

Bancos de dados da AWS

RDS

O Amazon Relational Database Service (Amazon RDS) é um serviço gerenciado que facilita a configuração, operação e dimensionamento de um banco de dados relacional na nuvem.

Os bancos de dados relacionais são conhecidos como bancos de dados SQL (Structured Query Language).

Os bancos de dados não relacionais são conhecidos como bancos de dados NoSQL.

RDS é um tipo de banco de dados OLTP (Online Transaction Processing).

Recursos e benefícios do RDS:

- Tipo de banco de dados SQL.
- Pode ser usado para realizar consultas e junções complexas.
- Fácil de configurar, altamente disponível, tolerante a falhas e escalável.
- Usado quando os dados são claramente definidos.
- Os casos de uso comuns incluem lojas online e sistemas bancários.

O Amazon RDS oferece suporte aos seguintes mecanismos de banco de dados:

Servidor SQL. Oráculo. Servidor MySQL. PostgreSQL. Aurora. MariaDB.

Aurora é o banco de dados proprietário da Amazon.

O RDS é um serviço totalmente gerenciado e você não tem acesso à instância do EC2 subjacente (sem acesso root).

O serviço RDS inclui o seguinte:

- Segurança e correção das instâncias de banco de dados.
- Backup automatizado para as instâncias de banco de dados.
- Atualizações de software para o mecanismo de banco de dados.

- Dimensionamento fácil para armazenamento e computação.
- Opção Multi-AZ com replicação síncrona.
- Failover automático para a opção Multi-AZ.
- Opção de réplicas de leitura para cargas de trabalho pesadas de leitura.
- Uma instância de banco de dados é um ambiente de banco de dados na nuvem com os recursos de computação e armazenamento que você especifica.

Criptografia:

- Você pode criptografar suas instâncias e snapshots do Amazon RDS em repouso ativando a opção de criptografia para sua instância de banco de dados do Amazon RDS.
- A criptografia em repouso é compatível com todos os tipos de banco de dados e usa o AWS KMS.
- Você não pode criptografar um banco de dados existente, você precisa criar um instantâneo, copiá-lo, criptografar a cópia e construir um banco de dados criptografado a partir do instantâneo.

Grupos de sub-rede de banco de dados:

- Um grupo de sub-redes de banco de dados é uma coleção de sub-redes (geralmente privadas) que você cria em uma VPC e designa para suas instâncias de banco de dados.
- Cada grupo de sub-redes de banco de dados deve ter sub-redes em pelo menos duas zonas de disponibilidade em uma determinada região.
- Recomenda-se configurar um grupo de sub-redes com sub-redes em cada AZ (mesmo para instâncias independentes).

Cobrança da AWS para:

- Horas da instância de banco de dados (horas parciais são cobradas como horas completas).
- Armazenamento GB/mês.
- Solicitações de E/S/mês – para armazenamento magnético.
- IOPS provisionado/mês – para SSD IOPS provisionado por RDS.
- Transferência de dados de saída.
- Armazenamento de backup (backups de banco de dados e instantâneos manuais).

Escalabilidade:

- Você só pode aumentar o RDS (computação e armazenamento).

- Você não pode diminuir o armazenamento alocado para uma instância do RDS.
- Você pode dimensionar o armazenamento e alterar o tipo de armazenamento para todos os mecanismos de banco de dados, exceto MS SQL.

O RDS fornece multi-AZ para recuperação de desastres que oferece tolerância a falhas em zonas de disponibilidade:

- O RDS Multi-AZ cria uma réplica em outra AZ e replica de forma síncrona para ela (somente DR).
- Há uma opção para escolher multi-AZ durante o assistente de inicialização.
- A AWS recomenda o uso de armazenamento IOPS provisionado para instâncias de banco de dados RDS multi-AZ.
- Cada AZ é executada em sua própria infraestrutura fisicamente distinta e independente e é projetada para ser altamente confiável.
- Você não pode escolher qual AZ na região será escolhida para criar a instância de banco de dados em espera.

Réplicas de leitura – fornecem desempenho aprimorado para leituras:

- As réplicas de leitura são usadas para bancos de dados pesados de leitura e a replicação é assíncrona.
- As réplicas de leitura são para compartilhamento e descarregamento de carga de trabalho.
- As réplicas de leitura fornecem DR somente leitura.
- As réplicas de leitura são criadas a partir de um instantâneo da instância mestre.
- Deve ter backups automatizados habilitados no primário (período de retenção > 0).

DynamoDB

O Amazon DynamoDB é um serviço de banco de dados NoSQL totalmente gerenciado que oferece desempenho rápido e previsível com escalabilidade perfeita.

Recursos e benefícios do Dynamo DB:

- Tipo de banco de dados NoSQL (não relacional).
- Rápido, altamente disponível e totalmente gerenciado.
- Usado quando os dados são fluidos e podem mudar.
- Casos de uso comuns incluem redes sociais e análise da web.

O dimensionamento por botão significa que você pode dimensionar o banco de dados a qualquer momento sem incorrer em tempo de inatividade.

Baseado em SSD e usa indexação limitada em atributos para desempenho.

O DynamoDB é um serviço da Web que usa HTTP sobre SSL (HTTPS) como transporte e JSON como formato de serialização de mensagens.

O Amazon DynamoDB armazena três réplicas distribuídas geograficamente de cada tabela para permitir alta disponibilidade e durabilidade dos dados.

Os dados são replicados de forma síncrona em 3 instalações (AZs) em uma região.

A replicação entre regiões permite replicar entre regiões:

- As tabelas globais do Amazon DynamoDB fornecem uma solução totalmente gerenciada para implantar um banco de dados multirregional e multimestre.
- Ao criar uma tabela global, você especifica as regiões da AWS em que deseja que a tabela esteja disponível.
- O DynamoDB executa todas as tarefas necessárias para criar tabelas idênticas nessas regiões e propagar alterações de dados contínuas para todas elas.

Fornece baixa latência de leitura e gravação.

Aumente ou diminua o armazenamento e a taxa de transferência conforme necessário, sem alterações de código ou tempo de inatividade.

O DynamoDB não tem esquema.

O DynamoDB pode ser usado para armazenar o estado da sessão.

Fornece dois modelos de leitura.

Leituras eventualmente consistentes (Padrão):

- A opção de consistência eventual maximiza sua taxa de transferência de leitura (melhor desempenho de leitura).
- Uma leitura eventualmente consistente pode não refletir os resultados de uma gravação concluída recentemente.
- Consistência em todas as cópias alcançadas em 1 segundo.

Leituras fortemente consistentes:

- Uma leitura fortemente consistente retorna um resultado que reflete todas as gravações que receberam uma resposta bem-sucedida antes da leitura (consistência mais rápida).

O Amazon DynamoDB Accelerator (DAX) é um cache de memória totalmente gerenciado e altamente disponível para o DynamoDB que oferece uma melhoria de desempenho de até 10 vezes – de milissegundos a microssegundos – mesmo em milhões de solicitações por segundo.

RedShift

O Amazon Redshift é um data warehouse rápido e totalmente gerenciado que torna simples e econômica a análise de todos os seus dados usando SQL padrão e ferramentas de Business Intelligence (BI) existentes.

RedShift é um data warehouse baseado em SQL usado para aplicativos de análise .

RedShift é um banco de dados relacional usado para casos de uso de OLAP (Online Analytics Processing).

O RedShift é usado para executar consultas analíticas complexas em petabytes de dados estruturados, usando otimização de consulta sofisticada, armazenamento colunar em discos locais de alto desempenho e execução de consulta maciçamente paralela.

O RedShift é ideal para processar grandes quantidades de dados para inteligência de negócios.

RedShift é 10x mais rápido que um banco de dados SQL tradicional.

RedShift usa armazenamento de dados colunar:

- Os dados são armazenados sequencialmente em colunas em vez de linhas.
- O banco de dados baseado em colunas é ideal para armazenamento e análise de dados.
- Requer menos E/Ss, o que melhora muito o desempenho.

O RedShift fornece compactação avançada:

- Os dados são armazenados sequencialmente em colunas, o que permite um desempenho muito melhor e menos espaço de armazenamento.
- RedShift seleciona automaticamente o esquema de compressão.
- O RedShift usa replicação e backups contínuos para aumentar a disponibilidade e a durabilidade e pode se recuperar automaticamente de falhas de componentes e nós.

O RedShift sempre mantém três cópias de seus dados:

- O original.
- Uma réplica em nós de computação (dentro do cluster).

- Uma cópia de backup no S3.

O RedShift fornece backups contínuos/incrementais:

- Várias cópias em um cluster.
- Backups contínuos e incrementais para S3.
- Backups contínuos e incrementais entre regiões.
- Restauração de streaming.

O RedShift fornece tolerância a falhas para as seguintes falhas:

- Falhas de disco.
- Falhas de nós.
- Falhas de rede.
- Desastres em nível de AZ/região.

ElastiCache

O ElastiCache é um serviço da Web que facilita a implantação e a execução de nós de servidor compatíveis com o protocolo Memcached ou Redis na nuvem.

O cache na memória fornecido pelo ElastiCache pode ser usado para melhorar significativamente a latência e a taxa de transferência para muitas cargas de trabalho de aplicativos de leitura intensa ou cargas de trabalho de computação intensiva.

Melhor para cenários em que a carga do banco de dados é baseada em transações OLAP (Online Analytics Processing).

Os nós do Elaticache EC2 não podem ser acessados pela Internet nem por instâncias do EC2 em outras VPCs.

Também podem ser instâncias sob demanda ou reservadas (mas não instâncias spot).

O Elasticache pode ser usado para armazenar o estado da sessão.

Existem dois tipos de mecanismo ElastiCache:

- Memcached – modelo mais simples, pode executar nós grandes com vários núcleos/threads, pode ser dimensionado para dentro e para fora, pode armazenar em cache objetos como bancos de dados.

- Redis – modelo complexo, suporta criptografia, replicação mestre/escravo, cross AZ (HA), failover automático e backup/restauração.

EMR

O Amazon EMR é um serviço da web que permite que empresas, pesquisadores, analistas de dados e desenvolvedores processem grandes quantidades de dados de maneira fácil e econômica.

O EMR utiliza uma estrutura do Hadoop hospedada em execução no Amazon EC2 e no Amazon S3.

Estrutura gerenciada do Hadoop para processar grandes quantidades de dados.

Também suporta Apache Spark, HBase, Presto e Flink.

Mais comumente usado para análise de log, análise financeira ou atividades de extração, tradução e carregamento (ETL).

Aurora

É um sistema mais rápido que o RDS.

Document DB

Banco de dados não relacional, compatível com MongoDB, banco de dados baseado em documentos.

Glue

- O AWS Glue é um serviço de extração, transformação e carregamento (ETL) totalmente gerenciado que facilita aos clientes a preparação e o carregamento de dados para análise.
- Você pode criar e executar um trabalho ETL com apenas alguns cliques no Console de gerenciamento da AWS.
- Você simplesmente aponta o AWS Glue para seus dados armazenados na AWS, e o AWS Glue descobre seus dados e armazena os metadados associados (por exemplo, definição de tabela e esquema) no Catálogo de dados do AWS Glue.
- Uma vez catalogados, seus dados são imediatamente pesquisáveis, consultáveis e disponíveis para ETL.
- O AWS Glue gera o código para executar suas transformações de dados e processos de carregamento de dados.

Athena

- O Amazon Athena é um serviço de consulta interativa que facilita a análise de dados no Amazon S3 usando SQL padrão.
- O Athena não tem servidor, portanto, não há infraestrutura para gerenciar e você paga apenas pelas consultas executadas.
- Com alguns cliques no Console de gerenciamento da AWS, os clientes podem apontar o Athena para seus dados armazenados no S3 e começar a usar o SQL padrão para executar consultas ad-hoc e obter resultados em segundos.
- Você pode usar o Athena para processar logs, realizar análises ad hoc e executar consultas interativas
- O Athena é dimensionado automaticamente – executando consultas em paralelo – para que os resultados sejam rápidos, mesmo com grandes conjuntos de dados e consultas complexas.