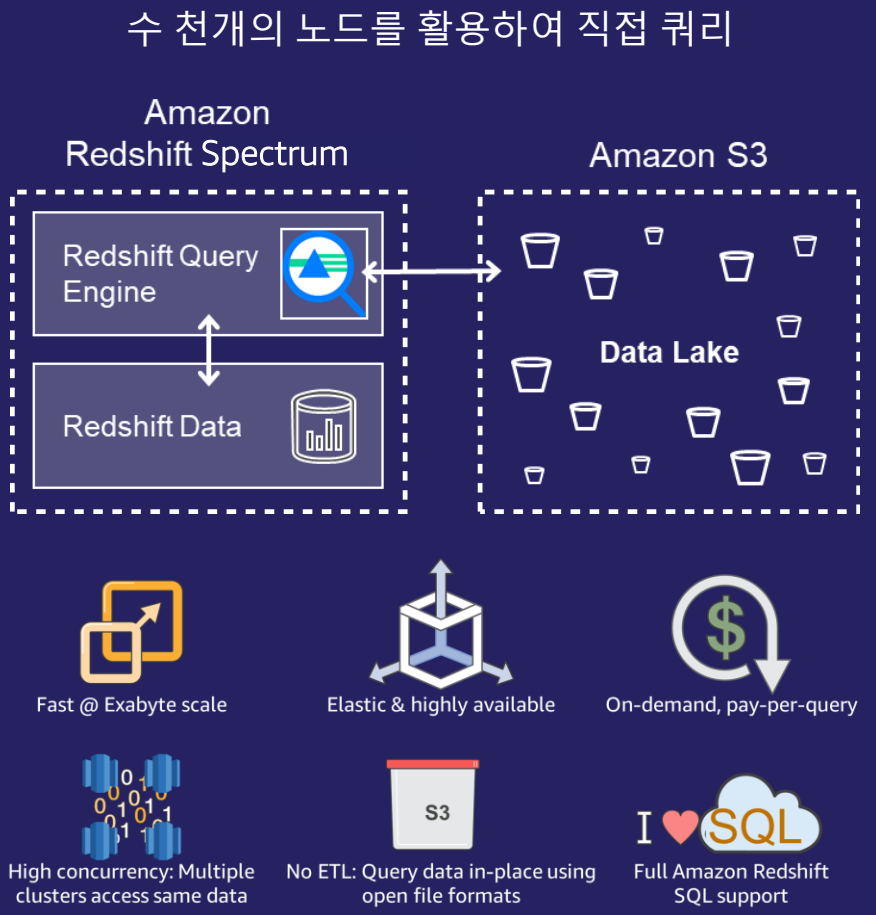
- 데이터 이동/변경없이 DW 환경에서 데이터 레이크 쿼리
- Warm 데이터는 Redshift, Cold 데이터는 S3
- 정형, 비정형 데이터 통합 분석 지원



Redshift Spectrum 개요

S3 데이터를 테이블에 적재하지 않고도 정형 및 비정형 데이터를 효율적으로 쿼리

- S3 데이터 레이크를 활용하여 엑사바이트 데이터 처리
- 데이터 로딩 없이 즉각적인 쿼리 지원
- 데이터 생애주기 티어링(warm/cold) 구축 가능
- AWS Glue, Amazon Athena와 호환되는 데이터 카탈로그를 사용하여 외부 테이블 생성 및 관리
- Amazon Redshift Spectrum 전용 노드 활용
- S3를 활용한 Materialized View 적용 가능



Redshift Spectrum 유연한 확장

Advanced

- Redshift Spectrum를 활용하여 기존 SQL과 BI 애플리케이션에 적용 가능
- 외부 테이블을 활용하여 Redshift 쿼리로 S3에 직접 쓰기 가능
- Join, nested 쿼리, 윈도우 함수 지원
- Date, Time, 지정키를 활용한 데이터 파티셔닝 지원
- Amazon Glue 데이터 카탈로그 활용 가능

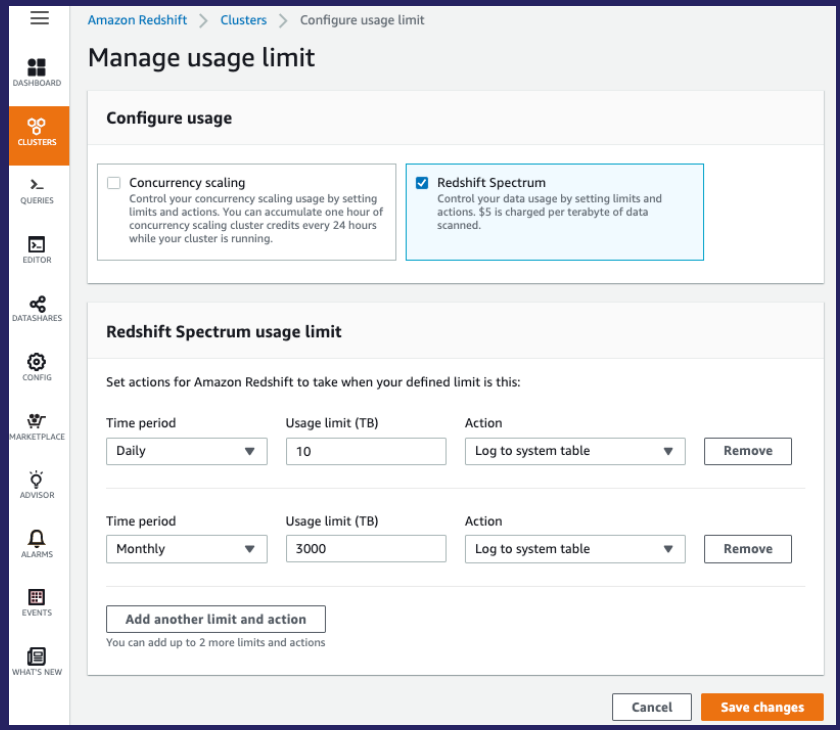
Redshift Spectrum 비용 효율성

Advanced

- 간소화된 가격 정책
  - $5 per TB from S3
- Redshift 콘솔에서 Spectrum 비용 모니터링 및 예산 통제
- 쿼리 당 수 천개의 Spectrum 노드 사용 가능
- 쿼리 비용 최적화 가능
  - 파티셔닝
  - Columnar 데이터 포맷
  - 압축

Redshift spectrum cost controls

Redshift spectrum costs are pro-rated based on the amount of S3 data scanned, so if a query scans 10 GB of S3 data, it costs $0.05. If a spectrum query scans 1 TB of S3 data, it costs $5, irrespective of the compute requirements associated with either of those queries



외부 스키마 및 테이블 생성

1. AWS Glue Data Catalog 또는 Apache Hive Metastore를 활용하여 외부 스키마 정의

```
CREATE EXTERNAL SCHEMA <schema_name>
```

2. Athena, Hive Metastore Client, Redshift [Create External Table] 구문을 활용하여 외부테이블 등록

```
CREATE EXTERNAL TABLE <table_name>
[PARTITIONED BY <column_name, data_type, …>]
STORED AS file_format
LOCATION s3_location
[TABLE PROPERTIES property_name=property_value, …];
```

3. 외부 스키마, 테이블을 활용한 쿼리

```
<schema_name>.<table_name>
```

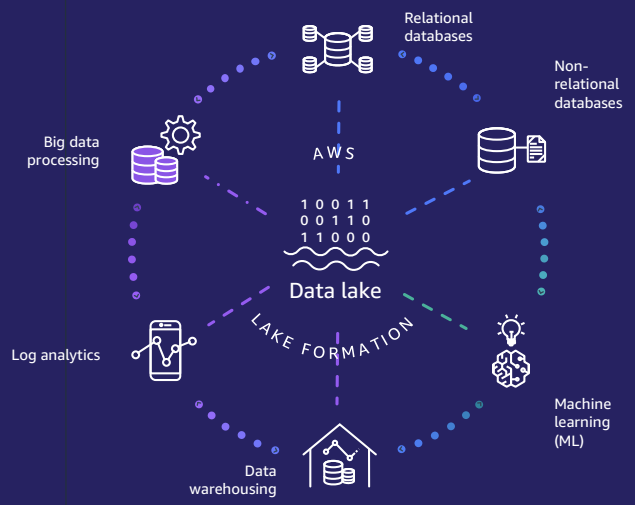
예시

```
create external schema hive_schema
from hive metastore
database 'hive_db'
uri '172.10.10.10' port 99
iam_role 'arn:aws:iam::123456789012:role/MySpectrumRole';
```

```
create external table lakehouse.sales(
salesid integer,
listed integer,
saledate date,
qtysold smallint,
pricepaid decimal(8,2),
saletime timestamp)
row format delimited
fields terminated by '\t'
stored as textfile
location 's3://sampledbusw2/tickit/lakehouse/sales/'
table properties ('numRows'='170000');
```

```
select lakehouse.sales_event.salesmonth, event.eventname,
sum(lakehouse.sales_event.pricepaid) from
lakehouse.sales_event, event where
lakehouse.sales_event.eventid = event.eventid and
salesmonth = '2008-02' and (event = '101' or event = '102')
group by event.eventname, lakehouse.sales_event.salesmonth
order by 3 desc;
```

데이터 레이크 서비스 선택



- 데이터 웨어하우스, 관계형, 복잡한 조인
- 낮은 Latency 필요
- 기존 DW와 S3 통합 분석 필요

Amazon Redshift

- 직관적인 Ad-hoc 쿼리 수행
- 서버리스
- 로그 분석

Amazon Athena

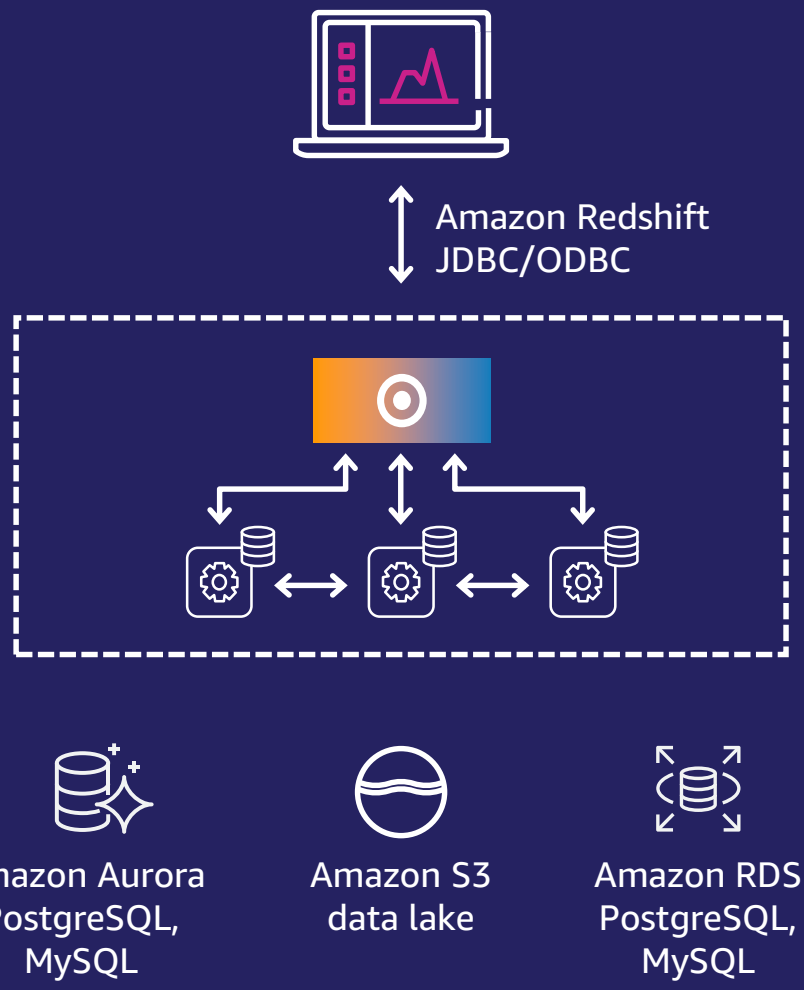
- 대규모 데이터 처리
- Hadoop Eco 기반 통합 분석
- Jupyter 기반 EMR 노트북 필요

Amazon EMR

엽합 쿼리(Federated Query) 활용

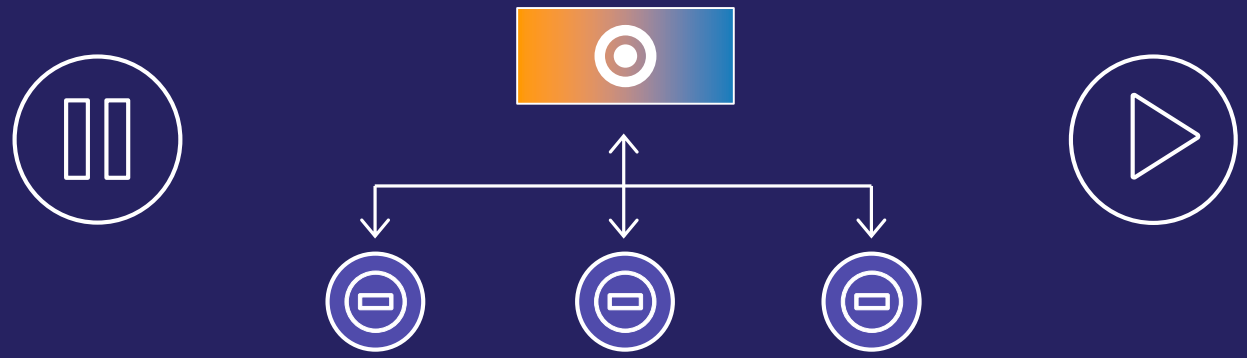
데이터베이스, 데이터 레이크 , DW 통합 분석

- 운영 데이터베이스와 실시간 데이터 통합 분석
- ETL 처리로 인한 지연없이 분석 작업 수행
- 복잡한 ETL 작업 없이 데이터 수집 가능
- Amazon RDS, Aurora PostgreSQL/MySQL 지원
- 성능 최적화를 위하여 원격지 데이터를 분산 처리 실행



클러스터 일시 중지, 다시 시작

비용 절감을 위해 클러스터 중지 및 시작 가능



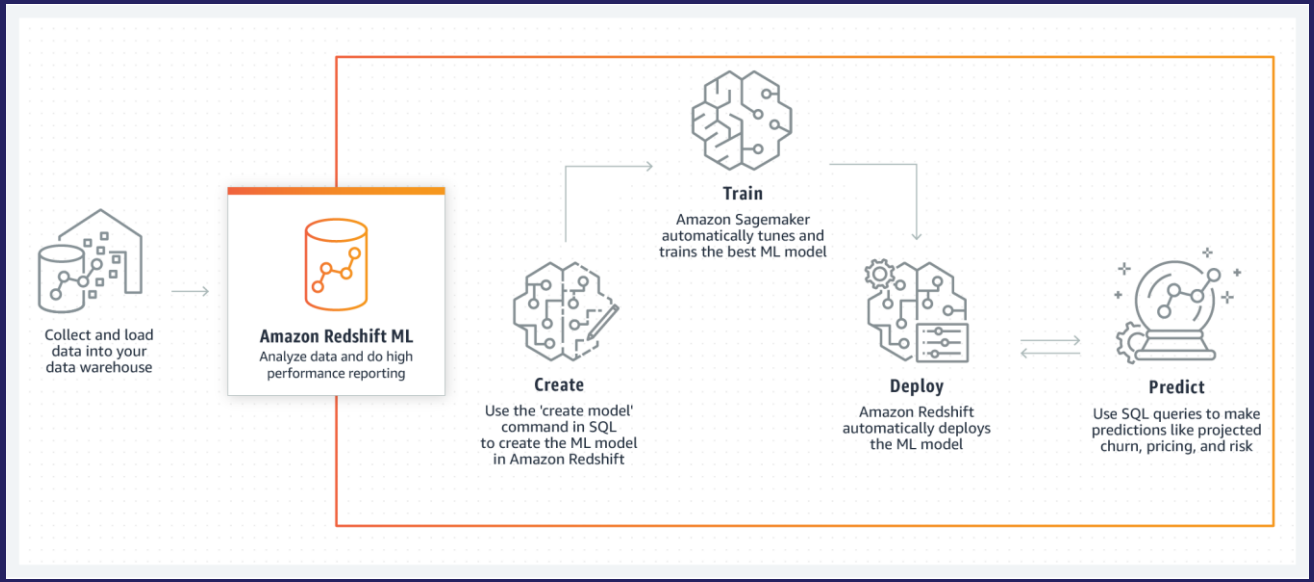
- 콘솔 또는 API를 활용하여 클러스터 일시중지, 다시시작 실행
- 사용자 정의 스케줄러를 활용하여 조정 가능
- 업무 시간 외 개발 워크로드 중지
- ETL 워크로드를 위한 클러스터의 경우 사용 후 중지 가능
- 예약 인스턴스(RI)를 다수의 클러스터가 사용 시간을 상이에게 적용하여 활용

신규 기능

Amazon Redshift ML

SQL을 활용한 머신러닝 모델 생성 및 학습

- SQL을 사용하여 ML 모델 생성, 학습, 배포
- 적합한 머신러닝 알고리즘 자동 선택
- BYOM(Bring Your Own Model) 지원
- 모델에 적합한 컬럼 선택 및 전처리 지원
- 활용사례: 상품추천, 사기예방, 고객이탈방지




```
CREATE MODEL demo_ml.customer_churn
FROM (SELECT c.age, c.zip, c.monthly_spend,
c.monthly_cases, c.active FROM
customer_info_table c)
TARGET c.active;
```

스칼라 Lambda UDF

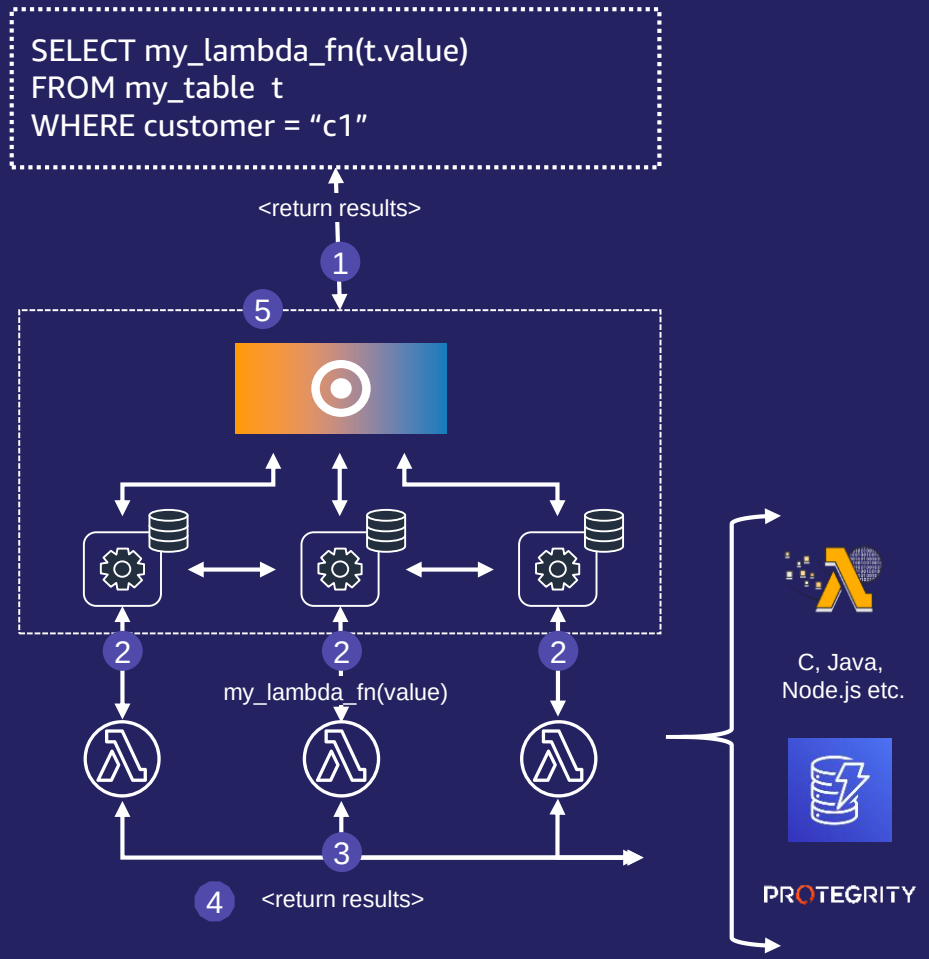
AWS Lambda를 활용한 사용자 정의 함수 생성

참조 문서

- Creating a scalar Lambda UDF

관련 블로그

- Data Tokenization with Amazon Redshift and Protegrity



SQL 쿼리 수행을 통한 AWS Lambda 프로그램 실행

외부 서비스 확장

- C++, Java, Go, Python 등 다양한 프로그램 언어 지원
- Amazon DynamoDB, Amazon SageMaker 접속

동시성 및 배치 프로세스 지원

비용 및 오류 컨트롤

반정형(Semi-Structured) 데이터 지원

데이터 타입 **SUPER**

빠르고 효율적인 JSON 데이터 처리

검색을 위한 유연한 쿼리 PartiQL

구체화된 뷰(Materialized View)를 활용하여
빠르게 컬럼 기반 분석 가능

Redshift Spectrum과 통합하여 활용

id INTEGER	name SUPER	phones SUPER
1	{"given":"Jane", "family":"Doe"}	[{"type":"work", "num":"9255550100"}, {"type":"cell", "num": 6505550101}]
2	{"given":„Richard", "family":„Roe"},	[{"type":"work", "num": 5105550102}]

```
SELECT name.given AS firstname, ph.num
FROM customers c, c.phones ph
WHERE ph.type = 'cell';

firstname | num
----------+---------------
"Jane"    | 6505550101
```

Data API 사용

웹 서비스 기반 애플리케이션으로 데이터 액세스

Python, Go, Java, Node.js 등 다양한 프로그램 언어 지원

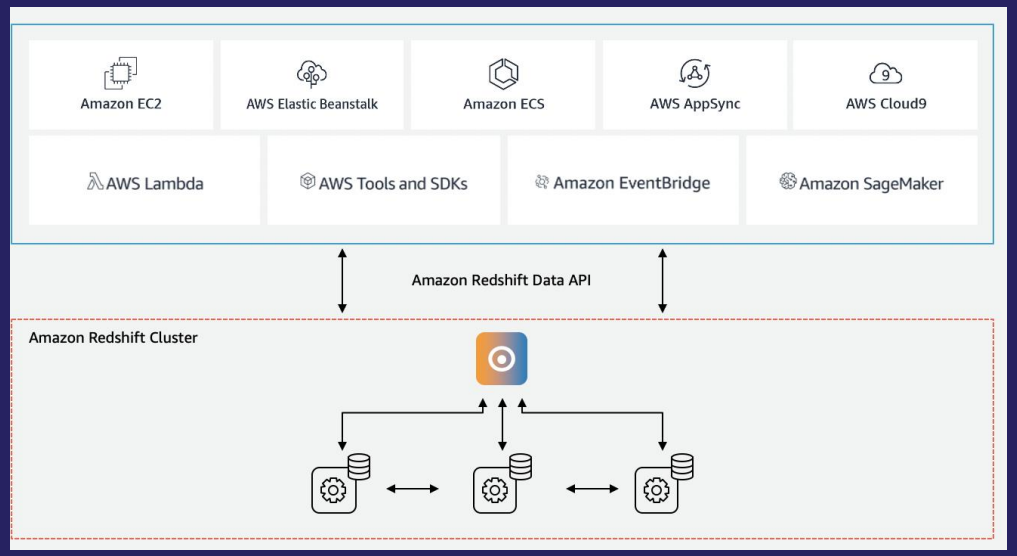
CLI/SDK 활용한 Query, Load, Unload 데이터

클러스터에 지속적인 연결 불필요

Secrets manager의 보안 암호를 활용한 인증 자격 증명

SageMaker NoteBook 활용

```
aws redshift-data execute-statement
--database [DATABASE]
--query [QUERY]
--secret-arn [CREDENTIALS_ARN]
```



자동 테이블 최적화 작업

인공지능 기반 테이블 디자인을 자동으로 최적화

- 자동 정렬키 및 배포키 적용
- 워크로드 패턴을 스캔하여 디자인 자동 최적화
- 데이터와 워크로드 확장에 따른 성능 최적화
- 머신러닝을 활용하여 워크로드 변화에 따라 최적화
- 테이블 기준 적용 가능
- svv_alter_table_recommendations 테이블에 권장 사항 로깅
- svl_auto_worker_action 모든 작업에 대한 감사 로깅 및 이전 상태 표시

Automatic vacuum delete

Automatic distribution keys

Automatic sort keys

Auto workload manager

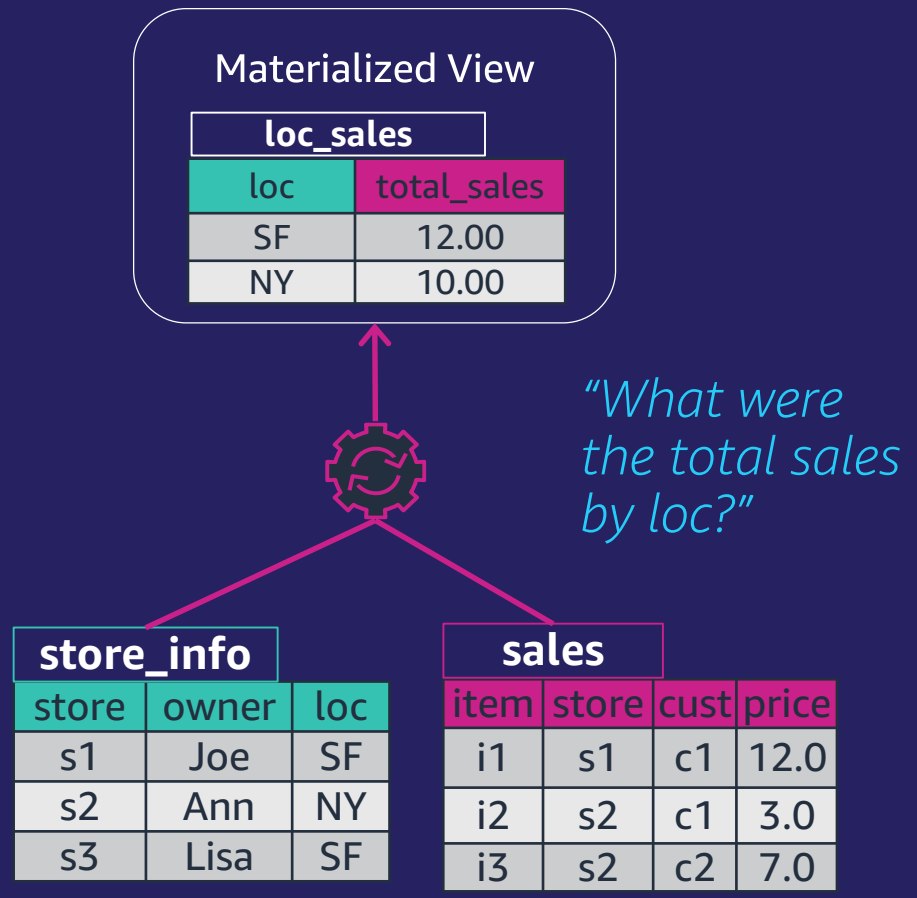
Automatic table sort

MV auto-refresh and rewrite

구체화된 뷰(Materialized views)

Compute once, query many times

- 예측 가능하고 반복되는 쿼리 활용
  - Joins, filters, aggregations, projections
- ETL/BI pipeline을 단순하고 빠르게
  - Incremental refresh
  - Auto refresh
- 자동 쿼리 재작성
- Redshift 로컬, Spectrum, Federated 쿼리 지원



실습 마무리 및 설문 참여 방법

- 실습이 모두 끝난 후에는 자원 삭제를 잊지 마세요. 직접 준비하신 AWS 계정으로 실습을 진행하신 고객 분들의 경우, 가이드에 따라 자원 삭제를 진행하셔야 합니다. 또한, 기존에 사용하시던 자원이 있으신 고객 분들의 경우, **오늘 생성한 자원만 삭제**하는 것에 주의 부탁드립니다.

- **가이드: https://bit.ly/3L5G75q**

- 마지막으로 세션이 끝난 후, **GoToWebinar 창을 종료하면 설문 조사 창**이 나옵니다.
  이때, **설문 조사를 진행해 주셔야 AWS 크레딧**(1인당 $50 크레딧) 을 제공받으실 수 있습니다.

AWS는 고객 피드백을 기반으로 의사 결정을 수행하며 이러한 피드백은 추후에 진행할 세션 방향을 결정합니다.
더 나은 세션을 위하여 여러분들의 소중한 의견을 부탁드립니다.

감사합니다.

실습 마무리 및 설문 참여 방법

- **실습이 모두 끝난 후에는 자원 삭제를 잊지 마세요.** 직접 준비하신 AWS 계정으로 실습을 진행하신 고객 분들의 경우, 가이드에 따라 자원 삭제를 진행하셔야 합니다. 또한, 기존에 사용하시던 자원이 있으신 고객 분들의 경우, **오늘 생성한 자원만 삭제**하는 것에 주의 부탁드립니다.

- **가이드: https://bit.ly/3L5G75q**

- 마지막으로 세션이 끝난 후, **GoToWebinar 창을 종료하면 설문 조사 창**이 나옵니다.
  이때, **설문 조사를 진행해 주시고 '크레딧 제공요청' 을 표기해주셔야 AWS 크레딧**(1인당 $50 크레딧) 을 제공받으실 수 있습니다.

AWS는 고객 피드백을 기반으로 의사 결정을 수행하며 이러한 피드백은 추후에 진행할 세션 방향을 결정합니다.
더 나은 세션을 위하여 여러분들의 소중한 의견을 부탁드립니다.

감사합니다.

더 나은 세미나를 위해 여러분의 의견을 남겨주세요!

- 질문에 대한 답변 드립니다.

Thank you!