



# AWS 스터디

## 스터디 문제풀이(3/28)

Amazon S3 버킷에 정적 사진을 저장하는 웹사이트를 만들고 있습니다. 이 회사의 목표는 향후 모든 요청에 대한 지연 시간과 비용을 줄이는 것입니다. 솔루션 설계자는 서비스 구성을 어떻게 제안해야 할까요?

- A. Amazon S3 앞에 NAT 서버를 배치합니다.
- B. Amazon S3 앞에 Amazon CloudFront를 배포합니다.
- C. Amazon S3 앞에 네트워크 로드 밸런서를 배포합니다.
- D. 웹 사이트의 용량을 자동으로 조정하도록 자동 확장을 구성합니다.

제출 답안 : B. CloudFront를 배포한다 (정답)

풀이 : 참고링크

콘텐츠 배포 네트워크(CDN) 서비스가 등장하기 전에는 특정 지역에 서버를 일일이 두고 서비스를 해야 했음.  
AWS S3와 AWS CloudFront를 이용하여 정적 콘텐츠를 대규모로 저장하고 보호, 전송할 수 있음

AWS S3 : 정적 콘텐츠를 “버킷(Bucket)”에 무제한으로 저장하고 손쉽게 확장이 가능한 고가용성 데이터 스토리지를 저렴한 비용으로 제공

- > S3 버킷은 자동으로 확장 및 축소되어 스토리지 공간을 사전에 특정할 필요가 없음
- > 애플리케이션에 서버가 필요하더라도 정적 콘텐츠에 대한 요청을 처리할 필요가 없어 서버 크기 줄일 수 있음

AWS CloudFront : 전 세계의 정적/동적 웹 콘텐츠, 비디오 스트림 및 API를 안전하게 대규모로 전송할 수 있는 콘텐츠 전송 네트워크(CDN) 서비스

- > 엣지 로케이션을 통해 최종 사용자가 위치한 곳에 더 가깝게 콘텐츠를 전송가능(성능 향상)
- > 사용자가 요청하는 파일이 CloudFront에 캐싱되어 있으면 빠르게 응답 / 파일이 없으면 오리진 서버(버킷 등)에서 가져와서 캐싱한 뒤 다음엔 캐시 데이터로 빠르게 응답 (캐시 히트)
- > CloudFront는 캐싱 파일을 쓰므로 S3의 부하를 줄이고 비용도 줄일 수 있음(이미 캐싱된 파일 사본으로 응답하므로 S3 -> CloudFront로 전송되는 파일이 없기때문)

결론 : 콘텐츠를 S3에서 바로 불러오는 것보다 CloudFront를 통해 사용자에게 제일 가까운 엣지 로케이션에 콘텐츠의 캐싱 파일을 만든 뒤 이를 통해 불러오는 것이 훨씬 빠르고 저

회사에 데이터를 수집하여 온프레미스 NFS 서버에 저장하는 온프레미스 애플리케이션이 있습니다.

이 회사는 최근 10Gbps AWS Direct Connect 연결을 설정했습니다.

하지만 회사의 레거시 스토리지 용량이 부족합니다.

회사는 온프레미스 애플리케이션의 데이터에 대한 저지연 액세스를 유지하면서 애플리케이션 데이터를 온프레미스에서 AWS 클라우드로 마이그레이션 해야합니다.

이러한 요구사항을 충족하기 위한 조치는 무엇인가요?

- A. 애플리케이션 데이터용 AWS Storage Gateway를 배포하고 파일 게이트웨이를 사용하여 데이터를 Amazon S3에 저장합니다.  
NFS를 사용해 온프레미스 애플리케이션 서버를 파일 게이트웨이에 연결합니다.

- B. Amazon Elastic File System (Amazon EFS) 파일 시스템을 NFS 서버에 연결하고 애플리케이션 데이터를 EFS 파일 시스템에 복사합니다.  
그런 다음 온프레미스 애플리케이션을 Amazon EFS에 연결합니다.

- C. AWS Storage Gateway를 볼륨 게이트웨이를 구성합니다.  
NFS 서버 및 Amazon Elastic Block Store (Amazon EBS) 스냅샷을 통해 온프레미스 애플리케이션에서 애플리케이션 데이터를 사용할 수 있도록 합니다.

- D. NFS서버를 소스 위치로 사용하고 Amazon Elastic File System(Amazon EFS) 파일 시스템을 애플리케이션 데이터 전송 대상으로 사용하여 AWS DataSync 에이전트를 생성합니다.  
그 후 온프레미스 애플리케이션을 EFS 파일 시스템에 연결합니다.

제출 답안 : A. AWS Storage Gateway를 배포하고 - (정답)

풀이 : 참고링크

AWS 하이브리드 클라우드 스토리지 아키텍처는 온프레미스 애플리케이션과 시스템을 클라우드 스토리지로 연결해서 비용을 절감하고, 관리 부담을 최소화한다

AWS Storage Gateway(참고) : 무제한의 클라우드 스토리지 액세스를 온프레미스에 제공하는 하이브리드 클라우드 스토리지 서비스 세트  
온프레미스 소프트웨어 어플라이언스를 클라우드 기반 스토리지에 연결

1) File Gateway(참고) : Amazon S3 File Gateway를 사용하여 Amazon S3 클라우드 스토리지에 객체로 저장 /  
Amazon FSx File Gateway를 사용하여 온프레미스에서 짧은 대기 시간으로 완전 관리형 파일 공유에 액세스

AWS Direct Connect : 사용자의 네트워크와 AWS Direct Connect 위치 중 하나 사이에 전용 네트워크 연결을 설정

Amazon Elastic File System(참고) : 파일을 추가하고 제거할 때 자동으로 확장되고 축소 / 한 번만 설정하면 되는 단순한 탄력적 서버리스 파일 시스템(관리 / 프로비저닝

AWS DataSync : 온프레미스와 AWS 스토리지 서비스 사이 데이터 마이그레이션 툴

비즈니스에는 Amazon RDS for MySQL을 데이터베이스로 사용하는 여러 앱이 있습니다.  
최근에 조직은 새로운 사용자 지정 보고 응용 프로그램이 데이터베이스의 쿼리 수를 늘렸다는 것을 깨달았습니다.  
이로 인해 성능이 저하되었습니다.  
솔루션 설계자는 가능한 가장 적은 수의 애플리케이션 수정으로 이 문제를 어떻게 해결할 수 있습니까?

- A. 다중 AZ를 사용하여 보조 DB 인스턴스를 추가합니다.
- B. Amazon RDS에서 읽기 복제본 및 다중 AZ를 설정합니다.
- C. Amazon RDS에 대기 복제본 및 다중 AZ를 설정합니다.
- D. Amazon RDS에서 캐싱을 사용하여 전체 성능을 개선합니다.

제출 답안 : D. Amazon RDS 캐싱을 사용 (오답)

정답 : B. Amazon RDS 읽기 복제본 및 다중 AZ

풀이 : 참고링크

Amazon RDS 다중 AZ 배포 : 자동으로 프라이머리 DB 인스턴스를 생성 + 다른 AZ의 인스턴스에 데이터를 복제  
장애를 감지하면 수동 개입 없이 자동으로 대기 인스턴스로 장애 조치

읽기 가능한 대기가 2개인 Amazon RDS 다중 AZ

- > 쿼리를 쓰기 서버와 적절한 읽기 전용 복제본 대기 인스턴스로 경로 지정하여 성능과 확장성을 극대화
- > 최대 2배 빠른 트랜잭션 커밋 대기 시간 확보
- > 2개의 읽기 가능한 대기 인스턴스에 트래픽을 분산하여 읽기 확장성을 확보'

결론 : 여러 AZ에 읽기 가능한 복제본을 준비하여 대기 시간을 더 빠르게 하고 트래픽 분산까지 달성

각각의 자원들을 콘솔로 제어하는 대신 코드로 작성해서 인프라를 구축하는 서비스는?

- A. AWS Elastic Beanstalk
- B. AWS CloudFormation
- C. AWS Elastic Container Service (ECS)
- D. AWS Cloud Front
- E. AWS Outpost

제출 답안 : B. CloudFormation (정답)

AWS Elastic Beanstalk : 애플리케이션을 업로드하기만 하면 용량 프로비저닝, 로드 밸런싱, 조정, 애플리케이션 상태 모니터링에 대한 세부 정보를 자동으로 처리  
AWS CloudFormation : 인프라를 코드로 처리하여 AWS 및 서드 파티 리소스를 모델링, 프로비저닝 및 관리  
Amazon Elastic Container Service : 컨테이너화된 애플리케이션의 손쉬운 배포, 관리 및 크기 조정을 지원하는 완전관리형 컨테이너 오케스트레이션 서비스  
AWS CloudFront : 전 세계의 정적/동적 웹 콘텐츠, 비디오 스트림 및 API를 안전하게 대규모로 전송할 수 있는 콘텐츠 전송 네트워크(CDN) 서비스  
AWS Outpost : 온프레미스에서 AWS 인프라 및 서비스 실행

## 스터디 문제풀이(3/29)

AWS 클라우드에서 기업은 다계층 전자 상거래 웹 애플리케이션을 운영하고 있습니다. 애플리케이션은 Amazon RDS MySQL 다중 AZ 데이터베이스에 연결된 Amazon EC2 인스턴스에서 호스팅됩니다. Amazon RDS는 Amazon Elastic Block Store(Amazon EBS)의 범용 SSD(gp2) 볼륨에 최신 세대 인스턴스와 2,000GB의 스토리지로 설정됩니다. 수요가 많은 순간에는 데이터베이스 성능이 애플리케이션에 영향을 미칩니다. Amazon CloudWatch Logs의 로그를 연구한 후 데이터베이스 관리자는 읽기 및 쓰기 IOPS 수가 6,000을 초과하면 애플리케이션의 성능이 지속적으로 떨어지는 것을 발견했습니다.

솔루션 설계자는 애플리케이션의 성능을 최적화하기 위해 무엇을 해야 할까요?

- A. 볼륨을 마그네틱 볼륨으로 교체합니다.
- B. gp2 볼륨의 IOPS 수를 늘립니다.
- C. 볼륨을 프로비저닝된 IOPS(PiOPS) 볼륨으로 교체합니다.
- D. 2,000GB gp2 볼륨을 2개의 1,000GBgp2 볼륨으로 교체합니다.

제출 답안 : C. 볼륨을 프로비저닝된 IOPS 볼륨으로 교체 (정답)

풀이 :

Amazon EBS (EBS, 볼륨 유형) : 내구성이 있는 블록 수준 스토리지 디바이스 / 인스턴스의 수명에 관계없이 유지되는 오프 인스턴스 스토리지 EBS의 SSD 유형  
1) 범용 SSD  
-> 99.8%~99.9% 내구성  
-> 지연 시간이 짧은 대화형 앱  
2) Provisioned IOPS SSD  
-> 99.999% 내구성  
-> 지속적인 IOPS 성능 또는 16,000 IOPS 이상이 필요한 워크로드

Amazon CloudWatch : 하이브리드 및 온프레미스 애플리케이션 및 인프라 리소스에 대한 데이터와 실행 가능한 인사이트를 제공하는 모니터링 및 관리 서비스 전체 스택 모니터링, 로그 등 이벤트 데이터를 사용하여 자동화된 조치하여 MTTR(Mean Time to Resolution)을 단축

결론 : 더 높은 IOPS를 감당하기 위해서는 기존 범용 SSD에서 프로비저닝된 SSD 볼륨으로 교체하면 됨

기업은 전 세계적으로 300개 이상의 웹사이트와 앱을 호스팅합니다. 조직은 매일 30TB 이상의 클릭스트림 데이터를 분석할 수 있는 플랫폼을 원합니다.

솔루션 설계자는 전송 및 처리 중에 클릭스트림 데이터로 무엇을 해야 할까요?

- A. AWS Data Pipeline을 설계하여 데이터를 Amazon S3 버킷에 보관하고 데이터로 Amazon EMR 클러스터를 실행하여 분석을 생성합니다.
- B. Amazon EC2 인스턴스의 Auto Scaling 그룹을 생성하여 데이터를 처리하고 Amazon Redshift가 분석에 사용할 수 있도록 Amazon S3 데이터 레이크로 보냅니다.
- C. 데이터를 Amazon CloudFront에 캐시합니다. Amazon S3 버킷에 데이터를 저장합니다. S3 버킷에 객체가 추가되면 AWS Lambda 함수를 실행하여 분석할 데이터를 처리합니다.
- D. Amazon Kinesis Data Streams에서 데이터를 수집합니다. Amazon Kinesis Data Firehose를 사용하여 Amazon S3 데이터 레이크로 데이터를 전송합니다. 분석을 위해 Amazon Redshift에 데이터를 로드합니다.

제출 답안 : D (정답)

풀이 :

300개 이상의 웹사이트와 앱을 호스팅하며 30TB의 클릭스트림 데이터를 분석하기 위해서는 큰 데이터를 관리할 수 있는 서비스가 필요

AWS RedShift (참고) : 페타바이트급 데이터 웨어하우스 서비스  
-> SQL을 사용하여 여러 데이터 웨어하우스, 운영 데이터베이스 및 데이터 레이크에서 정형 데이터 및 반정형 데이터를 분석

Amazon Kinesis Data Streams (참고) : 모든 규모의 데이터 스트림을 쉽게 캡처, 처리 및 저장할 수 있는 서버리스 스트리밍 데이터 서비스  
-> 클릭스트림과 같은 빈도가 높은 이벤트 데이터용 애플리케이션을 구축  
-> AWS Lambda / Amazon Kinesis Data Analytics 과 연동하여 데이터 인사이트 확보

Amazon Kinesis Data Analytics(참고) : 실시간 분석을 포함한 다양한 사용 사례에 특화된 Java, Scala, Python 및 SQL 기반의 유연한 API를 제공  
Amazon Kinesis Data Firehose(참고) : 스트리밍 데이터를 안정적으로 캡처하고 변환하여 데이터 레이크, 데이터 스토어, 분석 서비스에 전달하는 추출, 변환 로드 서비스

AWS Data Pipeline(참고) : 온프레미스 데이터 소스뿐 아니라 여러 AWS 컴퓨팅 및 스토리지 서비스 간에 데이터를 안정적으로 처리하고 지정된 간격으로 이동할 수 있게 지원

\*\* 궁금한 점  
30TB의 데이터를 분석하는데는 빅데이터 관련 서비스인 EMR이 더 낫지 않은지?  
EMR이 페타바이트 급 데이터를 분석하는 툴이라서 TB 정도 데이터는 취급을 안하는건지? (참고: EMR)

## 스터디 문제풀이(3/30)

솔루션 설계자는 새로운 정적 웹사이트의 구현을 설계하는 책임을 맡습니다. 솔루션은 비용 효율적이어야 하며 최소 99%의 가용성을 유지해야 합니다.

어떤 솔루션이 이러한 기준을 충족합니까?

- A. 버전 관리가 비활성화된 한 AWS 리전의 Amazon S3 버킷에 애플리케이션을 배포합니다.
- B. 2개의 AWS 리전과 2개의 가용 영역에서 실행되는 Amazon EC2 인스턴스에 애플리케이션을 배포합니다.
- C. 버전 관리 및 교차 리전 복제가 활성화된 Amazon S3 버킷에 애플리케이션을 배포합니다.
- D. 하나의 AWS 리전과 하나의 가용 영역에서 실행되는 Amazon EC2 인스턴스에 애플리케이션을 배포합니다.

제출 답안 : C(오답) -> 정답 A

풀이 :  
S3에서 정적 웹 사이트를 호스팅 할 수 있다.  
버전 관리가 비활성화 되어있어야 한다.

비즈니스에서 애플리케이션을 개발 중입니다. 이 프로그램은 Amazon API Gateway를 통해 데이터를 수신하고 AWS Lambda 함수를 사용하여 Amazon Aurora PostgreSQL 데이터베이스에 저장합니다.

개념 증명 단계에서 회사는 데이터베이스에 로드해야 하는 많은 양의 데이터를 관리하기 위해 Lambda 할당량을 대폭 늘려야 합니다. 솔루션 설계자는 확장성을 최대화하고 설정 노력을 줄이는 새로운 설계에 대한 권장 사항을 제공해야 합니다.

어떤 솔루션이 이러한 기준을 충족할까요?

- A. Lambda 함수 코드를 Amazon EC2 인스턴스에서 실행되는 Apache Tomcat 코드로 리팩터링합니다. 네이티브 JDBC(Java Database Connectivity) 드라이버를 사용하여 데이터베이스를 연결합니다.
- B. 플랫폼을 Aurora에서 Amazon DynamoDB로 변경합니다. DynamoDB Accelerator(DAX) 클러스터를 프로비저닝합니다. DAX 클라이언트 SDK를 사용하여 DAX 클러스터에서 기존 DynamoDB API 호출을 가리킵니다.
- C. 두 개의 Lambda 함수를 설정합니다. 정보를 수신할 하나의 기능을 구성하십시오. 정보를 데이터베이스에 로드하도록 다른 기능을 구성하십시오. Amazon Simple Notification Service(Amazon SNS)를 사용하여 Lambda 함수를 통합합니다.
- D. 두 개의 Lambda 함수를 설정합니다. 정보를 수신할 하나의 기능을 구성하십시오. 정보를 데이터베이스에 로드하도록 다른 기능을 구성하십시오. Amazon Simple Queue Service(Amazon SQS) 대기열을 사용하여 Lambda 함수를 통합합니다.

제출 답안 : D(오답) -> 정답 B

풀이 : (참조)  
DynamoDB : 모든 규모에서 빠르고 유연한 비관계형 데이터베이스 페타바이트까지 핸들 가능  
Amazon Aurora : 클라우드용으로 구축된 MySQL 및 PostgreSQL 호환 관계형 데이터베이스  
Amazon Simple Queue Service(SQS): 분산 시스템 및 서버리스 애플리케이션을 쉽게 분리하고 확장할 수 있도록 지원하는 완전관리형 메시지 대기열 서비스  
Amazon Simple Notification Service(SNS) : 클라우드에서 손쉽게 알림을 설정, 운영 및 전송할 수 있도록 하는 웹 서비스

결론: 대규모 데이터에 적합한 것은 다이나모DB