재해복구 & 마이그레이션

일시	2023.05.21. (일)	작성자	김민경

1. 재해복구

1) 개요

-재해: 회사의 사업 지속 or 재정에 부정적인 영향 미치는 이벤트

-재해복구: 재해에 대비하고 재해 발생 시 복구하느 ㄴ작업

-재해복구 종류

① 온프레미스 → 온프레미스: 비쌈

② 온프레미스 → AWS Cloud: 하이브리드 복구

(온프레미스를 기본 데이터 센터로 두고 재해 발생 시 클라우드 사용하는 방식)

③ AWS Cloud Region A → AWS Cloud Region B

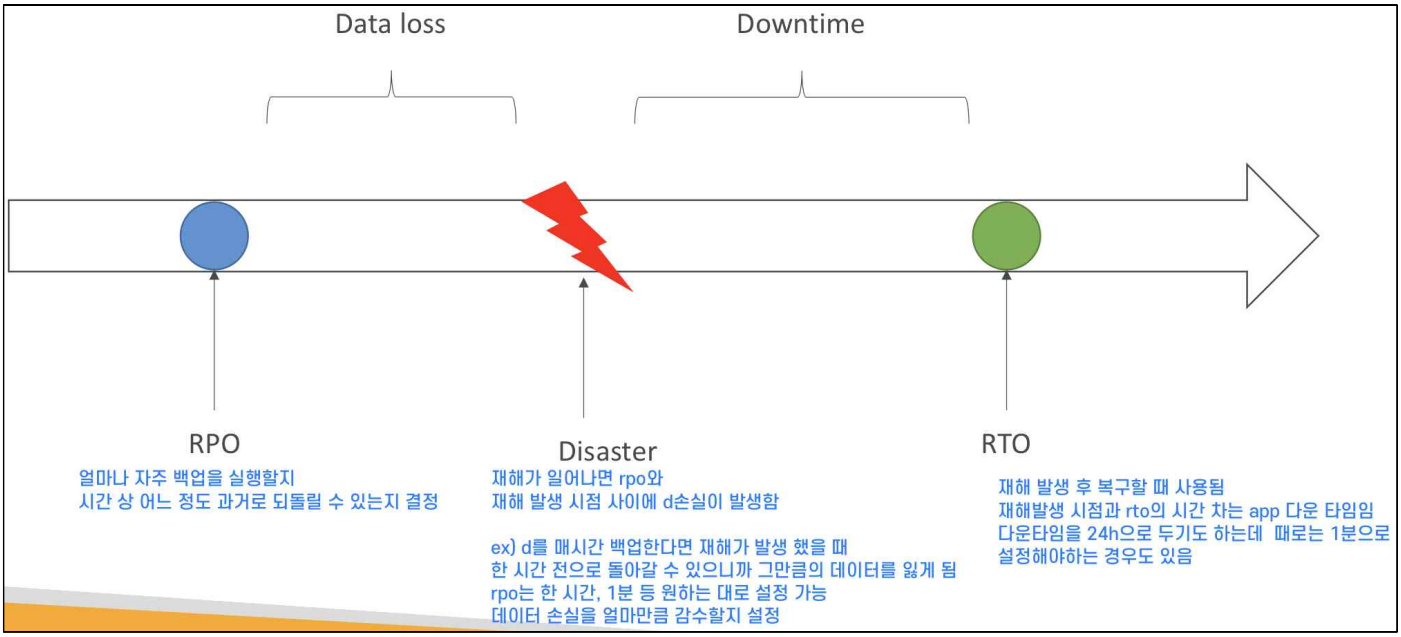
- RTO : Recovery Time Objective(RTO)
  - 얼마나 빨리 백업할 수 있는가?
  - 업무 중단 시점부터 복구되어 가동될 때까지의 시간 목표
- RPO : Recovery Point Objective(RPO)
  - 데이터 손실을 어느 정도까지 허용할 수 있는가?
  - 업무 중단 시점부터 데이터 손실을 수용할 수 있는 시점

2) RPO, RTO

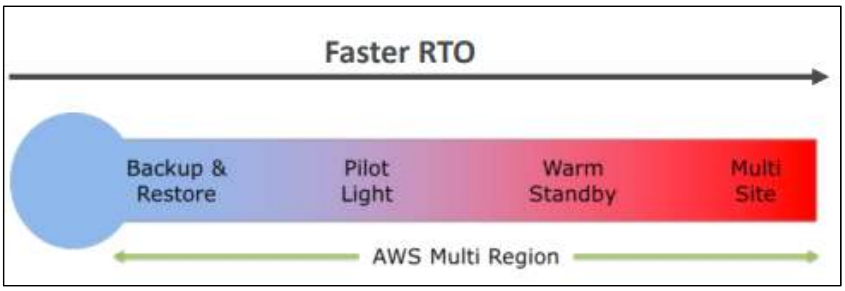
-RPO(Recovery Point Objective): 복구 시점 목표

-RTO(Recovery Time Objective): 복구 시간 목표

-RPO, RTO 최적화 ⇒ 솔루션 아키텍처 결정하는 요인, h 간격 짧을수록 비용 ⇑



3) 재해복구 전략



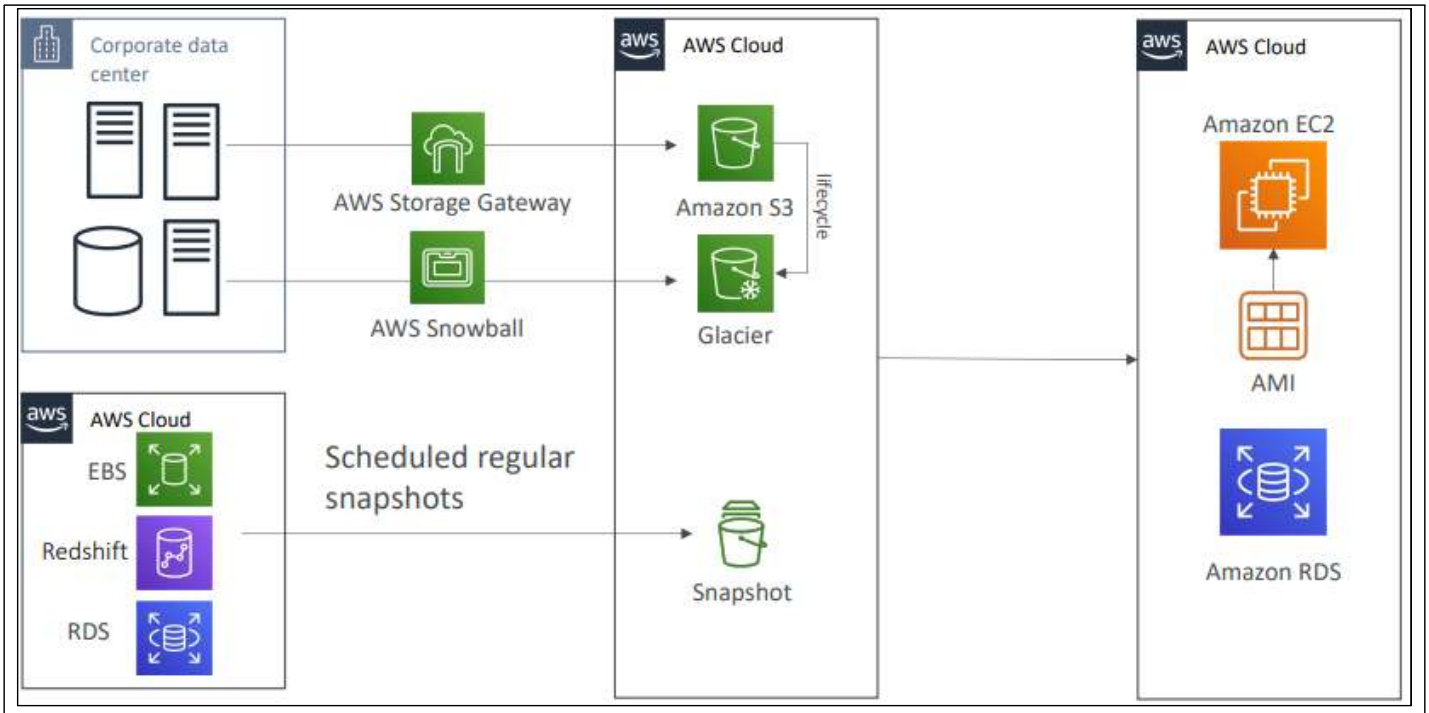
(1) <u>백업 및 복구</u> (Backup and Restore)

-백업 및 복구는 아주 쉽고, 비용 저렴

(중간에 인프라 관리 필요 x이 재해 발생 시 인프라 재생산할 수 있으니
백업 저장 비용 외에는 따로 돈이 들지 x)

-RPO가 높음

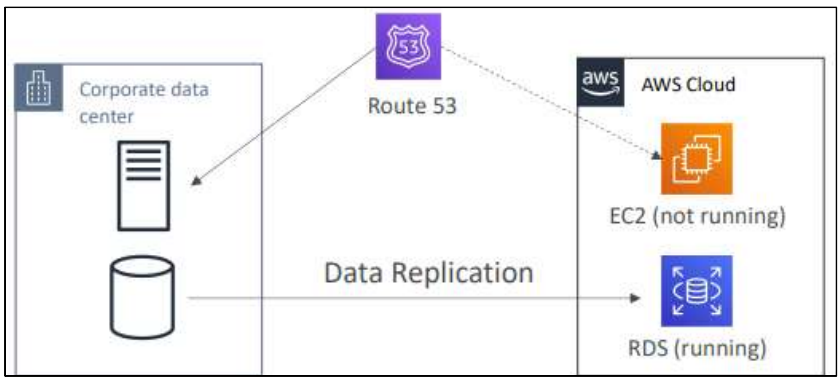
-데이터 복구는 h이 오래 걸려 RTO값도 커짐



• 기업 DC & S3 사용시
-h에 따라 데이터 백업하고 싶으면 Storage Gateway 사용.
수명주기 정책 만들어 비용 최적화 목적으로 Glacier에 데이터 입력
or Snowball 사용해 일주일에 한 번씩 多 데이터를 Glacier에 전송 가능(RPO 대략 일주일)
• Cloud 사용 시
-정기적으로 스냅샷 예약, 스냅샷 만드는 간격(24h, 1h등)에 따라 RPO 달라짐
-재해 발생 시 all 데이터 복구해야하므로 AMI 사용해 EC2 인스턴스 다시 만들고
app을 스핀 업 하거나 EBS, Redshift, RDS 등을 바로 복원 및 재생산

> What does spin up mean?
>
> (computing, transitive) To power up, launch, or instantiate. We spun up a virtual server in the cloud to handle the additional load. (figurative, transitive) To fabricate.

(2) 파일럿 라이트 (Pilot Light)
-app 축소 버전이 클라우드에서 항상 실행되고 보통 크리티컬 코어가 됨(=파일럿 라이트)
-백업 및 복구와 비슷 but 속도 더 빠름(크리티컬 시스템이 이미 가동 中이라)



AWS Pilot light

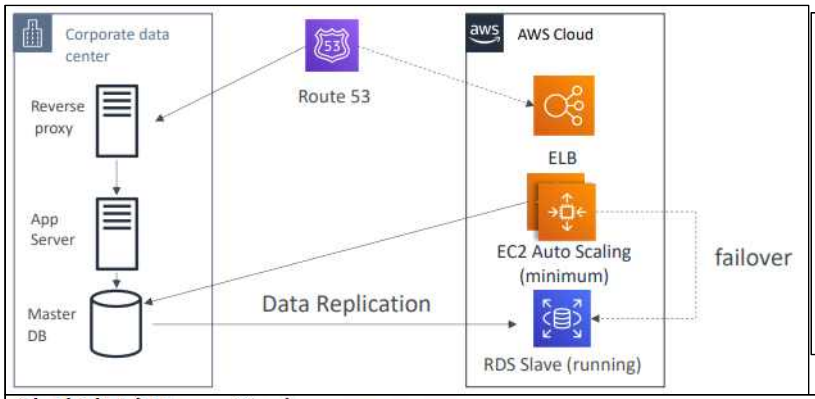
- AWS의 재해 복구 서비스
- Primary Site의 Database는 주기적으로 AWS로 복제되며 Web Server, App Server의 경우 해당 서버들로부터 복제된 AMI를 미리 대기시켜둠
- 장애 발생시 Web AMI / App AMI로부터 빠르게 인스턴스를 생성하며 복제된 Database로부터 데이터를 제공받음
  - Autoscaling를 설정하여 원하는 수의 최소 EC2를 실행할 수 있음
- Route 53 또한 Failover를 실시하여 Primary Site가 아닌 AWS로 라우팅을 실시함

• 크리티컬 DB에서 RDS로 데이터 계속 복제하면⇒ 언제든지 실행할 수 있는 RDS DB 확보하게 됨
• 재해 발생시 Route 53이 DC 서버에 장애 조치를 허용해 클라우드에 EC2 인스턴스 재생산하고 실행하도록 처리함(RDS DB는 이미 준비된 상태임, 나머지는 작동하지 x아서 재해 복구시 EC2 인스턴스만 불러오게 됨) (∴RTO, RPO ⇓짐)
-파일럿 라이트 ⇒ 크리티컬 코어 보조에만 사용됨

(3) **웜 대기** (Warm Standby)

-시스템 전체를 실행하되 최소한의 규모로 가동해서 대기하는 방법

-재해 발생 시 프로덕션 로드로 확장할 수 있음



Warm Standby

- Warm Standby라는 이름처럼 AWS 내에서 이미 실행된 상태에서 대기하는 것
- 실제 워크로드를 감당할 정도의 용량을 갖고 있지 않음
- 장애 발생 후 Failover시 Scale up을 통해 성능을 워크로드를 감당할 수준으로 비약적으로 향상시킴
- Restore Process
  - Route 53 Failover Routing(With Health Check)
  - 워크로드 처리를 위한 Autoscaling
  - 이미 가동중인 상태이기 때문에 RTO는 낮지만, 성능을 올리는데 시간이 소요됨

웜 사이트(Warm Site) 편집 원본 편집

부분적으로 설비를 가지고 있는 백업 사이트로서, 대개 디스크 드라이브, 테이프 드라이브와 같이 가격이 저렴한 주변기기를 가지고 있으나, 주 컴퓨터는 가지고 있지 않다.

- Route 53이 기업 DC로 가리키는 모습
- 클라우드에서는 RDS Slave DB로 데이터 복제가 이루어지고 있음
- EC2 오토 스케일링 그룹이 최소 용량으로 가동해 기업 DC DB와 소통함, ELB 준비된 상태임
- 재해 발생 시 Route 53을 사용해 ELB로 장애 조치해 app이 데이터를 가져오는 곳을 변경하는 작업 가능
  ex) RDS Slave에서 데이터 취하도록 변경한 뒤 효과적으로 대기했다가 오토스케일링 사용하면 app 빠르게 확장함.

-비용이 조금 더 듦(ELB & EC2 오토스케일링이 동시에 실행되서)

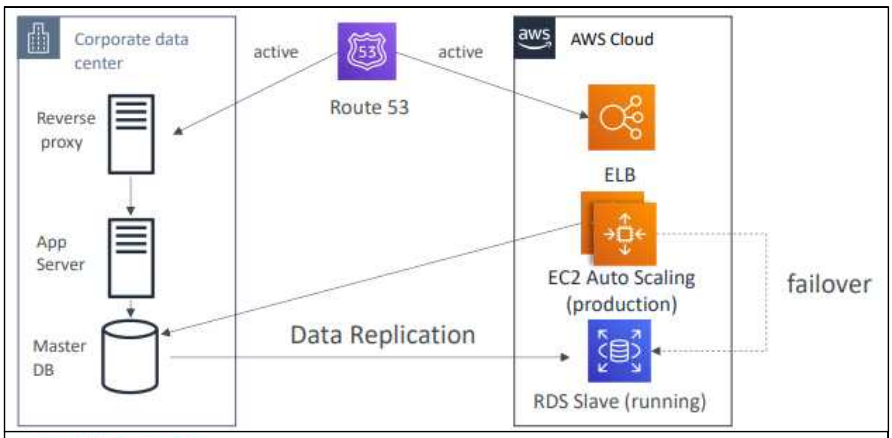
-RTO, RPO ⇓짐

(4) **핫 사이트 / 다중 사이트 접근** (Hot Site / Multi Site Approach)

-RTO 정말⇓ (몇분, 몇초)

-AWS & 온프레미스에서 완전 프로덕션 스케일을 얻음

프로덕션 환경은 소프트웨어 및 기타 제품이 실제로 최종 사용자가 의도 한 용도로 작동하는 설정을 설명하기 위해 개발자가 주로 사용하는 용어입니다. 프로덕션 환경은 프로그램을 실행하고 조직 또는 상업용 일일 운영을 위해 하드웨어 설정을 설치하고 사용하는 실시간 설정으로 생각할 수 있습니다.



2) 핫 사이트(Hot Site)란?

주 센터와 동일한 수준으로 자원을 원격지에 구축하지만, 대기 상태로 유지합니다.

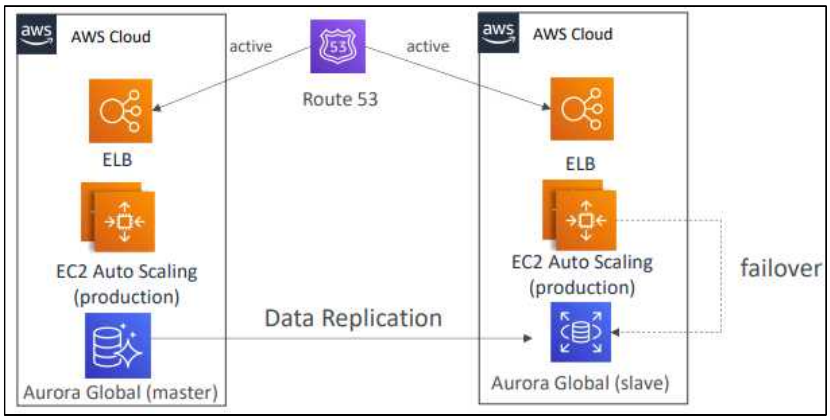
주 센터에 재해 발생 시 원격지 시스템을 운영으로 상태 전환하여 서비스를 제공하게 됩니다.

복구 목표 시간(RTO)는 수시간 이내이고, 미러 사이트와 마찬가지로 데이터의 최신성과 높은 안정성, 신속한 업무 재개를 제공합니다.

핫 사이트도 미러 사이트와 마찬가지로 초기 투자 비용과 유지 보수 비용이 높다는 단점이 있습니다.

- active-active 유형: 서버 두 대 운영 방식

(5) 다중 리전 (Multi Region)



-리전에 마스터 DB가 있고, 다른 리전에 slave로 복제된 Aurora Global DB가 있음
-두 리전 모두 잘 작동함. 장애 발생시 필요에 따라 완전 프로덕션 스케일이 다른 리전에서 가능

여러 AWS 리전을 활용한 백업

AWS Backup 은 중앙 집중적인 관리 기능을 제공할 뿐만 아니라, [그림 2] 와 같이 여러 리전(Region)에 백업의 복제본을 생성 할 수 있습니다. 여러 리전(Region)에 백업 데이터를 복제하여, 특정 리전(Region) 전체의 장애와 같은 상황도 대비 할 수 있습니다. [그림 3] 과 같이 장애가 발생한 리전(Region)과 다른 리전(Region)의 백업 복제본을 이용하여, 시스템을 복구하고 운영 할 수 있습니다.

여러 AWS 리전을 활용한 백업 방법에서는 원본 리전(Region)과 백업 리전(Region)은 서로 다른 계정 (AWS Account) 을 사용하는 것을 권장 합니다. 이렇게 하는 이유는 인적 오류나 장애가 발생한 리전(Region)의 특정 리소스가 백업 리전(Region)에 영향을 주는 것을 최소화 하기 위해서 입니다.

4) 재해복구 tips

(1) 백업
-EBS 스냅샷, RDS로 자동화된 스냅샷 & 백업 등을 사용
-S3, S3 IA, Glacier 등에 스냅샷을 규칙적으로 푸시 가능, 수명 주기 정책 가능, 리전간 복제 가능
-온프레미스 → cloud: : Snowball or Storage Gateway가 유용함

(2) HA
-HA 위해 Route53을 사용해 DNS를 다른 리전으로 옮기면 됨(아주 쉽고 유용한 방법)
-RDS Multi-AZ, ElastiCache Multi-AZ, EFS, S3등이 있음
-네트워크 HA⇒ 기업 DC에서 AWS로 연결할 때 Direct Connect를 실행했을 수도 있음
만약 연결 끊기면 Site to Site VPN을 네트워크 복구 옵션으로 사용할 수 있음

(3) 복제
-RDS 리전간 복제, AWS Aurora Global DB로 복제 가능
-온프레미스 DB를 RDS로 복제
-Storage Gateway

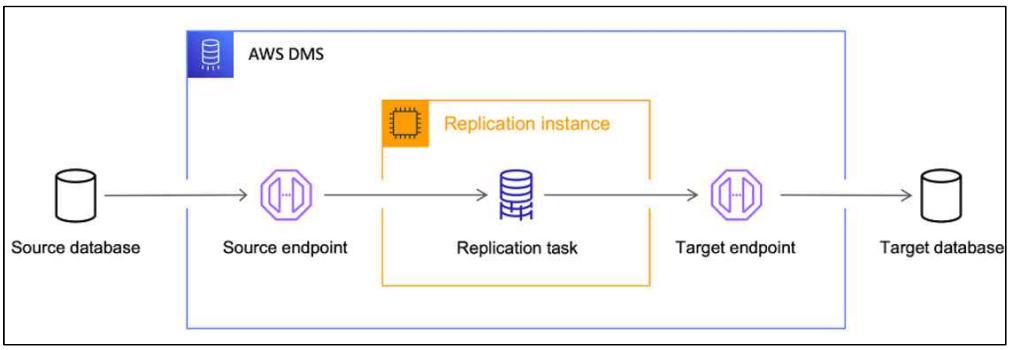
(4) 자동화
-CloudFormation & Elastic Beanstalk ⇒ 클라우드에 새로운 환경을 빠르게 재생산하도록 도움
-CloudFormation를 사용한다면 CloudWatch 경보가 실패했을 때 EC2 인스턴스를 복구하거나 다시 시작 가능
-Lambda⇒ 사용자 맞춤 자동화에 유용함(AWS 인프라 전체를 자동화할 때 효과적)

(5) 카오스 (Chaos)
-카오스 테스트
-재해를 만들어서 대처해 보는 것
-요즘 넷플릭스⇒ simian-army를 만들어 EC2 인스턴스 무작위로 종료함
-인프라 기반을 다져 어떤 장애에도 끄떡없도록

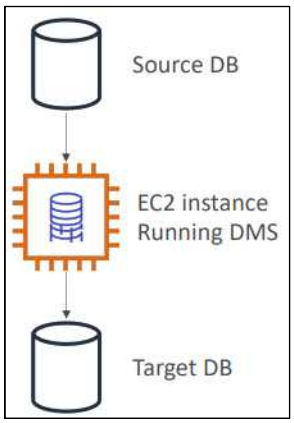
카오스 엔지니어링은 프로덕션 시스템에서 실제 중단 시나리오를 테스트하거나 가능한 프로덕션에 가까운 분산 시스템에서 약점을 찾는 원칙입니다. 즉 프로덕션 환경에서 실험을 통해 시스템에 영향을 끼치는 취약점을 포착하고 개발자가 서비스에 장애를 주입하여 결함을 수정 및 추적합니다.

2. 마이그레이션

1) DMS (Database Migration Service)



(1) 개요



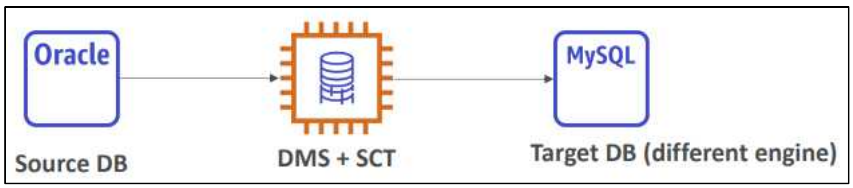
-온프레미스 시스템 → AWS 클라우드로 마이그레이션할 때 사용
-빠르고 안전한 DB 서비스
-복원성 좋고, 자가 복구 가능
-마이그레이션 과정에서 소스 DB도 여전히 사용 가능
-지원
  • 동종 마이그레이션 지원 (ex_Oracle → Oracle)
  • 이기종 마이그레이션 지원(ex_Microsoft SQL → Aurora)
-CDC(Change Data Capture) 사용한 지속적 복제 지원
-EC2 인스턴스 생성해서 복제를 처리하도록 해야함

(2) 대상 & 타겟

SOURCES:	TARGETS:
• On-Premises and EC2 instances databases: *Oracle, MS SQL Server, MySQL, MariaDB, PostgreSQL, MongoDB, SAP, DB2*	• On-Premises and EC2 instances databases: Oracle, MS SQL Server, MySQL, MariaDB, PostgreSQL, SAP
• Azure: *Azure SQL Database*	• Amazon RDS
• Amazon RDS: all including Aurora	• Redshift, DynamoDB, S3
• Amazon S3	• OpenSearch Service
• DocumentDB	• Kinesis Data Streams
	• Apache Kafka
	• DocumentDB & Amazon Neptune
	• Redis & Babelfish

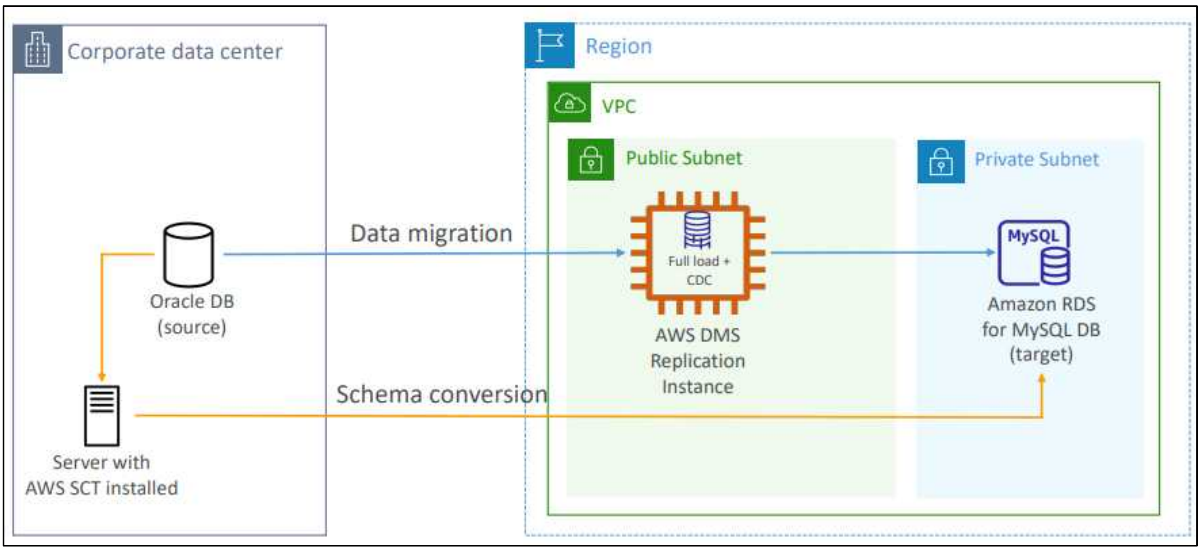
(3) SCT (Schema Conversion Tool)

-만약 소스 DB와 대상 DB가 같은 엔진 갖고 있지 x면⇒ SCT 사용
(같은 엔진 쓰는 DB의 마이그레이션에는 사용x, ex_On-Premise PostgreSQL ⇝ RDS PostgreSQL)



-DB의 스키마를 다른 엔진으로 변환
ex) SQL Server or Oracle ⇝ MySQL, PostgreSQL, Aurora 마이그레이션 가능
Teradata or Oracle 등의 분석과정에서 Amazon Redshift로 변환 가능

(4) DMS - 지속적인 복제 설정 방법



- Oracle DB (소스) ⇝ MySQL DB의 Amazon RDS DB (대상)
  ⇒ 서로 다른 DB
  ⇒ SCT 필요
- DMS에 복제 인스턴스를 설정해
  full load(전체 로드), CDC(진행 중인 복제 또는 변경 데이터 캡처)를 사용할 수 있음

AWS DMS를 사용하여 지속 복제를 위한 작업 생성

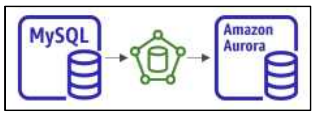
PDF | RSS

원본 데이터 저장소에서 진행 중인 변경 사항을 캡처하는AWS DMS 작업을 생성할 수 있습니다. 데이터를 마이그레이션하는 동안에도 이 변경 사항을 캡처할 수 있습니다. 작업을 생성하여 지원된 대상 데이터 스토어로 초기(전체 로드) 마이그레이션을 완료한 후 지속적 변경 사항을 캡처할 수도 있습니다. 이 프로세스를 진행 중인 복제 또는 변경 데이터 캡처(CDC)라고 합니다. AWS DMS에서는 원본 데이터 스토어에서 지속적 변경 사항을 복제할 때 이 프로세스를 사용합니다. 이 프로세스는 데이터베이스 엔진의 기본 API를 사용하여 데이터베이스 로그에 대한 변경 사항을 수집합니다.

2) RDS & Aurora MySQL Migrations

(1) RDS MySQL ⇝ Aurora MySQL

option 1) RDS MySQL DB에 스냅샷 생성해 Aurora MySQL에 복원



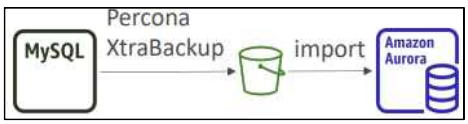
⇒ 다운타임 발생
(∵ 가동 중자하고 마이그레이션해야 되니까)

option 2) Aurora 읽기 전용 복제본을 RDS MySQL에 생성

⇒ 더 지속적인 방법
⇒ 복제본의 지연이 0이 되면 Aurora 복제본이 MySQL과 완전히 일치한다는 뜻
so 복제본을 DB 클러스터로 승격시킴
⇒ h 多걸림, 복제본 생성과 관련된 net 비용 발생

(2) 외부 MySQL ⇝ Aurora MySQL

option 1) Percona XtraBackup 기능을 사용해 백업



⇒ 먼저 백업 파일을 S3에 둔 후에, 이를 새로운 Aurora MySQL DB 클러스터로 가져옴

■ Percona XtraBackup 소개

Percona에서 만든 백업 유틸리티로 MySQL에 사용되는 온라인 백업입니다. mysqldump처럼 논리적인 백업이 아니라 아예 물리적인 파일을 통째로 특정 디렉토리에 복사하는 방법을 사용합니다. 풀백업, 증분백업, 암호화 백업, 압축백업을 지원합니다. MySQL 엔터프라이즈 라이센스에 포함된 백업 도구의 기능을 모두 제공할 뿐만 아니라 더 유용한 기능들도 제공합니다.

option 2) mysqldump 기능 사용해 출력값 내보내기

⇒ mysqldump 기능을 MySQL에서 실행해 기존 Aurora DB에 출력값 내보내는 것
⇒ h 多소요

(3) DMS 이용해 두 DB가 가동된 채로 DB 간 지속적 복제 진행

3) RDS & Aurora PostgreSQL Migrations

(1) RDS PostgreSQL ⇝ Aurora PostgreSQL

option 1) RDS PostgreSQL DB에 스냅샷 생성해 Aurora PostgreSQL에 복원

option 2) PostgreSQL 읽기 전용 복제본을 Aurora에 생성

⇒ 복제 지연이 0이 될 때까지 기다렸다가 DB 클러스커로 승격

(2) 외부 PostgreSQL ⇝ Aurora PostgreSQL

-백업 생성 후 S3에 두고,

데이터를 가져오기 위해 aws_s3 Aurora extension(확장자)를 사용해 새로운 DB 생성 방법

(3) DMS 이용해 두 DB가 가동된 채로 DB 간 지속적 복제 진행

4) On-premise strategy with AWS

(1) Amazon Linux 2 AMI를 가상머신으로 다운로드 (.iso 형식)

- 직접 VM을 통해 온프레미스 인프라에서 Linux 2를 실행할 수 있음

(2) VM Import / Export (가져오기/내보내기)

- 기존 VM과 app을 EC2로 마이그레이션 가능
- 재해복구 리포지토리 전략도 생성 가능
- 즉, 온프레미스 VM이 多 경우 이를 클라우드에 백업하고 싶을 때 가져오기/내보내기 기능을 통해 VM을 EC2에서 온프레미스 환경으로 다시 빼올 수도 있음

(3) AWS Application Discovery Service

- 온프레미스의 정보를 모아주고 마이그레이션을 계획할 수 있게 해주는 서비스
- 서버 사용량 정보 & 종속성 매핑에 대한 정보 제공 (온프레미스 ⇝ 클라우드로 대량의 마이그레이션할 때 유용)
- AWS Migration Hub를 통해 모든 마이그레이션 추적 가능

(4) AWS Database Migration Service (DMS)

- 온프레미스 ⇝ AWS, AWS ⇝ AWS, AWS ⇝ 온프레미스 복제 허용
- 다양한 DB들과 함께 작용해서 사용 편리 (ex_MySQL ⇝ DynamoDB)

(5) AWS Server Migration Service (SMS)

- 온프레미스의 라이브 서버들을 AWS로 증분 복제할 때 사용

***증분 백업**
:마지막 전체 백업 이후에 변경/추가된 데이터만 백업하는 방식

5) AWS Backup

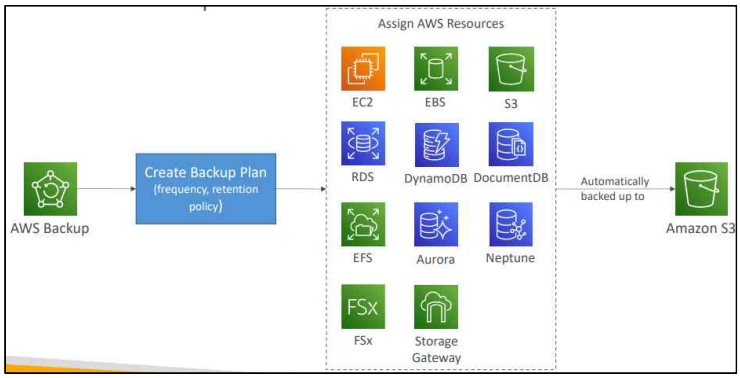
-완전 관리형 서비스

-서비스 간의 백업을 중점적으로 관리하고 자동화할 수 있게 도와줌

-사용자 지정 스크립트 or 매뉴얼 만들 필요x

-다양한 서비스 지원

- Amazon EC2 / Amazon EBS
- Amazon S3
- Amazon RDS (all DBs engines) / Amazon Aurora / Amazon DynamoDB
- Amazon DocumentDB / Amazon Neptune
- Amazon EFS / Amazon FSx (Lustre & Windows File Server)
- AWS Storage Gateway (Volume Gateway)



-리전간 백업 지원

-계정간 백업 지원

-PITR(지정시간 복구) 지원

-온디멘드와 함께 에약된 백업 지원

-태그 기반 백업 정책(ex_프로덕션 태그가 지정된 리소스만 백업 가능)

-백업 정책에서 백업 플랜 만듦

- 백업 빈도 정의
- 백업을 콜드 스토리지로 이전할지의 여부(보내지x, 며칠, 몇주/달/년 후에 보냄)
- 백업 보유기간(계속 보유, 일/주/월)

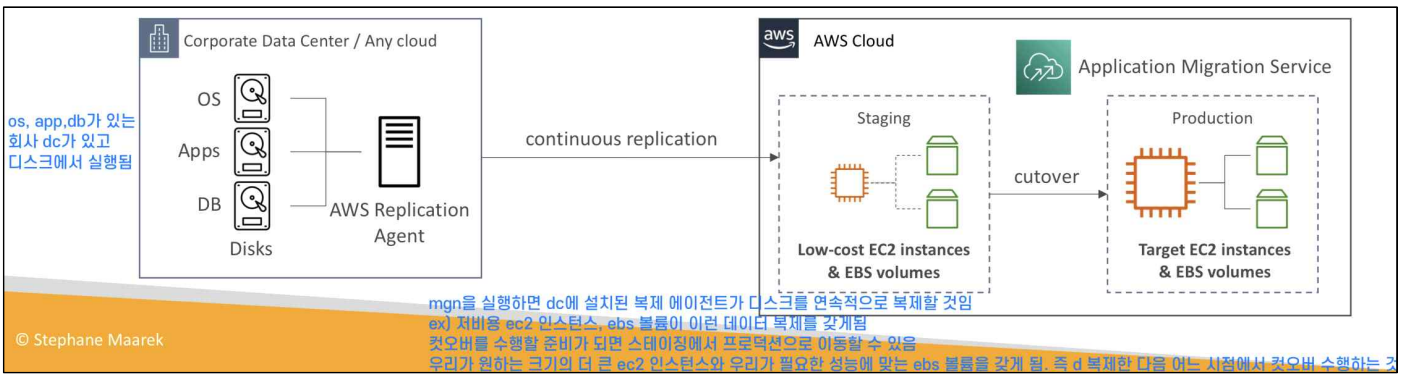
-백업 Vault Lock (볼트 잠금)
⇒ WORM(Write Once Read Many) 정책 시행하면 백업 볼트에 저장한 백업 삭제 x
⇒ 루트 사용자도 백업 삭제 x
⇒ ex) 의도치 않거나 악의적인 삭제 막고, 백업 유지기간 축소 or 변경 작업 방지

6) AWS Application Discovery Service

-온프레미스 서버 or DC 있어서 클라우드로 마이그레이션할 때 계획하는 것
-이동해야 할 항목 & 그것들이 내부적으로 어떻게 상호연결되어 있는지 파악하는데 유용
-서버 스캔하고 마이그레이션에 중요한 서버 설치 데이터 및 종속성 매핑에 대한 정보 수집
-마이그레이션 방법
① Agentless Discovery (AWS Agentless Discovery Connector 사용)
-가상머신, 구성, CPU, 메모리, 디스크 사용량 같은 성능 기록에 대한 정보 제공
-VMware VM (가상 머신)에 대한 정보만 수집할 수 있는 VMware 어플라이언스
② Agent-based Discovery (AWS Applicatio Discovery Agent 실행)
-가상 머신 내에서 더 많은 업데이트와 정보 얻을 수 있음
ex) 시스템 구성, 성능, 실행 중인 프로세스, 시스템 사이의 네트워크 연결에 대한 세부 정보 등
종속성 매핑을 얻는데 좋음
-모든 결과 데이터를 AWS Migration Hub에서 볼 수 있음

7) AWS Application Migration Service(MGN)

-온프레미스 ⇝ AWS로 이동하는 가장 간단한 방법
-AWS로의 애플리케이션 마이그레이션에 권장되는 기본 마이그레이션 서비스
-AWS Server Migration Service(SMS)의 대체



-다운타임 최소(서비스 자동 수행→ 엔지니어 고용x → ∴비용⇓

***컷 오버**
:프로젝트 수행 중 개발환경에서 실 운영환경으로 전환하는 단계

8) 대규모 데이터를 전송하는 방법

(1) 공용 인터넷 사용해 Site-to-Site VPN
- 설치 빠름, 바로 연결 가능
- 데이터 크기에 따라 적합할수도 아닐수도 있음
ex) 200TB ⇝ 1000GB ⇝ 1000BM = 반년 걸림

(2) Direct Connect
- 연결라인 초기설치 h이 오래걸림
- 연결 후에는 (1)보다 빠름
ex) 200TB ⇝ 1000GB ⇝ ~ = 18.5일

(3) Snowball
- 대용량 일회성 전송에 多 사용
- DMS와 결합 가능

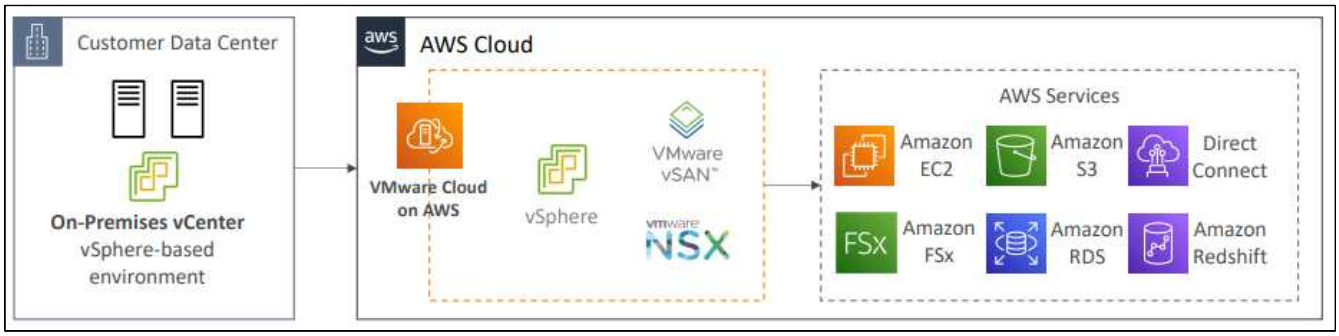
(4) 지속적 복제
- Site-to-Site VPN or DX & DMS or DataSync

***DX**
:Direct Connect
***DataSync**
:온프레미스와 AWS 스토리지 서비스 사이에서 데이터 이동을 자동화하는 서비스

9) **VMware Cloud on AWS**

-AWS와 VMware가 공동으로 개발한 통합 클라우드 솔루션

-온프레미스 VMware vSphere 기반 환경을 EC2에서 실행되는 AWS Cloud에 마이그레이션하고 확장하는 확장성이 뛰어난 보안 서비스를 제공



-VMware cloud의 인프라를 AWS에서 확장함으로써 vSphere 등 사용 가능

-재해 복구 전략으로도 사용할 수 있음

***VMware**

:컴퓨터 가상화 소프트웨어 등 각종 제품을 생산하는 IT 기업