AWS를 사용하는 백업 및 복구 접근 방식

2016년 6월

고지 사항

이 문서는 정보 제공 목적으로만 제공됩니다. 본 문서의 발행일 당시 AWS의 현재 제품 및 실행방법을 설명하며, 예고 없이 변경될 수 있습니다. 고객은 본 문서에 포함된 정보나 AWS 제품 또는 서비스의 사용을 독립적으로 평가할 책임이 있으며, 각 정보 및 제품은 명시적이든 묵시적이든 어떠한 종류의 보증 없이 "있는 그대로" 제공됩니다. 본 문서는 AWS, 그 계열사, 공급업체 또는 라이선스 제공자로부터 어떠한 보증, 표현, 계약 약속, 조건 또는 보증을 구성하지 않습니다. 고객에 대한 AWS의 책임 및 의무는 AWS 계약에 준거합니다. 본 문서는 AWS와 고객 간의 어떠한 계약도 구성하지 않으며 이를 변경하지도 않습니다.

목차

개요 4

서론 4

데이터 보호 플랫폼으로 AWS 를 사용해야 하는 이유 4

데이터 보호를 위한 AWS 스토리지 서비스 5

- Amazon S3 6
- Amazon Glacier 6
- AWS Storage Gateway 7
- AWS 전송 서비스 7

백업 및 복구 솔루션의 설계 7

클라우드 네이티브 인프라 8

- EBS 스냅샷 기반 보호 9
- 데이터베이스 백업 접근 방식 13

온프레미스-AWS 인프라 16

하이브리드 환경 20

- AWS 기반 애플리케이션을 데이터 센터로 백업 21
- 가용성 확보를 위한 백업 관리 기능의 클라우드 전환 22
- 하이브리드 시나리오 예 23
- AWS 를 이용한 데이터 아카이브 24

AWS 의 백업 데이터 보안 25

결론 25

기고자 26

문서 수정 26

개요

본 백서는 엔터프라이즈 솔루션 아키텍트, 백업 아키텍트 및 기업의 IT 환경에서 데이터 보호를 담당하고 있는 IT 관리자를 대상으로 작성되었습니다. 여기에서는 백업 및 복구 솔루션을 강화하거나 대체할 목적으로 AWS를 통해 구현할 수 있는 프로덕션 워크로드 및 아키텍처에 대해 설명하였습니다. 이러한 접근 방식은 비용 절감, 확장성 개선 및 내구성 향상으로 이어져 복구 목표 시간(RTO), 복구 목표 지점(RPO) 및 규정 준수 요건을 해결할 수 있습니다.

서론

엔터프라이즈 데이터의 증가 속도가 점차 빨라지면서 데이터 보호 작업의 어려움도 더욱 가중되고 있습니다. 다음과 같이 백업 방식의 내구성과 확장성에 대한 질문은 이제 어딜 가나 쉽게 들을 수 있습니다. 클라우드가 백업 및 아카이브 요건을 해결하는 데 어떤 역할을 할 수 있습니까?

본 백서에서는 다양한 백업 아키텍처(클라우드 네이티브 애플리케이션, 하이브리드 및 온프레미스 환경)를 비롯해 확장 가능하고 안정적인 데이터 보호 솔루션을 개발하는 데 필요한 AWS 서비스에 대해서 살펴보겠습니다.

데이터 보호 플랫폼으로 AWS를 사용해야 하는 이유

Amazon Web Services(AWS)는 사용이 간편한 클라우드 컴퓨팅 플랫폼으로서 안전성, 고성능, 유연성 및 비용 효율성을 모두 갖추었습니다. AWS는 "매우 어렵긴 해도 공통적으로 적용할 수 있는 서비스(Undifferentiated heavy lifting)"를 표방하며 확장 가능한 백업 및 복구 솔루션을 개발하는 데 사용할 수 있는 도구와 리소스를 제공합니다.

AWS를 데이터 보호 전략으로 사용하는 이점은 아래와 같이 많습니다.

- **내구성:** Amazon Simple Storage Service(Amazon S3)와 Amazon Glacier 저장 객체의 내구성을 99.999999999%까지 보장할 수 있도록 설계되었습니다. 두 플랫폼 모두 백업 데이터의 저장 위치를 안정적으로 제공합니다.

- **보안:** AWS는 접근을 제어하거나, 전송 또는 저장된 데이터를 암호화할 수 있는 다양한 옵션을 지원합니다.
- **글로벌 인프라:** AWS 서비스는 전 세계 어디에서든 사용이 가능하기 때문에 기업의 규정 준수 요건을 만족하는 리전에 데이터를 백업하여 저장할 수 있습니다.
- **규정 준수:** AWS 인프라는 SOC(Service Organization Controls), SSAE(Statement on Standards for Attestation Engagements) 16, ISO(International Organization for Standardization) 27001, PCI DSS(Payment Card Industry Data Security Standard), HIPPA(Health Insurance Portability and Accountability Act), SEC[1], 그리고 FedRAMP(Federal Risk and Authorization Management Program) 같은 표준에 따라 규정 준수 인증을 받았기 때문에 백업 솔루션을 기존 규정 준수 체계에 따라 쉽게 조정할 수 있습니다.
- **확장성:** AWS는 용량 걱정이 필요 없습니다. 관리 오버헤드 없이 요건 변화에 따라 사용량을 확장 또는 축소할 수 있기 때문입니다.
- **TCO 절감:** AWS 운영 규모의 조정으로 서비스 비용이 줄어들기 때문에 스토리지의 총 소유비용(TCO)을 절감하는 데 효과적입니다. AWS는 이러한 비용 절감 효과를 가격 인하의 형태로 고객에게 돌려드립니다.
- **종량 과금제:** 필요할 때 및 사용하려고 계획한 기간에 한해 AWS 서비스를 구매합니다. AWS 요금 정책에는 선불 비용, 해지 위약금 또는 장기 계약이 없습니다.

데이터 보호를 위한 AWS 스토리지 서비스

Amazon S3와 Amazon Glacier는 백업 및 아카이브를 위한 이상적인 서비스입니다. 둘 모두 내구성이 뛰어난 저가형 스토리지 플랫폼입니다. 용량 제한이 없기 때문에 백업 데이터가 증가하더라도 볼륨 또는 미디어를 관리할 필요가 없습니다. 종량 과금제 모델과 1GB당 낮은 월 사용료는 이 서비스가 왜 데이터 보호 사용 사례로 적합한지 잘 보여주는 이유입니다.

[1] https://aws.amazon.com/about-aws/whats-new/2015/09/amazon-glacier-receives-third-party-compliance-assessment-for-sec-rule-17a-4f-from-cohasset-associates-inc/

Amazon S3

Amazon S3는 강력한 보안 및 확장성의 객체 스토리지를 제공합니다.

Amazon S3를 사용하면 웹을 통해 언제 어디서든 원하는 양의 데이터를 저장하고 검색할 수 있습니다. Amazon S3는 데이터를 *버킷*이라는 리소스에 객체의 형태로 저장합니다. AWS Storage Gateway를 비롯한 타사 백업 솔루션들은 사용자를 대신해 Amazon S3 객체를 관리할 수 있습니다. 버킷에 원하는 만큼 객체를 저장할 수 있으며, 버킷에서 객체를 쓰고, 읽고, 삭제할 수도 있습니다. 단일 객체의 크기는 최대 5TB입니다.

Amazon S3는 사용 사례에 따라 다양하게 설계된 스토리지 클래스를 제공하고 있습니다. 클래스는 다음과 같습니다.

- **Amazon S3 Standard** - 액세스 횟수가 잦은 데이터를 위한 범용 스토리지에 적합
- **Amazon S3 Standard - Infrequent Access** 아카이브 기간이 길지만 액세스 횟수가 비교적 적은 데이터에 적합
- **Amazon Glacier** 장기 아카이브에 적합

Amazon S3 역시 구성이 가능한 수명 주기 정책을 제공하기 때문에 수명 주기 내내 데이터를 관리할 수 있습니다. 정책 구성이 완료되면 애플리케이션을 변경하지 않고도 데이터가 해당 스토리지 클래스로 마이그레이션됩니다. 자세한 내용은 S3 스토리지 클래스를 참조하십시오.

Amazon Glacier

Amazon Glacier는 데이터 아카이브 및 온라인 백업을 위한 안전하고 내구성 있는 저장 공간을 제공하는 매우 저렴한 클라우드 아카이브 스토리지 서비스입니다. 이 서비스는 액세스 횟수가 적고, 검색하는 데 2~3시간 정도까지 허용되는 데이터에 최적화되어 비용을 낮게 유지할 수 있습니다. 또한 GB당 월 0.007 USD의 낮은 가격으로 대량 또는 소량의 데이터를 안정적으로 저장할 수 있으므로 온프레미스 솔루션에 비해 상당한 비용 절감 효과가 있습니다. Amazon Glacier는 저장 요건이 장기이거나 무제한인 백업 데이터 스토리지와 장기 데이터 아카이브에 매우 적합합니다. 자세한 내용은 Amazon Glacier를 참조하십시오.

AWS Storage Gateway

AWS Storage Gateway는 온프레미스 소프트웨어 어플라이언스와 클라우드 기반 스토리지를 서로 연결하여 온프레미스 IT 환경과 AWS 스토리지 인프라 간 원활하고 안전한 통합을 보장합니다. 자세한 내용은 AWS Storage Gateway를 참조하십시오.

AWS 전송 서비스

타사의 게이트웨이 및 커넥터 외에도 AWS Direct Connect, AWS Snowball, AWS Storage Gateway 및 Amazon S3 Transfer Acceleration 같은 AWS 옵션을 사용하여 데이터 전송 속도를 높일 수 있습니다. 자세한 내용은 클라우드 데이터 마이그레이션을 참조하십시오.

백업 및 복구 솔루션의 설계

데이터 백업 및 복구를 위한 종합적인 전략을 세울 때는 먼저 발생 가능성이 있는 장애 또는 재해 상황과 그에 따른 잠재적 비즈니스 영향을 파악해야 합니다. 일부 산업의 경우에는 데이터 보안, 프라이버시 및 기록 보존 같은 규제 요건에 대해서도 고려해야 합니다.

백업 프로세스를 구현할 때는 다음을 포함해 비즈니스 RTO 및 RPO를 충족할 수 있도록 적절하게 세분화해야 합니다.

- 파일 수준 복구
- 볼륨 수준 복구
- 애플리케이션 수준 복구(데이터베이스 등)
- 이미지 수준 복구

다음 섹션에서는 인프라 구성에 따른 백업, 복구 및 아카이브 접근 방식에 대해서 설명하겠습니다. IT 인프라는 클라우드 네이티브, 온프레미스 및 하이브리드로 넓게 분류할 수 있습니다.

클라우드 네이티브 인프라

이 시나리오는 AWS 전체에 존재하는 워크로드 환경에 대한 내용입니다. 아래 그림에서도 볼 수 있듯이 이 인프라는 웹 서버, 애플리케이션 서버, 모니터링 서버, 데이터베이스 및 Active directory로 구성됩니다.

AWS의 서비스를 모두 실행할 경우에는 기본적으로 제공되는 다수의 기능을 이용해 데이터 보호 및 복구 요건을 충족할 수 있습니다.

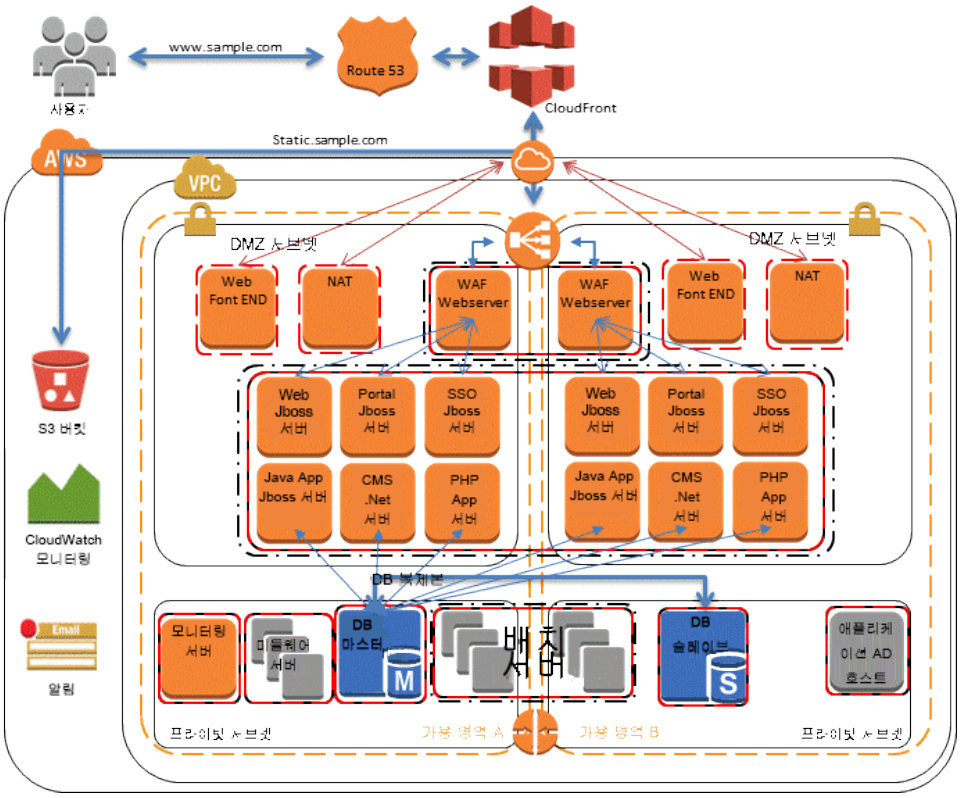

그림 1: AWS 네이티브 시나리오

EBS 스냅샷 기반 보호

서비스를 Amazon Elastic Compute Cloud[2](Amazon EC2)에서 실행할 경우 컴퓨팅 인스턴스가 Amazon Elastic Block Store(Amazon EBS) 볼륨을 사용하여 기본 데이터를 저장하고 액세스합니다. 데이터베이스 같은 정형 데이터 또는 EBS 볼륨 파일 시스템의 파일 같은 비정형 데이터 모두 이 블록 스토리지를 사용할 수 있습니다.

Amazon EBS는 모든 Amazon EBS 볼륨의 스냅샷(백업 파일)을 생성할 수 있는 기능을 지원합니다. 먼저 볼륨 복사본을 만들어 Amazon S3의 여러 가용 영역으로 중복 저장합니다. 첫 번째 스냅샷은 전체 볼륨 복사본이며, 이후 스냅샷은 블록 수준의 증분 변경 사항만 저장합니다.

이 방법은 전체 볼륨 데이터를 빠르고 안전하게 복구할 수 있습니다. 부분 복구만 원할 경우에는 볼륨을 다른 장치 이름에서 실행 중인 인스턴스에 연결하여 마운트한 다음 운영 체제 복사 명령을 사용하여 데이터를 백업 볼륨에서 프로덕션 볼륨으로 복사하면 됩니다.

Amazon EBS 스냅샷은 콘솔이나 명령줄에서 Amazon EBS 스냅샷 복사 기능을 사용해 다른 AWS 리전으로도 복사할 수 있습니다. 자세한 내용은 Amazon Elastic Cloud Compute 사용 설명서를[3] 참조하십시오. 이 기능을 사용하면 기본 복제 기술을 관리할 필요 없이 백업 파일을 다른 리전에 저장할 수 있습니다.

---

[2] http://aws.amazon.com/ec2/

[3] http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ebs-copy-snapshot.html

EBS 스냅샷 만들기

스냅샷을 만들 때는 데이터를 내구성이 좋은 디스크 기반 스토리지에 직접 보호합니다. Amazon EBS 스냅샷은 AWS Management Console, 명령줄 인터페이스(CLI) 또는 API를 사용해 만들 수 있습니다.

Amazon EC2 콘솔의 [**Elastic Block Store Volumes**] 페이지에서 [**Actions**] 메뉴의 [**Create Snapshot**]을 선택합니다. [**Create Snapshot**] 대화 상자에서 [**Create**]를 선택하고 Amazon S3에 저장할 스냅샷을 생성합니다.

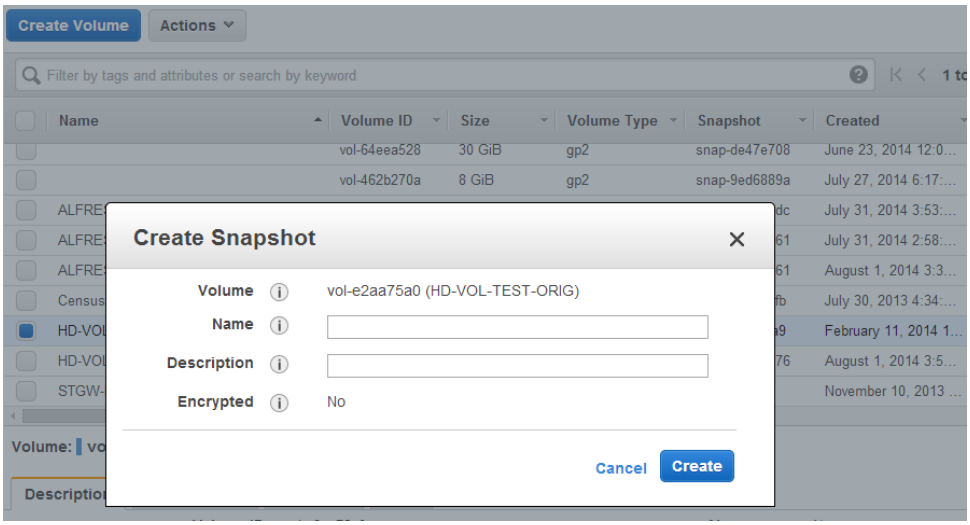

그림 2: EC2 콘솔을 사용하여 스냅샷 만들기

CLI 명령을 사용하여 스냅샷을 만들려면 다음 명령을 실행합니다.

```
➢  aws ec2 create-snapshot
```

`aws ec2 create-snapshot` 명령은 예약을 통해 정기적으로 실행하여 EBS 데이터를 백업할 수 있습니다. 이처럼 다수의 스냅샷을 생성하여 데이터를 보존할 수 있는 이유는 Amazon S3의 경제적 비용에서 비롯됩니다. 또한 스냅샷이 블록 기반이기 때문에 사용하는 저장 공간도 초기 스냅샷을 생성한 이후 변경된 데이터로 제한됩니다.

EBS 스냅샷에서 복구

스냅샷에서 데이터를 복구할 때도 AWS Management Console, CLI 또는 API를 사용하여 기존 스냅샷에서 볼륨을 생성합니다.

예를 들어 다음 단계에 따라 이전 백업 시점으로 볼륨을 복구할 수 있습니다.

1. 다음 명령을 사용하여 백업 스냅샷에서 볼륨을 생성합니다.

```
➢ aws ec2 create-volume --region us-west-1b --snapshot-
  id mysnapshot-id
```

2. Amazon EC2 인스턴스에서 기존 볼륨을 마운트 해제합니다.

   Linux의 경우에는 `umount`를 사용하십시오. Windows의 경우에는 LVM(Logical Volume Manager)을 사용하십시오.

3. 다음 명령을 사용하여 인스턴스에서 기존 볼륨을 분리합니다.

```
➢ aws ec2 detach-volume --volume-id oldvolume-id –
  instance-id myec2instance-id
```

4. 다음 명령을 사용하여 스냅샷에서 생성했던 볼륨을 연결합니다.

```
➢ aws ec2 attach-volume --volume-id newvolume-id
  --instance-id myec2instance-id --device /dev/sdf
```

5. 볼륨을 실행 중인 인스턴스에 다시 마운트합니다.

일관적 또는 핫 백업 파일 만들기

백업을 실행할 때는 시스템이 I/O 연산을 실행하지 않는 상태에서 하는 것이 가장 좋습니다. 이처럼 시스템이 트래픽을 수신하지 않는 것이 가장 이상적이지만 연중무휴 IT 운영이 당연하게 여겨지면서 이러한 백업 방식은 점차 사라지고 있습니다.

이러한 이유로 전체 백업을 실행하기 위해서는 파일 시스템이나 데이터베이스의 작업을 거부해야 합니다. 작업 거부 방식은 데이터베이스나 파일 시스템에 따라 달라집니다.

데이터베이스인 경우 프로세스는 다음과 같습니다.

- 가능하다면 데이터베이스를 핫 백업 모드로 전환합니다.
- Amazon EBS 스냅샷 명령을 실행합니다.
- 데이터베이스의 핫 백업 모드를 해제하거나, 혹은 읽기 복제본을 사용하는 경우에는 읽기 복제본 인스턴스를 종료합니다.

파일 시스템의 프로세스도 비슷하지만 운영 체제나 파일 시스템의 기능에 따라 차이가 있습니다. 예를 들어 XFS는 일관된 백업을 위해 데이터를 플러시할 수 있는 파일 시스템입니다. 자세한 내용은 xfs_freeze[4]를 참조하십시오.

파일 시스템이 작업 거부 기능을 지원하지 않는 경우에는 파일 시스템 마운트를 해제하여 스냅샷 명령을 실행한 후 다시 파일 시스템을 마운트해야 합니다. 그 밖에 I/O 연산 중단 기능을 지원하는 논리 볼륨 관리자를 사용해도 이러한 프로세스에 도움이 됩니다.

스냅샷 프로세스가 백그라운드에서 계속 실행되면서 스냅샷을 빠르게 생성하여 시점을 포착하기 때문에 백업하려는 볼륨은 아주 잠시지만 마운트가 해제되어야 합니다. 백업이 허용되는 시간은 최대한 짧기 때문에 중단 시간을 예측하여 사전에 예약할 수 있습니다.

다중 볼륨 백업

경우에 따라 논리 볼륨 관리자를 사용하여 잠재적 처리량을 높임으로써 데이터를 다수의 Amazon EBS 볼륨으로 분산 저장할 수 있습니다. 논리 볼륨 관리자(mdadm 또는 LVM)를 사용할 때는 기본 EBS 볼륨이 아닌 볼륨 관리자 계층에서 백업을 실행하는 것이 중요합니다. 그래야만 모든 메타데이터가 일관적이고 하위 구성 요소 볼륨이 논리적이기 때문입니다.

---

[4] https://access.redhat.com/documentation/en-US/Red_Hat_Enterprise_Linux/6/html/Storage_Administration_Guide/xfsfreeze.html

이 백업 방법은 다양합니다. 예를 들어 alestic.com[5]에서 작성한 스크립트를 사용해도 됩니다. 이때 메모리 버퍼는 디스크로 플러싱되고, 디스크에 대한 파일 시스템 I/O는 중단되고, RAID 집합을 구성하는 모든 볼륨에서 동시에 스냅샷이 시작되어야 합니다. 볼륨 스냅샷이 시작된 후에도 약 1~2초 동안은 파일 시스템이 연산을 계속 실행할 수도 있습니다. 따라서 복구 도중 스냅샷을 집중 관리하려면 스냅샷에 태그를 지정하는 것이 좋습니다.

이러한 백업은 논리 볼륨 관리자나 파일-시스템 수준에서 실행할 수도 있습니다. 이 경우 기존의 백업 에이전트를 사용하더라도 네트워크를 통한 데이터 백업이 가능합니다. 에이전트 기반 백업 솔루션은 인터넷을 비롯해 AWS Marketplace[6]에서도 다양하게 구매할 수 있습니다. 단, 에이전트 기반 백업 소프트웨어는 서버 이름과 IP 주소가 일정합니다. 따라서 이처럼 인스턴스가 Amazon 가상 프라이빗 클라우드(VPC)[7]에 배포되는 도구를 사용하는 것이 가장 안정적인 방법입니다.

그 밖에 대용량 단일 볼륨에 존재하는 기본 시스템 볼륨의 복제본을 생성하는 것도 또 하나의 방법입니다. 이렇게 하면 대용량 볼륨 하나만 백업할 뿐 기본 시스템에서는 백업이 불필요하기 때문에 백업 프로세스가 간소화됩니다. 하지만 먼저 백업 도중 단일 볼륨의 실행 가능 여부와 애플리케이션에 대한 최대 볼륨 크기의 적합성 여부를 결정해야 합니다.

데이터베이스 백업 접근 방식

AWS는 데이터베이스 옵션이 많습니다. 자신의 데이터베이스를 EC2 인스턴스에서 실행하거나, 혹은 Amazon Relational Database Service[8](Amazon RDS)에서 제공하는 관리형 서비스 데이터베이스 옵션 중 하나를 사용할 수도 있습니다. 자신의 데이터베이스를 EC2 인스턴스에서 실행하는 경우에는 기본 도구(MySQL[9], Oracle[10], MSSQL[11], PostgreSQL[12])를 사용하여 데이터를 파일로 백업하거나 "EBS 스냅샷 기반 보호"에서 설명한 방법 중 하나를 사용하여 데이터가 저장된 볼륨의 스냅샷을 만드는 방법도 있습니다.

---

[5] https://github.com/alestic/ec2-consistent-snapshot

[6] https://aws.amazon.com/marketplace/

[7] http://aws.amazon.com/vpc/

[8] https://aws.amazon.com/rds/

[9] http://dev.mysql.com/doc/refman/5.7/en/backup-and-recovery.html

[10] http://docs.oracle.com/cd/E11882_01/backup.112/e10642/rcmbckba.htm#BRADV8003

[11] http://msdn.microsoft.com/en-us/library/ms187510.aspx

[12] http://www.postgresql.org/docs/9.3/static/backup.html

데이터베이스 복제본 백업

Amazon EBS 볼륨의 RAID 집합을 기반으로 구축된 데이터베이스의 경우 데이터베이스의 읽기 복제본을 생성하여 기본 데이터베이스의 백업 부담을 제거할 수 있습니다. 이것은 최신 데이터베이스 복제본으로 별도의 Amazon EC2 인스턴스에서 실행됩니다. 이 복제본 데이터베이스 인스턴스는 원본과 비슷한 다수의 디스크를 사용해 생성하거나, 혹은 데이터를 단일 EBS 볼륨으로 통합할 수도 있습니다. 그런 다음 "EBS 스냅샷 기반 보호"에서 설명한 절차 중 한 가지를 사용해 EBS 볼륨 스냅샷을 생성합니다. 이러한 접근 방식은 종종 연중 무휴 운영이 필요한 대용량 데이터베이스에서 사용됩니다. 이런 경우에는 백업에 필요한 기간이 너무 길어 장시간 프로덕션 데이터베이스를 중단할 수 없기 때문입니다.

Amazon RDS를 사용한 백업

Amazon RDS는 데이터베이스 백업을 자동화할 수 있는 기능을 지원합니다. Amazon RDS는 개별 데이터베이스가 아닌 전체 DB 인스턴스를 백업하여 DB 인스턴스의 스토리지 볼륨 스냅샷을 생성합니다.

Amazon RDS는 다음과 같이 DB 인스턴스 백업 및 복구를 위한 두 가지 방법을 제공합니다.

- **자동 백업** 기능은 DB 인스턴스의 특정 시점 복구가 가능합니다. 이 기능은 새로운 DB 인스턴스를 생성할 때 기본적으로 활성화됩니다. 그러면 DB 인스턴스를 생성할 때 정의한 시간에 Amazon RDS가 데이터 전체를 매일 백업합니다. 자동 백업은 최대 35일까지 보존 기간을 구성할 수 있습니다. Amazon RDS는 이러한 정기 데이터 백업과 트랜잭션 로그를 함께 사용하여 최대 `LatestRestorableTime`(일반적으로 마지막 5분)까지 DB 인스턴스를 보존 기간 중 원하는 시점(단위: 초)으로 복구할 수 있도록 지원합니다. DB 인스턴스의 최신 복구 가능 시간을 확인하려면 `DescribeDBInstances` API 호출을 사용하거나 AWS Management Console에서 데이터베이스 **[Description]** 탭을 살펴보십시오.

  특정 시점 복구를 시작하면 DB 인스턴스를 요청 시간으로 복구할 수 있도록 트랜잭션 로그가 가장 적합한 매일 백업으로 적용됩니다.

- **DB 스냅샷**은 사용자가 직접 실행하는 백업 방법으로서 DB 인스턴스를 원하는 만큼 자주 확인된 상태로 백업한 다음 언제든지 해당 상태로 복구할 수 있습니다. DB 스냅샷은 AWS Management Console 또는 `CreateDBSnapshot` API 호출을 사용해 생성할 수 있습니다. 스냅샷 보존 기간은 무제한입니다. 콘솔 또는 `DeleteDBSnapshot` API 호출을 사용해 명시적으로 삭제하기 전까지는 영구 보존됩니다.

데이터베이스를 특정 시점으로, 또는 DB 스냅샷에서 복구할 때는 새로운 데이터베이스 인스턴스가 새로운 엔드포인트와 함께 생성됩니다. 이러한 방식으로 특정 DB 스냅샷이나 특정 시점에서 다수의 데이터베이스 인스턴스를 생성할 수 있습니다.

이전 데이터베이스 인스턴스는 AWS Management Console 또는 `DeleteDBInstance` 호출을 사용해 삭제할 수 있습니다.

AMI를 사용한 EC2 인스턴스 백업

AWS는 Amazon Machine Image(AMI)라는 곳에 시스템 이미지를 저장합니다. 이 이미지는 인스턴스 실행에 필요한 루트 볼륨 템플릿으로 구성됩니다. 루트 볼륨은 AWS Management Console 또는 `aws ec2 create-image` CLI 명령을 사용해 AMI로 백업할 수 있습니다.

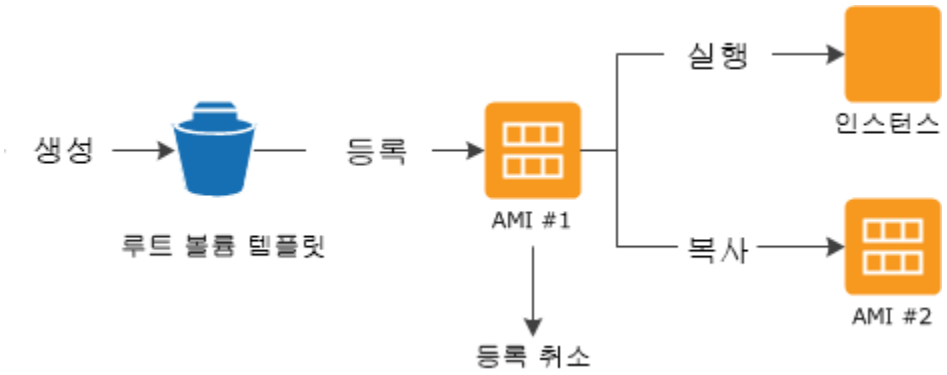

그림 3: AMI를 사용한 인스턴스 백업 및 실행

AMI를 등록하면 Amazon EBS 스냅샷으로 계정에 저장됩니다. 이 스냅샷은 Amazon S3에 상주하여 내구성이 매우 우수합니다.

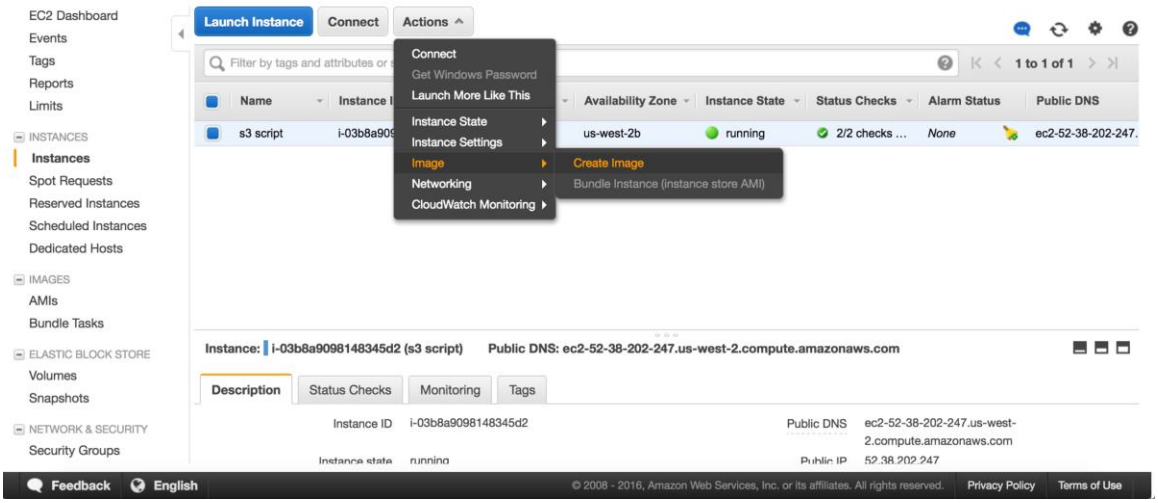

그림 4: EC2 콘솔을 사용한 Machine Image 생성

Amazon EC2 인스턴스의 AMI를 생성하였으면 이제 AMI를 사용하여 인스턴스를 재생성하거나 더 많은 인스턴스 복사본을 실행할 수 있습니다. 또한 AMI를 다른 리전으로 복사하여 애플리케이션 마이그레이션 또는 재해 복구도 가능합니다.

온프레미스-AWS 인프라

이 시나리오는 클라우드 구성 요소가 없는 워크로드 환경에 대한 내용입니다. 웹 서버, 애플리케이션 서버, 모니터링 서버, 데이터베이스, Active Directory 등을 포함한 모든 리소스가 고객 데이터 센터나 코로케이션을 통해 호스팅됩니다.

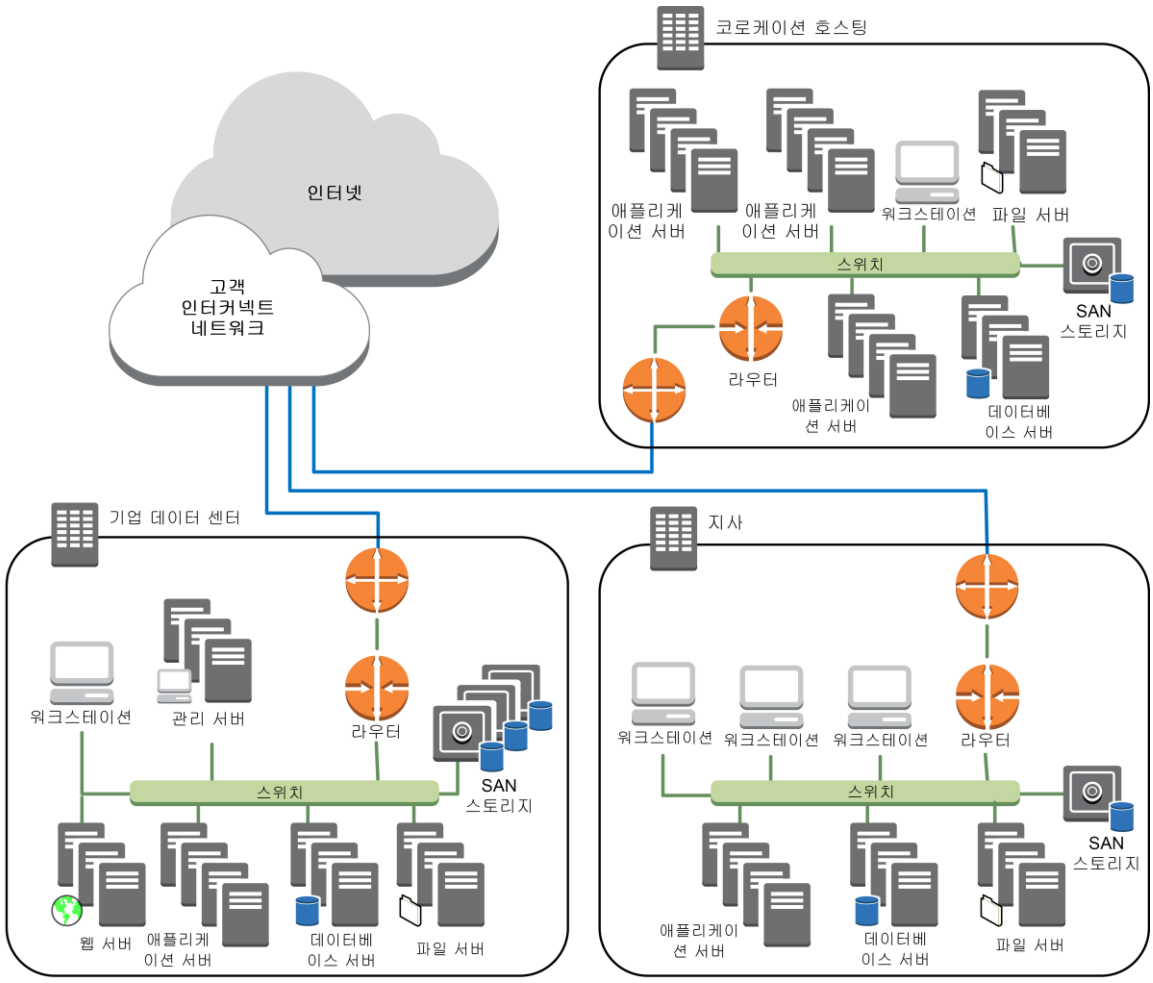

그림 5: 온프레미스 환경

이 시나리오에서는 AWS 스토리지 서비스를 사용하여 백업 및 아카이브 작업에 집중할 수 있습니다. 백업 작업에 따르는 스토리지 확장이나 인프라 용량에 대해 걱정할 필요가 없습니다.

Amazon S3 및 Amazon Glacier는 기본적으로 API 기반이기 때문에 인터넷에서도 사용할 수 있습니다. 백업 소프트웨어 공급업체가 아래 그림과 같이 자신의 애플리케이션을 AWS 스토리지 솔루션에 직접 통합할 수 있는 이유도 여기에 있습니다.

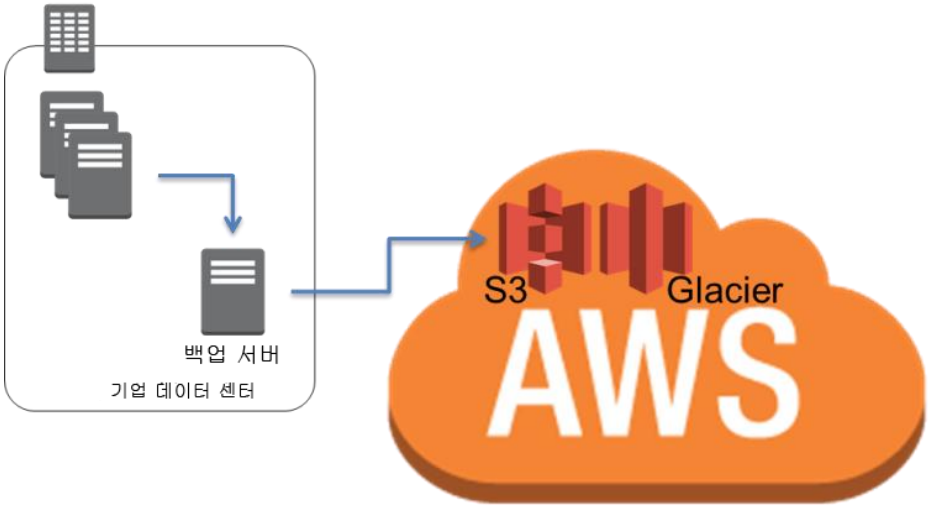

그림 6: Amazon S3 또는 Amazon Glacier에 연결되는 백업 커넥터

이 시나리오에서 백업 및 아카이브 소프트웨어는 API를 통해 AWS에 직접 연결됩니다. 백업 소프트웨어가 AWS를 인식하기 때문에 온프레미스 서버의 데이터를 Amazon S3 또는 Amazon Glacier로 직접 백업합니다.

기존 백업 소프트웨어가 기본적으로 AWS 클라우드를 지원하지 않을 경우에는 AWS 스토리지 게이트웨이 제품을 사용하십시오. AWS Storage Gateway[13]는 데이터 센터와 AWS 스토리지 인프라를 서로 원활하고 안전하게 통합할 수 있는 가상 어플라이언스입니다. 이 서비스를 사용하면 AWS 클라우드에 데이터를 안전하게 저장하여 확장 가능하면서 비용 효율적인 스토리지를 구현할 수 있습니다. Storage Gateway는 업계 표준 스토리지 프로토콜을 지원하여 모든 데이터를 암호화하고 Amazon S3 또는 Amazon Glacier에 안전하게 저장하는 동시에 기존 애플리케이션과도 호환이 가능합니다.

[13] http://aws.amazon.com/storagegateway/

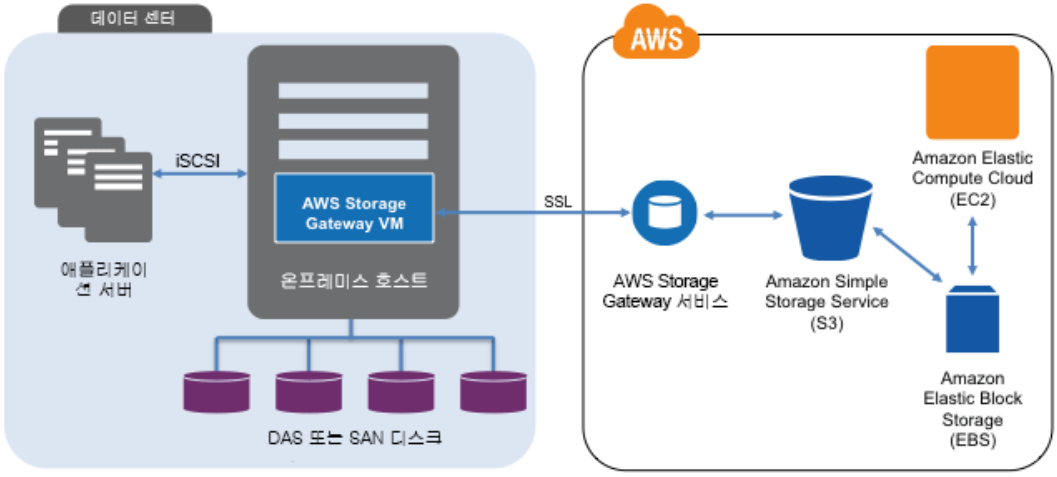

그림 7: 온프레미스와 AWS 스토리지의 연결

AWS Storage Gateway는 다음과 같은 구성을 지원합니다.

- **볼륨 게이트웨이:** 볼륨 게이트웨이는 온프레미스 애플리케이션 서버에서 iSCSI(Internet Small Computer System Interface) 장치로 마운트할 수 있도록 클라우드 기반 스토리지 볼륨을 제공합니다. 이 게이트웨이가 지원하는 볼륨 구성은 아래와 같습니다.
  - **게이트웨이 캐싱 볼륨:** 기본 데이터는 Amazon S3에 저장하고, 자주 액세스하는 데이터는 로컬에 저장합니다. 게이트웨이 캐싱 볼륨은 기본 스토리지에 미치는 비용 절감 효과가 매우 클 뿐만 아니라 온프레미스 스토리지의 확장 필요성을 최소화하고 자주 액세스하는 데이터에 대한 액세스 지연 시간을 낮게 유지합니다.
  - **게이트웨이 저장 볼륨:** 전체 데이터 세트에 대한 액세스 지연 시간을 낮게 유지해야 하는 경우에는 온프레미스 데이터 게이트웨이를 구성하여 기본 데이터를 로컬에 저장하고, 이 데이터의 특정 시점 스냅샷을 비동기 방식으로 Amazon S3에 백업할 수 있습니다. 게이트웨이 저장 볼륨은 내구성과 경제성이 좋은 오프사이트 백업 방식으로 로컬이나 Amazon EC2에서 복구할 수 있습니다.

- **게이트웨이 가상 테이프 라이브러리(게이트웨이-VTL):** 게이트웨이-VTL은 저장할 수 있는 가상 테이프 수가 무제한입니다. 가상 테이프는 각각 Amazon S3 기반 가상 테이프 라이브러리 또는 Amazon Glacier 기반 가상 테이프 선반에 저장됩니다. 가상 테이프 라이브러리는 산업 표준 iSCSI 인터페이스를 지원하기 때문에 백업 애플리케이션에서 가상 테이프에 온라인으로 액세스할 수 있습니다. 가상 테이프에 저장된 데이터에 더 이상 액세스할 필요가 없을 때는 백업 애플리케이션을 사용해 가상 테이프 라이브러리에서 가상 테이프 선반으로 데이터를 마이그레이션하여 스토리지 비용을 줄일 수도 있습니다.

이러한 게이트웨이는 표준 iSCSI 장치를 제공하는 플러그-앤-플레이 방식의 장치 역할을 하기 때문에 백업 또는 아카이브 프레임워크에 통합 가능합니다. iSCSI 디스크 장치를 스토리지 풀로서 백업 소프트웨어 또는 게이트웨이-VTL에 사용하여 테이프 기반 백업 또는 아카이브를 직접 Amazon S3 또는 Amazon Glacier로 오프로드할 수 있습니다.

이 방법을 사용하면 백업 데이터와 아카이브가 자동으로 오프사이트(규정 준수 목적)로 내구성이 좋은 미디어에 저장되기 때문에 오프사이트 테이프 관리의 복잡성과 보안 위험이 사라집니다.

하이브리드 환경

지금까지 언급한 두 가지 인프라 배포 방식인 클라우드 네이티브와 온프레미스는 하이브리드 시나리오로 결합할 수 있습니다. 그러면 온프레미스 구성 요소와 AWS 인프라 구성 요소를 모두 갖춘 워크로드 환경이 구축됩니다. 웹 서버, 애플리케이션 서버, 모니터링 서버, 데이터베이스, Active Directory 등을 포함한 리소스는 고객 데이터 센터나 AWS에 호스팅됩니다. 그리고 AWS 클라우드 기반 애플리케이션이 온프레미스 기반 애플리케이션에 연결됩니다.

이 방식은 오늘날 엔터프라이즈 워크로드 시나리오로 가장 많이 사용되고 있습니다. 대부분 기업들은 자체 데이터 센터에서 AWS를 사용해 용량을 늘립니다. 이러한 고객 데이터 센터는 고용량 네트워크 링크를 통해 AWS 네트워크에 연결되는 경우가 많습니다. 예를 들어 AWS Direct Connect[14]를 이용하면

[14] http://aws.amazon.com/directconnect/

온프레미스에서 AWS로 사설 전용 연결을 구성할 수 있습니다. 그러면 데이터 보호를 목적으로 데이터를 클라우드에 업로드할 수 있는 대역폭과 일정한 지연 시간을 확보할 뿐만 아니라 하이브리드 워크로드를 위한 성능과 지연 시간도 일관되게 유지할 수 있습니다.

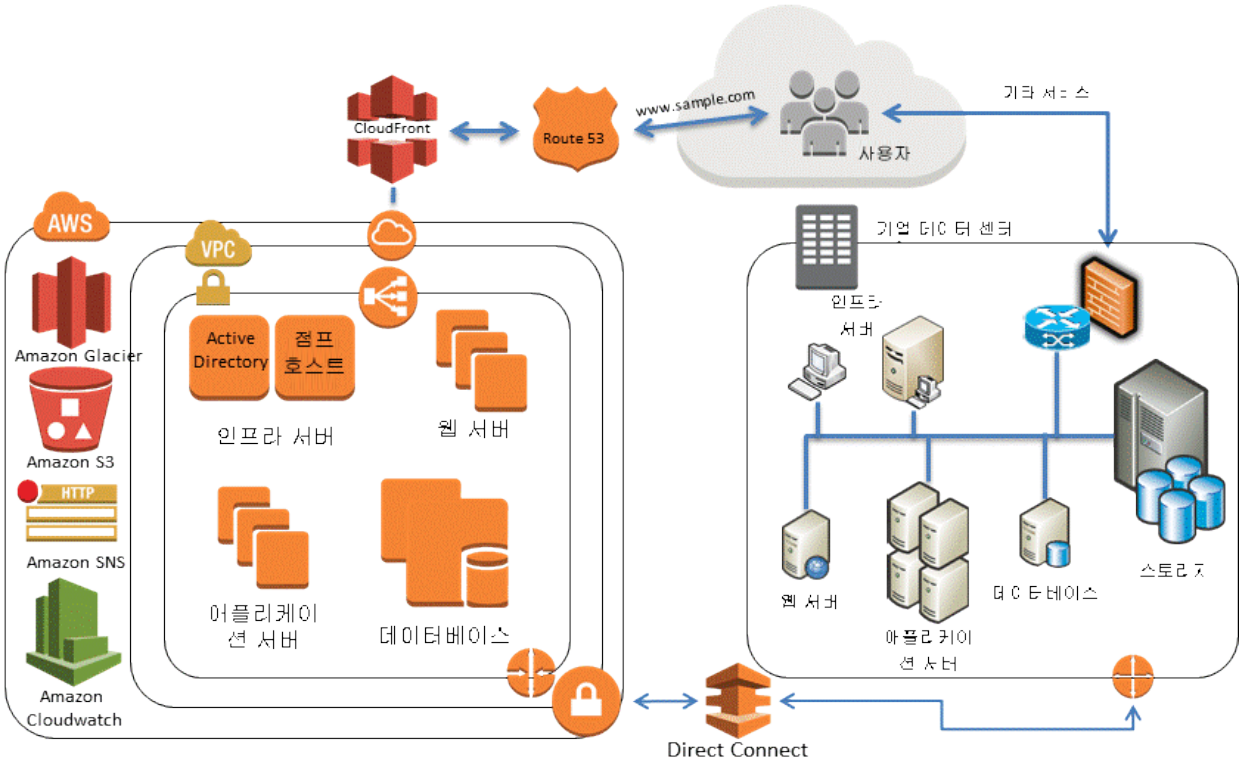

그림 8: 하이브리드 인프라 시나리오

효과적으로 설계된 데이터 보호 솔루션이라고 하면 일반적으로 클라우드 네이티브 및 온프레미스 솔루션에서 언급한 방법들을 조합하여 사용합니다.

AWS 기반 애플리케이션을 데이터 센터로 백업

온프레미스 서버를 위한 데이터 백업 프레임워크가 이미 구현되어 있다면 VPN 연결 또는 AWS Direct Connect를 통해 AWS 리소스로 확장하는 일이 쉽습니다. 백업 에이전트를 Amazon EC2 인스턴스에 설치한 후 데이터 보호 정책에 따라 백업하기만 하면 됩니다.

가용성 확보를 위한 백업 관리 기능의 클라우드 전환

백업 아키텍처에 따라 마스터 백업 서버 1개, 그리고 미디어 또는 스토리지 서버 1개 이상이 보호 서비스와 함께 온프레미스 환경에 설치되어 있는 경우가 많습니다. 이때 마스터 백업 서버를 Amazon EC2 인스턴스로 마이그레이션하여 온프레미스 재해에서 보호하는 동시에 백업 인프라의 고가용성을 확보해야 할 수도 있습니다.

또한 백업 데이터의 흐름을 관리하려면 미디어 서버 1개 이상을 Amazon EC2 인스턴스에 생성해야 할 수도 있습니다. 미디어 서버를 Amazon EC2 인스턴스에 가까이 설치하면 인터넷 전송 비용 절감은 물론이고 S3 또는 Amazon Glacier로 백업할 때 전체 백업 및 복구 성능이 향상됩니다.

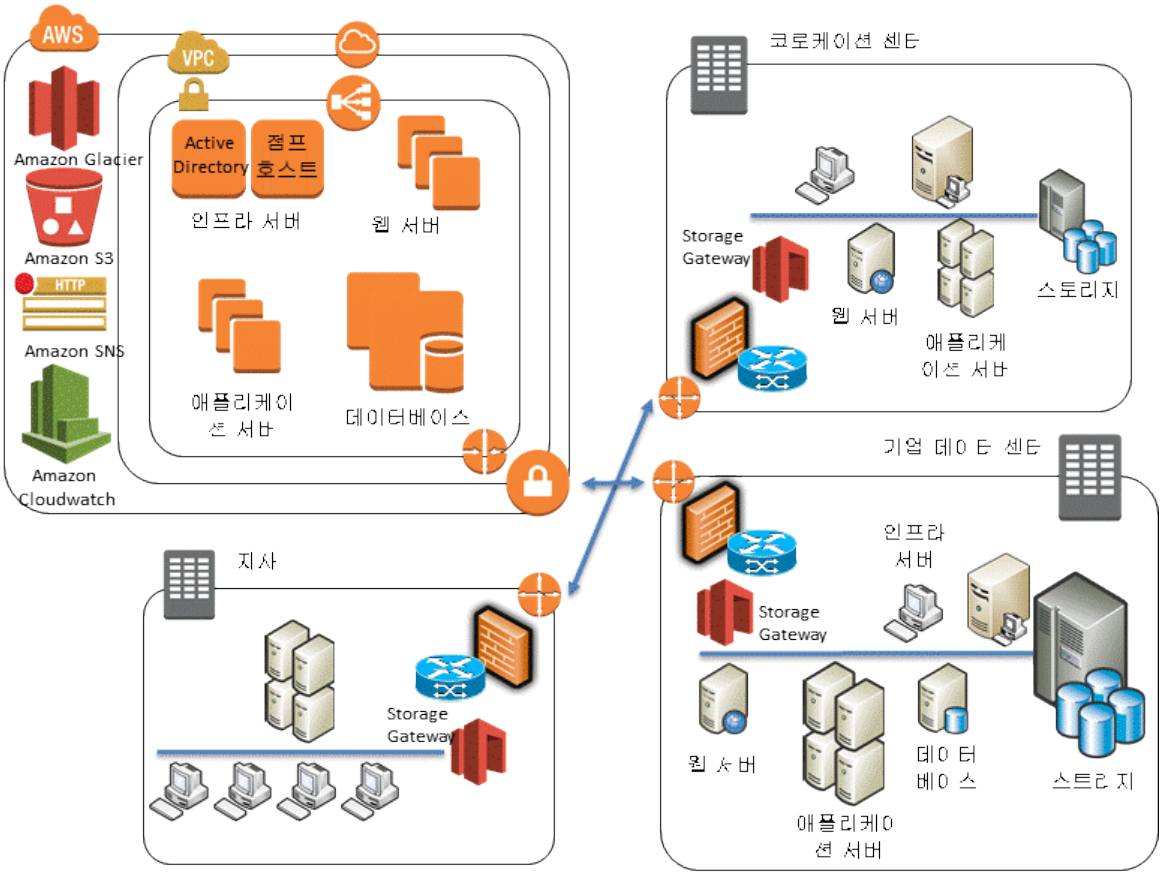

그림 9: 하이브리드 시나리오의 게이트웨이 사용

하이브리드 시나리오 예

Amazon EC2 인스턴스, 독립 서버, 가상 머신 및 데이터베이스의 백업 환경을 관리한다고 가정하겠습니다. 이 환경의 서버 수는 1,000개이며, 운영 체제, 파일 데이터, 가상 머신 이미지 및 데이터베이스를 백업합니다. 백업할 데이터베이스(MySQL, Microsoft SQL Server 및 Oracle)는 20개입니다.

백업 소프트웨어에는 운영 체제, 가상 머신 이미지, 데이터 볼륨, SQL Server 데이터베이스 및 Oracle 데이터베이스(RMAN 사용)를 백업할 수 있는 에이전트가 구성되어 있습니다. MySQL 같은 애플리케이션은 백업 소프트웨어에 에이전트가 없어서 mysqldump 클라이언트 유틸리티를 사용해 데이터베이스 덤프 파일을 디스크에 생성합니다. 그러면 디스크에서 표준 백업 에이전트가 데이터를 보호할 수 있습니다.

타사의 백업 소프트웨어라면 이러한 환경 보호를 위해 백업, 아카이브 및 복구 작업을 제어할 수 있는 글로벌 카탈로그 서버나 마스터 서버 1개와 디스크 기반 스토리지, LTO(Linear Tape-Open) 테이프 장치 및 AWS 스토리지 서버에 연결되는 미디어 서버 다수로 구성될 가능성이 높습니다.

AWS 스토리지 서비스를 이용해 백업 솔루션을 강화할 수 있는 가장 간단한 방법은 Amazon S3 또는 Amazon Glacier를 지원하는 백업 공급업체의 옵션을 이용하는 것입니다. 따라서 먼저 공급업체와 논의하여 통합 및 커넥터 옵션에 대해 알아보는 것이 좋습니다. AWS와 협력하는 백업 소프트웨어 공급업체의 목록을 보려면 파트너 디렉터리[15]를 참조하십시오.

기존 백업 소프트웨어가 기본적으로 백업 또는 아카이브용 클라우드 스토리지를 지원하지 않는 경우에는 브릿지 같이 백업 소프트웨어와 Amazon S3 또는 Amazon Glacier를 연결해주는 스토리지 게이트웨이 장치를 사용하면 됩니다.

타사의 게이트웨이 솔루션은 매우 다양합니다. 하지만 iSCSI 기반 볼륨과 가상 테이프 라이브러리(VTL) 같은 범용 기술을 사용하는 AWS Storage Gateway 가상 어플라이언스 역시 이 둘을 서로 연결하는 데 효과적입니다. 이 구성에는 지원되는 하이퍼바이저(VMware 또는 Microsoft Hyper-V)와 어플라이언스 호스팅용 로컬 스토리지가 필요합니다.

[15] http://www.aws-partner-directory.com/PartnerDirectory/PartnerSearch?type=ISV

AWS를 이용한 데이터 아카이브

규정 준수 또는 기업의 이유로 데이터를 보존해야 하는 경우 아카이브를 사용할 수 있습니다. 일반적으로 프로덕션 데이터 복사본을 단시간 저장하여 데이터 손상 또는 손실 시 복구하는 백업과 달리 아카이브는 보존 정책이 종료될 때까지 모든 데이터 복사본을 유지합니다.

우수한 아카이브가 되기 위한 기준은 다음과 같습니다.

- 장기 무결성을 위한 데이터 내구성
- 데이터 보안
- 복구 용이성
- 저렴한 비용

불변 데이터 스토어가 또 하나의 규제 또는 규정 준수 요건이 될 수 있습니다.

Amazon Glacier는 저렴한 비용, 기본적인 저장 데이터 암호화, 99.999999999%의 내구성, 그리고 무제한 용량을 지원하는 아카이브를 제공합니다.

Amazon S3 Standard - Infrequent Access는 빠른 데이터 검색이 필요한 사용 사례에 적합합니다. 반면 Amazon Glacier는 데이터 액세스 횟수가 적고 검색하는 데 2~3시간까지 허용되는 사용 사례에 적합합니다.

객체는 S3의 수명 주기 규칙이나 Amazon Glacier API를 통해 Amazon Glacier로 티어링(tiering)할 수 있습니다. Amazon Glacier 저장소 잠금 기능으로 각 Amazon Glacier 저장소마다 저장소 잠금 정책에 따라 규정 준수 제어 항목을 쉽게 배포하고 적용할 수 있습니다. 저장소 잠금 정책에서는 “write once, read many”(WORM) 같은 제어 항목을 지정하여 앞으로 편집하지 못하도록 정책을 잠글 수도 있습니다. 자세한 내용은 Amazon Glacier를 참조하십시오.

AWS의 백업 데이터 보안

데이터 보안은 누구나 갖는 공통 관심사입니다. AWS는 보안을 매우 중요하게 생각합니다. 그 이유는 보안이야 말로 모든 AWS 서비스의 토대이기 때문입니다. Amazon S3 같은 스토리지 서비스는 저장 및 전송 중인 데이터에 대한 강력한 액세스 제어 및 암호화 기능을 지원합니다. 모든 Amazon S3 및 Amazon Glacier API 엔드포인트는 전송 데이터에 대한 SSL 암호화를 지원합니다. Amazon Glacier는 기본적으로 모든 저장 데이터를 암호화합니다. Amazon S3에서는 고객이 AWS에게 암호화 키 관리를 허용하거나, 객체 업로드 시 자신의 키를 제공하거나, 암호화 키에 AWS 키 관리 서비스(AWS KMS)[16]를 통합하여 저장 객체를 서버에서 암호화하도록 선택할 수 있습니다. 그 밖에 고객이 AWS에 데이터를 업로드하기 전에 항상 암호화하는 방법도 있습니다. 자세한 내용은 Amazon Web Services: 보안 프로세스 개요를 참조하십시오.

결론

Gartner는 이미 AWS를 퍼블릭 클라우드 스토리지 서비스의 리더로 선정한 바 있습니다[17]. AWS는 기업이 차세대 백업 환경인 클라우드 기반 플랫폼으로 워크로드를 마이그레이션하는 데 도움이 될 수 있는 모든 준비를 마쳤습니다. 또한 비용 효율적이고 확장 가능한 솔루션을 제공하여 기업이 백업과 아카이브 요건의 균형을 유지할 수 있도록 도울 것입니다. AWS 서비스는 오늘날 많이 사용되는 기술이 효과적으로 통합되어 있습니다.

[16] http://docs.aws.amazon.com/AmazonS3/latest/dev/UsingKMSEncryption.html

[17] http://www.gartner.com/technology/reprints.do?id=1-1WWKTQ3&ct=140709&st=sb

기고자

다음은 이 백서의 작성에 도움을 준 개인입니다.

- Pawan Agnihotri, Amazon Web Services의 솔루션 아키텍트
- Lee Kear, Amazon Web Services의 솔루션 아키텍트
- Peter Levett, Amazon Web Services의 솔루션 아키텍트

문서 수정

2016년 5월 업데이트됨