PONTOS MAIS COBRADOS NA PROVA DE SAA-C03

1 ARQUITETURA COM ALTA DISPONIBILIDADE

(serviços Multi-AZ , auto Scaling , Load Balance)

1. **Discussão sobre Arquitetura Clássica**
    - Aplicativos da Web sem estado: horario-brasilia.com
    - Aplicativos da Web com estado: mercado livre,
    - Instanciando aplicativos rapidamente

Aplicativos da Web sem estado: Horario-brasilia.com

- Permite que as pessoas saibam que horas são
- Nenhum banco de dados necessário
- Queremos começar aos poucos e podemos aceitar o tempo de inatividade no início
- Queremos ser capazes de escalar totalmente vertical e horizontalmente sem tempo de inatividade
- Considerações ao construir a arquitetura:
    - IP público vs IP privado no caso de instâncias EC2
    - Elastic IP vs Route53 vs Load Balancers
    - Registros Route53 TTL A vs registros de alias
    - Manter o escalonamento de instância do EC2 manualmente em vez de usar um ASG
    - Implantações Multi AZ para recuperação de desastres
    - Verificações de saúde ELB
    - Regras de grupos de segurança
    - Reserva de capacidade para economia de custos quando possível
    - Estamos considerando os 5 pilares de uma aplicação bem arquitetada: custo, desempenho, confiabilidade, segurança e excelência operacional

Aplicativos da Web com estado:Mercado Livre

- ML: permite que as pessoas comprem roupas online

- Existe um carrinho de compras para cada usuário
- O aplicativo pode ter centenas de usuários usando o aplicativo ao mesmo tempo
- Precisamos escalar, manter a escalabilidade horizontal e manter o aplicativo o mais sem estado possível
- Os usuários não devem perder o carrinho de compras enquanto navegam no site
- Os usuários devem ter seus dados em um banco de dados
- Consideração ao construir a arquitetura:
    - Sessões fixas do ELB
    - Clientes da Web para armazenar cookies e tornar o aplicativo sem estado
    - ID de sessão + cache de sessão que pode ser ElastiCache ou DynamoDB
    - O ElastiCache pode ser usado para armazenar dados em cache do RDS
    - Multi AZ para recuperação de desastres
    - RDS:
        - Podemos usá-lo para armazenar dados do usuário
        - Podemos ter réplicas de leitura para escalar leituras
        - Podemos habilitar o Multi AZ para recuperação de desastres
- Segurança rígida usando grupos de segurança

Instanciando aplicativos rapidamente

- Ao iniciar um aplicativo full stack, pode levar algum tempo para:
    - Instale o aplicativo
    - Insira dados iniciais (ou de recuperação)
    - Configure tudo
    - Inicie o aplicativo
- Para resolver isso podemos usar:
    - Para instâncias EC2:
        - **Golden AMI** : instale o aplicativo, sistema operacional, dependências etc. com antecedência e inicie a instância EC2 a partir do Golden AMI
        - **Bootstrap usando dados do usuário** : para configuração dinâmica podemos usar scripts de dados do usuário

- Para bancos de dados RDS, podemos restaurar o banco de dados a partir de um instantâneo
- Para volumes EBS também podemos restaurar os dados de um snapshot

2_GLOBAL ACCELERATION X CLOUDFRONT

- **Ambos usam a mesma rede global AWS e seus pontos de presença em todo o mundo**
- **Ambos os serviços se integram ao AWS Shield para proteção DDoS**

CLOUDFRONT:

- **Melhora o desempenho de conteúdo armazenável em cache (como imagens e vídeos) e conteúdo dinâmico (como aceleração de API e entrega dinâmica de site)**
- **O conteúdo é servido a partir dos pontos de presença**

GLOBAL ACCELERATION

- **Melhora o desempenho para uma ampla gama de aplicações através de TCP ou UDP**
- **Os pacotes para o aplicativo são proxy dos pontos de presença**
- **Adequado para aplicativos não HTTP, como jogos (UDP), IoT (MQTT) ou voz sobre IP**
- **Bom primeiro para HTTP caso seja necessário ter endereços IP estáticos ou failover regional determinístico e rápido**

3_SHIELD X WAF (Firewall de aplicativos da Web)

AWS Shield usado para proteção DDoS

- **Vem em 2 TIPOS:**
  - **Padrão AWS Shield:**

        - **Serviço gratuito ativado por padrão para todos os clientes da AWS**
        - Fornece proteção contra ataques como inundações SYN/UDP, ataques de reflexão e outros ataques de camada 3/camada 4
    - AWS Shield Advanced Avançado:
        - Serviço opcional de mitigação de DDoS (US$ 3 mil por mês)
        - Protege contra ataques mais sofisticados em EC2, ELB, CloudFront, Global Accelerator, Route 53
        - Acesso 24 horas por dia, 7 dias por semana à equipe de resposta DDoS (DRP) da AWS
        - Protege contra taxas mais altas durante picos devido a DDoS

AWS WAF - Firewall de aplicativos da Web

- Protege aplicativos da web contra explorações comuns da web (camada 7)
- O WAF pode ser implantado no Application Load Balancer, API Gateway e CloudFront
- Para usar o WAF, precisamos definir uma Lista de Controle de Acesso à Web (ACL):
    - As regras podem incluir: endereços IP, cabeçalhos HTTP, corpo HTTP ou strings URI
    - Protege contra ataques comuns, como injeção de SQL e Cross-Site Scripting (XSS)
    - Restrição de tamanho
    - Geo-match, bloquear determinados países
    - Regras baseadas em taxas, para proteção DDoS

Gerenciador de firewall AWS

- Gerenciar regras em todas as contas de uma organização AWS
- Definimos um conjunto comum de regras de segurança no Firewall Manager. Estas regras podem conter regras WAF
- O gerenciador de firewall também pode gerenciar o AWS Shield Advanced

- **Ele também pode gerenciar grupos de segurança para recursos EC2 e ENI em VPC**

4_SQS X SNS

SQS - Simple Queue Service Serviço de Fila Simples

- Integração e Mensagens, SQS - Fila Padrão, Segurança SQS, Tempo limite de visibilidade de mensagens, Filas de mensagens mortas, Fila de atraso, Filas FIFO, SQS com grupo de Auto Scaling, Ordenação de dados em SQS

SNS - Simple Notification Service Serviço de Mensagens Simples

- Segurança, SNS + SQS Fan Out

5_RDS (Multi-AZ e Read Réplicas)

- **É um serviço de banco de dados gerenciado para bancos de dados relacionais**
- **Permite-nos criar bancos de dados na nuvem gerenciados pela AWS**
- **Ofertas RDS fornecidas pela AWS:**
  - **PostgreSQL**
  - **MySQL**
  - **Maria DB**
  - **Oráculo**
  - **Servidor SQL da Microsoft**
  - **aurora**
- **Vantagens do AWS RDS em relação à implantação de um banco de dados relacional no EC2:**
  - **RDS é um serviço gerenciado, o que significa:**
    - **Provisionamento automatizado, aplicação de patches de sistema operacional**
    - **Backups contínuos e restauração para carimbo de data/hora específico (restauração pontual)**
    - **Painéis de monitoramento**

        - Ler réplicas
        - Configuração multiAZ
        - Janelas de manutenção para atualizações
        - Capacidade de escala (vertical e horizontal)
        - Armazenamento apoiado por EBS (GP2 ou IO)
- Desvantagens:
    - Nenhum SSH na instância que hospeda o banco de dados

Backups RDS

- Os backups são habilitados automaticamente no RDS
- AWS RDS fornece backups automatizados:
    - Preencher backup diário do banco de dados (durante a janela de manutenção)
    - Os logs de transações são copiados pelo RDS a cada 5 minutos, o que fornece a capacidade de fazer restaurações pontuais
    - Há uma retenção de 7 dias para os backups que pode ser aumentada para 35 dias
- Instantâneos do banco de dados:
    - Existem backups acionados manualmente pelos usuários
    - A retenção pode durar o tempo que o usuário desejar
    - Útil para manter o estado do banco de dados por um longo período de tempo

Réplicas de leitura RDS

- As réplicas de leitura ajudam a dimensionar as operações de leitura
- Podemos criar até 5 réplicas de leitura
- Essas réplicas podem estar dentro de AZ, entre AZ ou em regiões diferentes
- Os dados entre o banco de dados principal e as réplicas de leitura são replicados de forma assíncrona => as leituras são eventualmente consistentes
- Réplicas de leitura podem ser promovidas em seu próprio banco de dados
- Caso de uso para réplicas de leitura:

    - O banco de dados de produção está instalado e funcionando assumindo carga normal
    - Há um novo recurso para executar alguns relatórios para análises que podem causar lentidão e sobrecarregar o banco de dados
    - Para corrigir isso, podemos criar réplicas de leitura para relatórios
- Réplicas de leitura são usadas para operações SELECT (não INSERT, UPDATE, DELETE)
- Custo de rede para réplicas de leitura:
    - Na AWS, há custo de rede se os dados forem de uma AZ para outra
    - No caso de replicação entre AZ, podem incorrer custos adicionais devido ao tráfego de rede
    - Para reduzir custos, poderíamos ter as réplicas de leitura na mesma AZ

RDS Multi AZ (recuperação de desastres)

- A replicação RDS Multi AZ é feita usando replicação síncrona
- No caso de configuração multi AZ, obtemos um nome DNS
- Caso o banco de dados principal fique inativo, o tráfego é automaticamente redirecionado para o banco de dados de failover
- Multi AZ não é usado para dimensionamento
- As réplicas de leitura podem ser configuradas como Multi AZ para recuperação de desastres

Segurança RDS

Criptografia

- AWS RDS fornece criptografia restante: possibilidade de criptografar o mestre e ler réplicas com AWS KMS - criptografia AES-256
    - A criptografia deve ser definida no momento do lançamento
    - Se o mestre não estiver criptografado, as réplicas de leitura não poderão ser criptografadas
    - A criptografia transparente de dados (TDE) está disponível para Oracle e SQL Server

- **Criptografia em voo: usa certificados SSL para criptografar dados do cliente para o RDS em voo**
    - **É necessário SSL um certificado confiável ao conectar ao banco de dados**
    - **Para impor SSL:**
        - **PostgreSQL: rds.force_ssl=1 no console AWS RDS (grupos de parâmetros)**
        - **MySQL:**`GRANT USAGE ON *.* To 'user'@'%' REQUIRE SSL;`
- **Criptografando backups RDS:**
    - **Instantâneos de bancos de dados RDS não criptografados não são criptografados**
    - **Instantâneos de bancos de dados RDS criptografados são criptografados**
    - **Podemos copiar um instantâneo não criptografado para um criptografado**
- **Criptografe um banco de dados RDS não criptografado:**
    - **Crie um instantâneo**
    - **Copie o instantâneo e ative a criptografia para ele**
    - **Restaure o banco de dados do instantâneo criptografado**
    - **Migre o aplicativo do banco de dados antigo para o novo e exclua o banco de dados antigo**

Segurança de rede e IAM

- **Segurança de rede:**
    - **Os bancos de dados RDS geralmente são implantados em uma sub-rede privada**
    - **A segurança RDS funciona aproveitando grupos de segurança (semelhantes ao EC2), eles controlam quem pode se comunicar com a instância do banco de dados**
- **Gerenciamento de acesso:**
    - **Existem políticas IAM que ajudam a controlar quem pode gerenciar um banco de dados AWS RDS (por meio da API RDS)**
    - **Nome de usuário/senha tradicional pode ser usado para fazer login no banco de dados**
    - **A autenticação baseada em IAM pode ser usada para fazer login no MySQL e PostgreSQL**

- **Autenticação IAM:**
  - **A autenticação de banco de dados IAM funciona com MySQL e PostgreSQL**
  - **Não precisamos de uma senha para autenticar, apenas um token de autenticação obtido por meio de chamadas de API IAM e RDS**
  - **O token tem vida útil de 15 minutos**
  - **Benefícios:**
    - **A entrada/saída da rede deve ser criptografada usando SSL**
    - **O IAM é usado para gerenciar usuários centralmente em vez de credenciais de banco de dados**
    - **Podemos gerenciar funções IAM e perfis de instância EC2 para fácil integração**

Resumo de segurança

- **Criptografia em repouso:**
  - **Isso é feito somente quando o banco de dados é criado**
  - **Para criptografar um banco de dados existente, criamos um instantâneo, copiá-lo como criptografado e criar um banco de dados criptografado a partir do instantâneo**
- **Nossa responsabilidade:**
  - **Verifique as regras de entrada de portas/IP/grupos de segurança**
  - **Cuide da criação e permissões de usuários do banco de dados ou gerencie-os através do IAM**
  - **Crie um banco de dados com ou sem acesso público**
  - **Certifique-se de que os grupos de parâmetros ou o banco de dados estejam configurados para permitir somente conexões SSL**
- **Responsabilidade da AWS:**
  - **Patch de banco de dados**
  - **Patches e atualizações subjacentes do sistema operacional**

6_Tipos de Instâncias EC2 (Spot , Reserved e On-Demand)

EC2

- **Consiste principalmente nos seguintes recursos:**
  - **Alugando máquinas virtuais na nuvem (EC2)**
  - **Armazenando dados em unidades virtuais (EBS)**
  - **Distribuindo carga entre várias máquinas (ELB)**
  - **Dimensionando os serviços usando um grupo de escalonamento automático (ASG)**

Introdução aos grupos de segurança (GS)

- **Grupos de segurança são fundamentais para a segurança de rede na AWS**
- **Eles controlam como o tráfego é permitido dentro ou fora das máquinas EC2**
- **Basicamente eles são firewalls**

Aprofundamento dos Grupos de Segurança

- **Os grupos de segurança regulam:**
  - **Acesso aos portos**
  - **Intervalos de IP autorizados – IPv4 e IPv6**
  - **Controle do tráfego de rede de entrada e saída**
- **Os grupos de segurança podem ser anexados a várias instâncias**
- **Eles estão bloqueados para uma combinação de região/VPC**
- **Eles residem fora das instâncias do EC2 - se o tráfego estiver bloqueado, a instância do EC2 não será capaz de vê-lo**
- ***É bom manter um grupo de segurança separado para acesso SSH***
- **Se a solicitação do aplicativo expirar, provavelmente é um problema do grupo de segurança**
- **Se para a solicitação a resposta for um erro de "conexão recusada", significa que é um erro de aplicação e o tráfego passou pelo grupo de segurança**
- **Por padrão, todo o tráfego de entrada é bloqueado e todo o tráfego de saída é autorizado**
- **Um grupo de segurança pode permitir o tráfego de outro grupo de segurança. Um grupo de segurança pode fazer referência a outro grupo de segurança, o que significa que não há necessidade de fazer**

referência ao IP da instância à qual o grupo de segurança está anexado

IP elástico

- Quando uma instância EC2 é interrompida e reiniciada, ela pode alterar seu endereço IP público
- Caso haja necessidade de um IP fixo para a instância, o Elastic IP é a solução
- Um Elastic IP é um IP público de propriedade do usuário, desde que o IP não seja excluído pelo proprietário
- Com o endereço Elastic IP, podemos mascarar a falha de uma instância remapeando rapidamente o endereço para outra instância
- A AWS fornece um número limitado de 5 IPs elásticos (limite flexível)
- No geral, é recomendado evitar o uso do Elastic IP porque:
    - Eles geralmente refletem decisões arquitetônicas de pool
    - Em vez disso, usamos um IP público aleatório e registramos um nome DNS nele

Dados do usuário EC2

- É possível inicializar (executar comandos para configuração) uma instância EC2 usando o script de dados do usuário EC2
- O script de dados do usuário é executado apenas uma vez na primeira inicialização da instância
- Os dados do usuário EC2 são usados para automatizar tarefas de inicialização, como:
    - Instalando atualização
    - Instalando software
    - Baixando arquivos comuns da internet
    - Qualquer outra tarefa de inicialização
- Os scripts de dados do usuário EC2 são executados com privilégios de usuário root

Tipos de inicialização de instância EC2

- **Instâncias sob demanda:** carga de trabalho curta, preços previsíveis
- **Reservado**: período de tempo conhecido (mínimo 1 ano). Tipos de instâncias reservadas:

    - Instâncias reservadas: cargas de trabalho longas recomendadas
    - Instâncias reservadas conversíveis: recomendadas para cargas de trabalho longas com tipos de instância flexíveis
    - Instâncias reservadas agendadas: instâncias reservadas por um período mais longo usadas em um determinado agendamento
- **Instâncias Spot**: para cargas de trabalho curtas, são baratas, mas há risco de perder a instância durante a execução
- **Instâncias dedicadas**: nenhum outro cliente compartilhará o hardware subjacente
- **Hosts dedicados**: reserve um servidor físico inteiro, pode controlar o posicionamento da instância

EC2 sob demanda

- Pague pelo que usamos, o faturamento é feito por segundo após o primeiro minuto
- Tem o custo mais alto, mas não exige pagamento antecipado
- Recomendado para cargas de trabalho ininterruptas e de curto prazo, quando não podemos prever como o aplicativo se comportará

Instâncias reservadas do EC2

- Desconto de até 75% em comparação com On-demand
- Pagar adiantado por um determinado período, implica compromisso de longo prazo
- O período reservado pode ser de 1 ou 3 anos
- Podemos reservar um tipo de instância específico
- Recomendado para aplicações de uso em estado estacionário (exemplo: banco de dados)
- Instâncias reservadas conversíveis :
    - O tipo de instância pode ser alterado
    - Desconto de até 54%
- Instâncias reservadas agendadas :
    - A instância pode ser iniciada dentro de uma janela de tempo
    - É recomendado quando é necessário que uma instância seja executada em determinados horários do dia/semana/mês

Instâncias Spot EC2

- Podemos obter até 90% de desconto em comparação com instâncias sob demanda
- É recomendado para cargas de trabalho resilientes a falhas, pois a instância pode ser interrompida pela AWS se nosso preço máximo for menor que o preço spot atual
- Não recomendado para trabalhos ou bancos de dados críticos
- Ótima combinação: instâncias reservadas para desempenho básico + instâncias sob demanda e spot para horários de pico

Hosts dedicados EC2

- **Servidor EC2 físico dedicado**
- Fornece controle total do posicionamento da instância EC2
- Ele fornece visibilidade aos soquetes/núcleos físicos subjacentes do hardware
- Requer uma reserva de período de 3 anos
- Útil para software que possuem modelos de licenciamento complicados ou para empresas que possuem fortes necessidades de conformidade regulatória

Instâncias dedicadas EC2

- Instâncias executadas em hardware dedicado a uma única conta
- As instâncias podem compartilhar hardware com outras instâncias da mesma conta
- Sem controle sobre o posicionamento da instância
- Fornece faturamento por instância

Instâncias Spot do EC2 – Aprofundamento

- Com uma instância spot podemos obter um desconto de até 90%
- Definimos um preço spot máximo e obtemos a instância se o preço spot atual <preço spot máximo
- O preço spot por hora varia de acordo com a oferta e capacidade
- Se o preço à vista atual ultrapassar o preço à vista máximo selecionado, podemos optar por interromper ou encerrar a instância nos próximos 2 minutos

- **Bloqueio Spot: bloqueia uma instância spot durante um período de tempo especificado (1 a 6 horas) sem interrupções. Em raras situações, uma instância pode ser recuperada**
- **Solicitação spot - com uma solicitação spot definimos:**
    - **Preço máximo**
    - **Número desejado de instâncias**
    - **Especificações de lançamento**
    - **Tipo de solicitação:**
        - **Solicitação única: assim que a solicitação spot for atendida, as instâncias serão lançadas e a solicitação desaparecerá**
        - **Solicitação de persistência: queremos que o número desejado de instâncias seja válido enquanto a solicitação spot estiver ativa. Caso as instâncias spot sejam recuperadas, a solicitação spot tentará reiniciar as instâncias assim que o preço cair**
- **Cancelar uma instância spot: podemos cancelar solicitações de instância spot se ela estiver no estado aberto, ativo ou desabilitado (não falhou, cancelada, fechada)**
- **O cancelamento de uma solicitação spot não encerra as instâncias executadas. Se quisermos encerrar uma instância spot para sempre, primeiro temos que cancelar a solicitação spot e podemos encerrar as instâncias associadas, caso contrário, a solicitação spot poderá reinicia-las**

Spot Fleet

- **Spot Fleet é um conjunto de instâncias spot e instâncias sob demanda opcionais**
- **A frota spot tentará cumprir a capacidade pretendida com restrições de preço**
- **A AWS lançará instâncias de um pool de lançamento, o que significa que temos que definir o tipo de instância, SO, AZ para um pool de lançamento**
- **Podemos ter vários pools de lançamento dentro do qual o melhor é escolhido**
- **Se um local em que uma frota atingir a capacidade ou o custo máximo, nenhuma nova instância será lançada**

- Estratégias para alocar instâncias spot em uma frota spot:
    - lowerPrice : as instâncias serão lançadas do pool com o menor preço
    - diversificado : as instâncias lançadas serão distribuídas de todos os pools definidos
    - capacidadeOptimized : lançamento com a capacidade ideal com base no número de instâncias

Tipos de instância EC2

- R: aplicativos que precisam de muita RAM – cache na memória
- C: aplicativos que precisam de boa CPU - computação/banco de dados
- M: aplicativos balanceados - aplicativo geral/web
- I: aplicações que precisam de boa E/S local - bancos de dados
- G: aplicativos que precisam de GPU – renderização de vídeo/ML
- T2/T3 – instâncias expansíveis
- T2/T3 ilimitado: burst ilimitado

Instâncias bustáveis (T2/T3)

- No geral, o desempenho da instância está bom
- Quando a máquina precisa processar algo inesperado (um pico de carga), ela pode estourar e a CPU pode ter um desempenho muito bom
- Se a máquina estourar, ela utilizará “créditos estourados”
- Se todos os créditos acabarem, a CPU fica ruim
- Se a máquina parar de estourar, os créditos serão acumulados ao longo do tempo
- O uso de crédito/saldo de crédito de uma instância expansível pode ser visto no CloudWatch
- Créditos de CPU: quanto maior a instância, mais rápido o crédito é obtido
- T2/T3 Ilimitado: poderá ser pago dinheiro extra caso sejam utilizados créditos burst. Não haverá perda de desempenho

AMI

- AWS vem com muitas imagens base

- As imagens podem ser personalizadas em tempo de execução com dados do usuário EC2
- No caso de personalização mais granular, a AWS permite a criação de imagens próprias - isso é chamado de AMI
- Vantagens de uma AMI personalizada:
    - Pré-instalar pacotes
    - Tempo de inicialização mais rápido (conforme necessidade da instância executar os scripts a partir dos dados do usuário)
    - Máquina configurada com software de monitoramento/empresarial
    - Preocupações de segurança – controle sobre as máquinas da rede
    - Controle sobre a manutenção
    - Active Directory pronto para uso
- Uma AMI é criada para uma região específica (NÃO GLOBAL!)

AMI pública

- Podemos aproveitar AMIs de outras pessoas
- Também podemos pagar pela AMI de outras pessoas por hora, basicamente alugando a AMI do AWS Marketplace
- Aviso: não use AMI que não seja confiável!

Armazenamento AMI

- Uma AMI ocupa espaço e é armazenada no S3
- As AMIs, por padrão, são privadas e protegidas por conta/região
- Podemos tornar nossas AMIs públicas e compartilhá-las com outras pessoas ou vendê-las no Marketplace

Compartilhamento de AMI entre contas

- É possível compartilhar AMI com outra conta AWS
- Compartilhar uma AMI não afeta a propriedade da AMI
- Se uma AMI compartilhada for copiada, a conta que fez a cópia se tornará a proprietária
- Para copiar uma AMI que foi compartilhada de outra conta, o proprietário da AMI de origem deve conceder permissões de leitura para o armazenamento que dá suporte à AMI, seja o snapshot do EBS associado ou um bucket do S3 associado.

- **Limites:**
    - **Uma AMI criptografada não pode ser copiada. Em vez disso, se o instantâneo subjacente e a chave de criptografia forem compartilhados, podemos copiar o instantâneo enquanto o criptografamos novamente com uma chave própria. O snapshot copiado pode ser registrado como uma nova AMI**
    - **Não podemos copiar uma AMI com um código billingProduct associado que foi compartilhado conosco de outra conta. Isso inclui AMIs do Windows e AMIs do AWS Marketplace. Para copiar uma AMI compartilhada com o código billingProduct , temos que iniciar uma instância EC2 de nossa conta usando a AMI compartilhada e, em seguida, criar uma AMI a partir da fonte**

Grupos de canais

- **Às vezes queremos controlar como as instâncias EC2 são colocadas na infraestrutura AWS**
- **Ao criarmos um grupo de canais, podemos especificar uma das seguintes estratégias de posicionamento:**
    - **Cluster - agrupar instâncias em um grupo de baixa latência em uma única AZ**
    - **Spread - espalha instâncias pelo hardware subjacente (máximo de 7 instâncias por grupo por AZ)**
    - **Partição - espalhe instâncias por muitas partições diferentes (que dependem de diferentes conjuntos de racks) dentro de uma AZ. Dimensione para centenas de instâncias do EC2 por grupo (Hadoop, Cassandra, Kafka)**

Grupos de veiculações - Cluster

- **Prós: Ótima rede (largura de banda de 10 Gbps entre instâncias)**
- **Contras: se o rack falhar, todas as instâncias falharão naquele momento**
- **Casos de uso:**
    - **Trabalho de big data que precisa ser concluído rapidamente**
    - **Aplicativo que precisa de latência extremamente baixa e alto rendimento de rede**

Grupos de canais - Spread

- Prós:
  - Pode abranger várias AZs
  - Reduz o risco de falha simultânea
  - As instâncias EC2 estão em hardware diferente
- Contras:
  - Limitado a sete instâncias por AZ por grupo de posicionamento
- Caso de uso:
  - Aplicativo que precisa maximizar a alta disponibilidade
  - Aplicações críticas onde cada instância deve ser isolada de falhas

Grupos de posicionamento - Partições

- Prós:
  - Até 7 partições por AZ
  - Pode ter centenas de instâncias EC2 por AZ
  - As instâncias em uma partição não compartilham racks com as instâncias de outras partições
  - Uma falha na partição pode afetar muitas instâncias, mas não afetará outras partições
  - As instâncias obtêm acesso às informações da partição como metadados
- Casos de uso: HDFS, HBase, Cassandra, Kafka

Interfaces de Rede Elástica - ENI

- Componente lógico em uma VPC que representa uma placa de rede virtual
- Uma ENI pode ter os seguintes atributos:
  - Endereço IPv4 privado primário, um ou mais endereços IPv4 secundários
  - Um Elastic IP (IPv4) por IPv4 privado
  - Um IPv4 público
- As instâncias ENI podem ser criadas independentemente de uma instância EC2
- Podemos anexá-los dinamicamente a uma instância EC2 ou movê-los de uma para outra (útil para failover)

- **ENIs estão vinculados a uma zona disponível específica**
- **ENIs podem ter grupos de segurança anexados a eles**
- **As instâncias EC2 geralmente têm uma ENI primária (eth0). Caso anexemos uma ENI secundária, a interface eth1 estará disponível. O ENI primário não pode ser desanexado.**

EC2 Hibernate Hibernação

- **Podemos parar ou encerrar instâncias EC2:**
  - **Se uma instância for interrompida: os dados no disco (EBS) são mantidos intactos**
  - **Se uma instância for encerrada: qualquer volume raiz do EBS também será destruído**
- **Na inicialização, acontece o seguinte no caso de uma instância EC2:**
  - **Primeiro início: o sistema operacional é inicializado e o script de dados do usuário EC2 é executado**
  - **A seguir inicia: o sistema operacional inicializa**
  - **Após a inicialização do sistema operacional, os aplicativos são iniciados, o cache é aquecido, etc., o que pode levar algum tempo**
- **Hibernação EC2:**
  - **Todos os dados da RAM são preservados no desligamento**
  - **A inicialização da instância é mais rápida**
  - **Nos bastidores: o estado da RAM é gravado em um arquivo no volume raiz do EBS**
  - **O volume raiz do EBS deve ser criptografado**
- **Tipos de instância compatíveis para hibernação: C3, C4, C5, M3, M4, M5, R3, R4, R5**
- **Tipos de SO suportados: Amazon Linux 1 e 2, Windows**
- **Tamanho da RAM da instância: deve ser inferior a 150 GB**
- **Instâncias bare metal não suportam hibernação**
- **Volume raiz: deve ser EBS, criptografado, não armazenamento de instância. E deve ser grande o suficiente**
- **O Hibernate está disponível para instâncias sob demanda e reservadas**
- **Uma instância não pode hibernar por mais de 60 dias**

EC2 para arquitetos de soluções

- As instâncias EC2 são cobradas por segundo, t2.micro é de nível gratuito
- No Linux/Mac podemos usar SSH, no Windows Putty ou SSH
- SSH está usando a porta 22, o grupo de segurança deve permitir que nosso IP possa se conectar
- No caso de um tempo limite, é mais provável que seja um problema do grupo de segurança
- Permissão para chave SSH => chmod 0400
- Os grupos de segurança podem fazer referência a outros grupos de segurança em vez de endereços IP
- A instância do EC2 pode ser personalizada na inicialização usando dados do usuário do EC2
- 4 modos de inicialização do EC2:
    - Sob demanda
    - Reservado
    - Spot
    - Dedicated host
- Podemos criar AMIs para pré-instalar software
- Uma AMI pode ser copiada por meio de contas e regiões
- As instâncias do EC2 podem ser iniciadas em grupos de posicionamento:
    - Cluster
    - Spread
    - Partition

GE TAVARES V2