

AWS_210805

데이터베이스 호스팅

Amazon EC2 대 Amazon RDS 비교

-RDS적합 = 비즈니스 본연의 가치에 집중하는 경우, 데이터베이스 관리 업무를 원치 않는 경우, 회사내 데이터베이스 관리 역량이 충분치 않은 경, 백업 및 복원 자동화를 원하는 경우.

-EC2적합 ▪ 데이터베이스 인스턴스에 대한 완전한 통제권을 원하는 경우, 백업, 복제, 클러스터링 등에 대한 완전한 통제권을 원하는 경우, 기업에서 요구하는 용량 및 성능이 Amazon RDS의 제공 수준을 넘어서는 경우.

Amazon RDS에서의 확장성 구현

- RDS에서 실행되는 데이터베이스의 확장성을 관리할 수 있는 방법 존재.
- 사용자는 필요에따라 스케일업이나 스케일다운이 필요.

인스턴스 타입 변경

- 확장성 조절을 위한 가장 간단한 방법 = 데이터베이스 인스턴스타입 변경.
- 사용자는 사용자는 Amazon RDS가 지원하는 어떤 인스턴스 클래스로도 자유롭게 스케일업 스케일다운 가능.
- 사용자는 AWS CLI 및 AWS API를 이용해 확장성 조절가능.

읽기 사본, Read Replica 활용

- 마스터 데이터베이스의 read-only 복사본으로 마스터 데이터베이스와 동기화 상태를 유지하며, RDS는 RDBMS 엔진에 따라 최대 15개의 읽기 사본을 지닐 수 있음.
- 읽기 사본은 read-only 쿼리의 부담을 줄여주며, 결과적으로 마스터 데이터베이스의 워크로드를 감소 시킴.
- 읽기 사본을 통해 read-only 트래픽의 부담을 줄일 수 있고, 마스터 데이터베이스는 중요한 트랜잭션 쿼리에 집중가능
- 사용자가 다른 국가에 있는 경우 해당 국가의 리전에 읽기 사본을 뒤 read-only 트래픽 부담을 줄일 수 있음.
- 사용자는 동일 리전의 다른 AZ에 읽기 사본을 생성하거나, 다른 리전에 읽기 사본을 생성해 크로스 리전 읽기 사본(cross-regional read replica)을 생성가능.
- 일부 RDBMS 엔진은 크로스 리전 읽기 사본 기능을 지원x

Amazon RDS에서의 보안성 구현

- Amazon RDS 인스턴스는 Amazon VPC 내에서 론칭.
- 데이터베이스는 항상 방화벽으로 보호해야 하므로, 데이터베이스를 처음부터 프라이빗 서브넷 내에 생성하는 것이 좋음.
- VPC 내에 데이터베이스를 생성하면 데이터베이스에 대한 유저와 애플리케이션의 접근을 효과적으로 관리가능.
- 데이터베이스에 대한 접근을 시큐리티 그룹의 프로토콜, 포트 범위, 트래픽 소스 등으로 관리할 수 있으며, 관리자는 특정 IP 주소, 다수의 IP 주소를 포함한 CIDR 블록의 트래픽을 차단하거나 다른 시큐리티 그룹의 트래픽도 차단할 수 있고, 오직 관리자가 허용한 인스턴스로부터의 트래픽만 유입하도록 설정가능.

Amazon RDS에서의 데이터 암호화

- 암호화는 다수의 기업 고객에게 중요한 요소이며, Amazon RDS는 데이터베이스 암호화 기능을 제공. 다수의 기업은 강행 규정으로 데이터베이스에 대한 암호화를 해야 하며, RDS는 저장 상태(data at rest)의 데이터에 대한 암호화를 제공함.
- RDS의 암호화된 인스턴스는 데이터 보안을 위한 추가적인 레이어를 제공해 비인가 접근차단.
- Amazon RDS 암호화는 클라우드에 배포된 애플리케이션의 데이터 보안 수준을 높이며, 저장 상태 데이터의 암호화 법규에 대한 준수 의무를 효과적으로 지킬 수 있음.

Amazon RDS에서의 데이터 암호화

- AWS RDS의 데이터베이스 암호화에는 다음 항목에 대한 암호화가 포함됨.
- Amazon RDS 인스턴스의 데이터 암호화에는 산업 표준인 AES-256 암호화 알고리즘이 사용되며, 일단 데이터 암호화를 진행한 뒤 RDS는 자동으로 복호화 과정을 처리.
- RDS 인스턴스를 생성하고 암호화 기능을 활성화하면, AWS KMS(Key Management Service) 내에 RDS를 위한 기본 키가 생성돼 Amazon RDS 인스턴스의 데이터 암호화 및 복호화에 사용할 수 있고, 사용자는 자신의 계정을 통해 키 관리가능.

8.모니터링

- 아마존은 RDS 에서 실행되는 데이터베이스를 모니터링 할 수 있는 다양한 방법을 제공.
- 표준 모니터링 : RDBMS 엔진에 따라 15-18 개의 성능 지표를 모니터링할 수 있으며, 주요 지표는 CPU 이용률, 스토리지, 메모리, 스왑 사용량, 데이터베이스 연결, I/O (읽기 및 쓰기), 지연, 처리성능, 읽기 사본 처리 지연 등이 있음.

- 강화 모니터링: 좀 더 세분화된 성능 지표가 필요한 때 선택하는 옵션으로 표준 모니터링의 성능지표 외 37개의 지표를 추가해, 총 50개의 성능 지표를 제공할 수 있음.
- 가용성, 백업, 환경 설정 변경, 생성, 삭제, 실패, 장애 대응, 유지 보수, 복구, (스냅샷 등을 이용한) 복원 등 17개 카테고리의 이벤트가 발생하면, 사용자에게 noti피케이션 전송.
- RDS는 모니터링된 모든 성능 지표를 Amazon CloudWatch로 전송하고, 사용자는 이를 RDS 콘솔, CloudWatch 콘솔 또는 CloudWatch API를 통해 확인할 수 있음. 퍼포먼스 인사이트 대시보드를 통해 데이터베이스 성능 정보를 제공하며, 사용자는 이를 통해 성능 이슈 해결가능.

9.Amazon Aurora

- Amazon Aurora는 클라우드에 최적화된 MySQL 및 PostgreSQL 호환 관계형 데이터베이스.
- 상용 데이터베이스로서의 성능 및 가용성을 제공하고, 오픈소스 데이터베이스로서의 간소함과 비용 효율성을 제공.
- Amazon Aurora는 완전 분산형 및 자가 진단형 스토리지 시스템을 통해 성능 및 신뢰성을 보장하고, 탄력성 및 AWS 클라우드관리 역량을 바탕으로한 가용성을 제공.
- Amazon Aurora는 MySQL 호환 타입 및 PostgreSQL 호환 타입 등 두 가지 타입이 있음.

10. Amazon Redshift

- Amazon Redshift는 관리형 데이터 웨어하우스 솔루션.
- 비즈니스 인텔리전스 활동에 맞게 설계되며, 기업 조직에서 요구하는 데이터 처리 및 분석 성능을 강화하는데 주로 사용됨.
- 데이터 웨어하우스는 OLTP 시스템과는 확연히 구분되는 차이점이 있으며, 데이터 웨어하우스를 이용해 트랜잭션 워크로드와 분석 워크로드를 분리해 처리 가능.
- 데이터 웨어하우스 환경에는 데이터 추출, 변환, 로딩의 전 과정을 일컫는 ETL 솔루션, 통계 분석, 리포트 작성, 데이터 마이닝, 클라이언트 분석 도구 등이 포함되고 데이터 수집 행동 가능 정보로의 변환, 비즈니스 의사 결정자에게 제공하기 위한 애플리케이션이 포함됨.

■ Amazon Redshift의 장점

-신속성 : Redshift는 칼럼형 스토리지로 높은 쿼리 성능제공. 쿼리는 병렬적으로 여러 노드에서 동시에 처리되므로 쿼리 결과가 신속하게 제공되고 IO 효율성 또한 높아짐.

-저렴함 : Redshift는 데이터 웨어하우스 솔루션 시장에서 가장 낮은 수준의 비용으로 이용할 수 있으며, 테라 바이트당 \$1,000 수준으로, 다른 경쟁 제품군의 1/10 가격에 제공되고있음.

-우수한 압축성 • Redshift는 세 배에서 네 배 정도 높은 압축 기능을 제공하므로 데이터 전송 등 업무에서 비용을 더욱 절감할 수 있다.

-관리형 서비스 : Redshift는 관리형 서비스로서 사용자는 클러스터, 네트워킹, 운영체제 등을 관리할 필요가 없고, Amazon이 운영체제 패치, 업그레이드, 백업, 복원 업무를 처리함. 또한 데이터 웨어하우스 관리, 모니터링, 확장성 등 관리 업무 대부분이 자동화되어있음.

-확장성 : Redshift는 분산형 병렬식 아키텍처를 이용해 처리 성능에 대한 요구사항을 수평적 확장 방식으로 제공.

-보안성 : Redshift는 저장 중인 데이터 및 전송 중인 데이터 모두에 대한 암호화 기능을 제공하며, VPC 내에 클러스터를 생성해 다른 리소스와 격리할 수 있고, AWS KMS(Key Management System)와 HSM(Hardware Security Modules)을 이용해 키를 관리함.

-보안 : Redshift는 RDBMS 중 하나이므로, 슈퍼유저(superuser) 또는 유저 퍼미션을 지닌 데이터베이스 유저(database users)를 생성해야함.

*이때 마스터 유저네임은 데이터베이스에 로그인할 수 있는 슈퍼유저를 가리키며, 마스터 유저 계정은 삭제 불가. Redshift 서비스를 사용하기 위해 IAM 유저를 그룹 등을 포함한 IAM 정책을 생성할 수 있으며, Redshift 관리형 정책을 통해 Redshift를 어드민 또는 리드 온리 권한으로 접근할 수 있고, 사용자가 직접 커스텀 정책을 만들어서 세분화된 권한을 부여가능.

-백업 및 복구 클러스터의 스냅샷 형태로 자동으로 백업하여 이를 Amazon S3에 저장함.

(8시간마다 또는 5GB의 블록이 추가될 때마다)

-자동화된 백업은 RPO(Recovery Point Objective) 즉 복구 시점 목표에 따라 스냅샷을 생성함.

-스냅샷을 자동으로 다른 리전에 복사가능.

-스냅샷을 통해 전체 데이터베이스를 복원하는 것 또한 가능.

11.Amazon DynamoDB

-확장성 및 개발의 신속성을 제공하는 고성능의 완전 관리형 NoSQL 데이터베이스.

-DynamoDB는 완전 관리형 클라우드 NoSQL 데이터베이스 서비스로 테이블 생성, 목표 처리 성능 설정, 저장 상태의 데이터 관리 등을 지원.

-사용자는 테이블 생성 시 애플리케이션에 요구되는 처리 용량만 지정하면됨.

-DynamoDB는 확장성을 자동으로 관리하므로 애플리케이션에 대한 처리 성능 요구 수준이 높아지기 전에 스케일업가능.

-DynamoDB의 주요 사용 방식 = 광고 서비스에서 브라우저의 쿠키 상태 정보 수집, 모바일 애플리케이션에서 애플리케이션 데이터, 세션 상태 저장, IoT 기기의 센서 데이터, 게임 애플리케이션 등등 다양하게 활용가능.

Amazon DynamoDB의 장점

-확장성

-오토 스케일링

-관리형 서비스(애플리케이션 동작에 따라 처리 성능을 높이거나 낮춤.)

DynamoDB는 지정 시점 복구(point-in-time recovery) 백업, 테이블 복원 등의 기능도 제공.

-신속성, 일관된 성능

-세분화된 접근 제어

-비용 효율성

-다른 AWS 서비스와의 통합 (AWS Lambda와 통합해 이벤트 트리거(triggers)를 생성가능, 사용자는 이들 트 리거를 이용해 애플리케이션에서 DynamoDB 테이블의 데이터변경가능.

12.Amazon ElastiCache

-클라우드에서 인메모리 캐시를 배포, 운영, 확장하기 위 한 웹 서비스.

-사용자는 불과 몇 분만에 ElastiCache의 API를 이용해 애플리케이션에 인메모리 캐시 레이어를 추가가능.

-ElastiCache는 Amazon EC2, Amazon RDS는 물론, AWS CloudFormation, AWS Elastic Beanstalk, AWS OpsWorks와 통합해 사용가능.

-ElastiCache는 클러스터를 지속적으로 모니터링해 워크로드가 정상적으로 실행 되도록 해 개발자가 고부가가치를 제공하는 애플리케이션 업무에 집중할 수 있도록 해줌.

-ElastiCache는 샤딩 기법을 통해 메모리 확장성을 관리.

-ElastiCache가 제공하는 인메모리 캐시는 중요한 데이터를 빠르게 접근할 수 있도록 메모리에 저장해 애플리케이션의 성능을 높이며, 사용자는 이를 이용해 소셜 네트워크, 게임, 미디어 콘텐츠 공유, Q&A 포털 등 서비스의 처리 지연 시간을 현격하게 줄이고 처리 성능을 높이는 것이 가능.