



Anhanguera

# COMPUTAÇÃO EM NUVEM

Qualidade de serviço em nuvem

Prof. Me Fernando S. Claro  
[fernando.claro@anhanguera.com](mailto:fernando.claro@anhanguera.com)

# ÍNDICE

**01.** Introdução

**02.** Qualidade de serviço

**03.** Matérias para as próximas provas

# 01.

## Introdução

- Qualidade de serviço em nuvem é medido através de métricas de desempenho de rede e a definição do conceito de **Acordo de Nível de Serviço (SLA)**.
- Para tanto, vamos conhecer alguns indicadores de Qualidade de Serviço que são utilizados pelos provedores de computação em nuvem.

- Vamos conhecer também mecanismos que visam aprimorar a escalabilidade, a confiabilidade e a disponibilidade de aplicações em nuvem.
- Por últimos vamos discutir a recuperação de desastres.

# 02.

## Qualidade de Serviço



- São métricas quantitativas para descrever os requisitos mínimos de desempenho.
- QoS (Quality of Service) pode ser entendida como uma abordagem utilizada para especificar parâmetros de desempenho das aplicações, assim como os mecanismos necessários para garantir os requisitos de desempenho estabelecido.

- As métricas de desempenho de redes que podem ser utilizadas para os modelos QoS são:
- **Atraso**: é o tempo total de transmissão de um pacote do nó remetente até o nó destinatário. Aplicações de chamadas de voz na Internet, por exemplo, requerem um atraso máximo de 150ms.





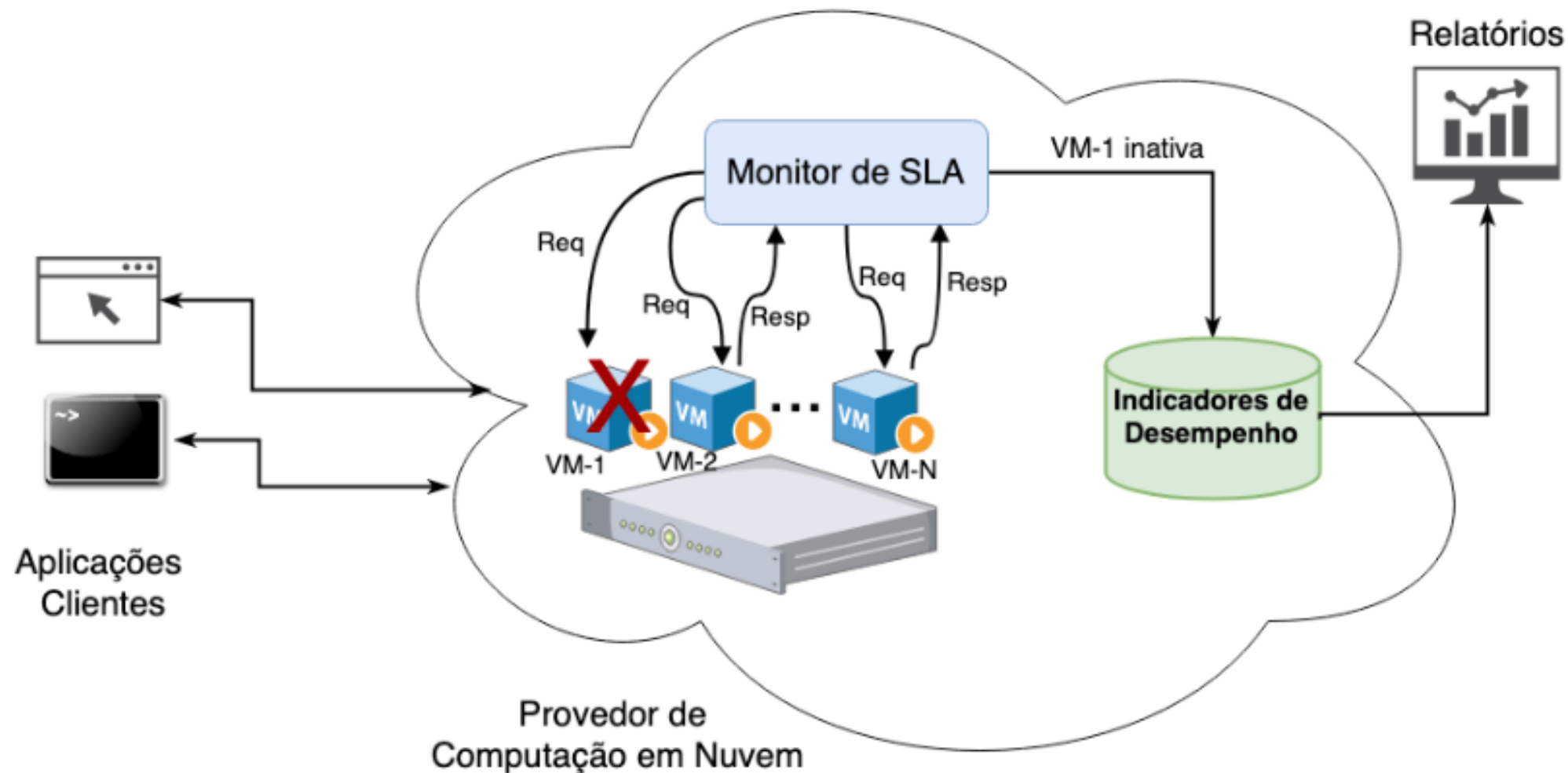
- **Jitter**: é uma medida da variação no atraso na transmissão dos pacotes. Quanto maior o jitter, pior é o desempenho de aplicações multimídia, como streaming de música na internet.
- **Taxa de transmissão**: é o volume de dados efetivamente transmitido entre o nó remetente e o nó destinatário. Em geral, é medida em termos de megabits por segundo (Mbps).

O Netflix estabelece que, para transmitir vídeo em resolução Ultra HD, é necessária uma conexão com uma taxa de pelo menos 25Mbps.

- **Taxa de perda:** é a porcentagem dos pacotes que não foram entregues com sucesso para o nó destinatário. Por exemplo, se foram transmitidos 50 pacotes e apenas 40 foram efetivamente entregues, então a taxa de perda é de 20%.

- Através destas métricas, entre outras, é que os provedores produzem um documento chamado Acordo de Nível de Serviço (SLA – Service Level Agreement).
- O SLA descreve de forma objetiva as garantias de QoS, a confiabilidade e o desempenho de cada serviço.

- Uma das principais métricas utilizadas pelos provedores é determinada como a porcentagem do tempo em que o serviço se mantém apto para responder corretamente as requisições de uma aplicação, cliente ou usuário.
- Por exemplo, se um serviço web ficou "fora do ar" por 8 dias durante um período de 100 dias, então pode-se afirmar que a sua disponibilidade for de apenas 92%.



- Outras métricas também são adotadas pelos provedores como por exemplo:
- **Escalabilidade**: diz respeito à capacidade de um sistema de ajustar a quantidade de recursos computacionais de acordo com a demanda.
- Um sistema é considerado escalável, quando ele é capaz de aumentar a carga de trabalho sem comprometer o desempenho.



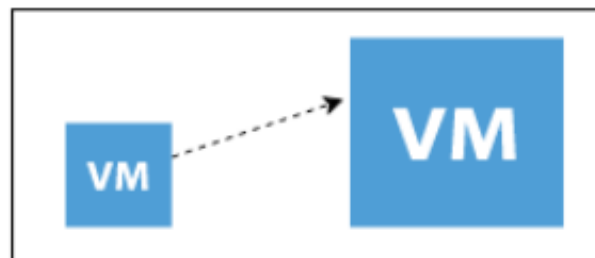
- **Performance:** é uma medida direta da capacidade de execução do serviço. Por exemplo, o tempo de resposta para um cliente receber os dados de uma requisição enviada ao servidor.
- **Confiabilidade:** consiste na capacidade do serviço de operar continuamente sem falhas. Pode ser medida em termos do tempo médio entre falhas, ou seja, o tempo esperado entre a ocorrência de falhas.

- **Resiliência:** representa uma medida da robustez de um serviço. Está relacionada ao grau de tolerância a falhas do serviço.
- Outros tipos de serviços que podem ser monitorados visando o QoS:
  - Dimensionamento automático;
  - Balanceamento de carga;
  - Recuperação de desastres.

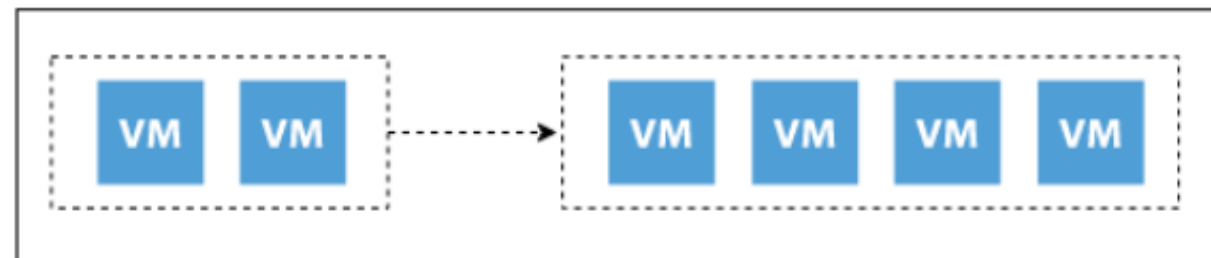


- **Dimensiosamento automático (automated scaling):** é responsável por ajustar a capacidade de um serviço em função das demandas. Se a carga de trabalho aumenta, o mecanismo aloca mais recursos, para manter a performance do serviço.
- O redirecionamento (escalonamento): pode ser vertical ou horizontal.

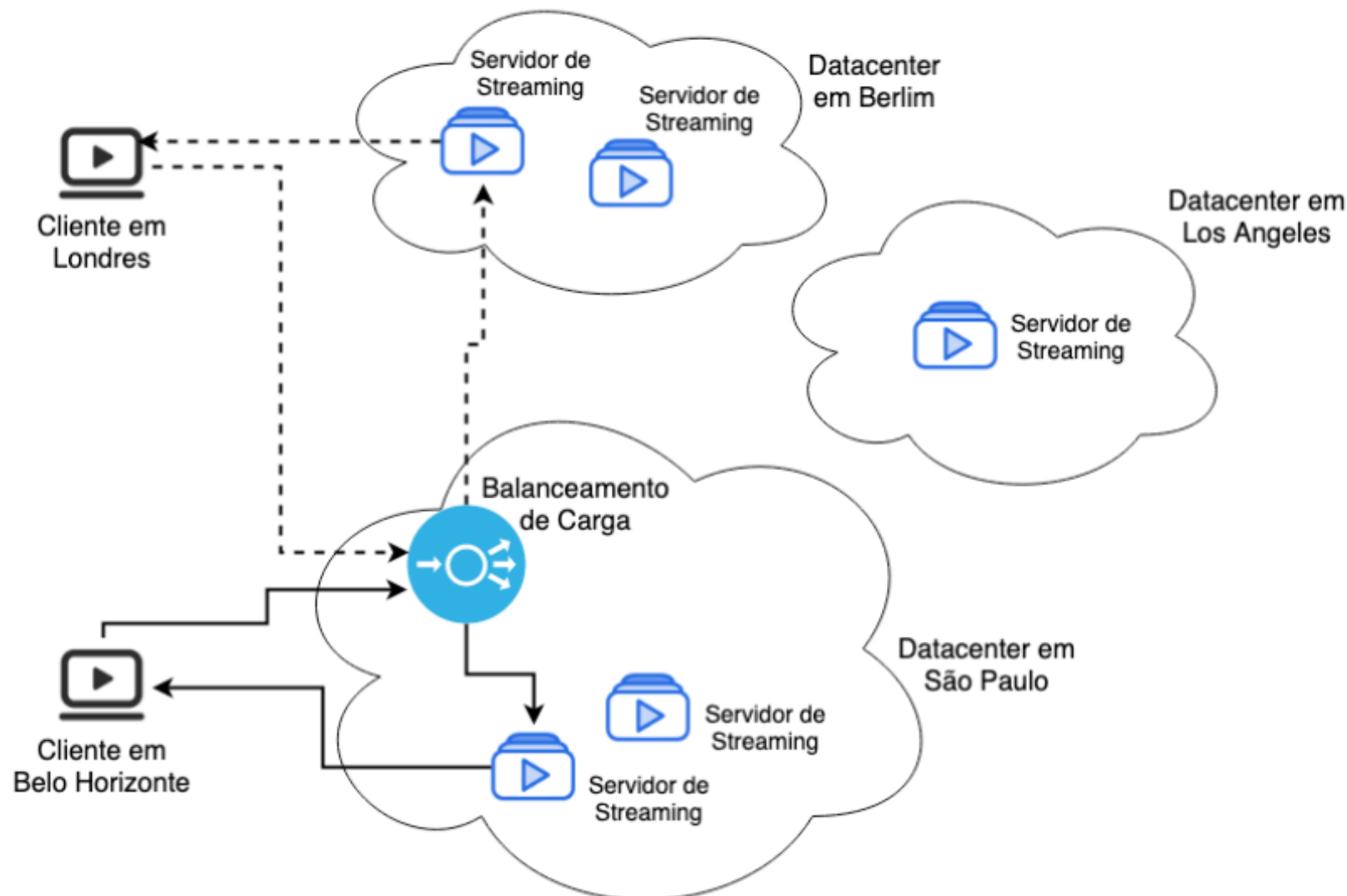
Escalonamento Vertical



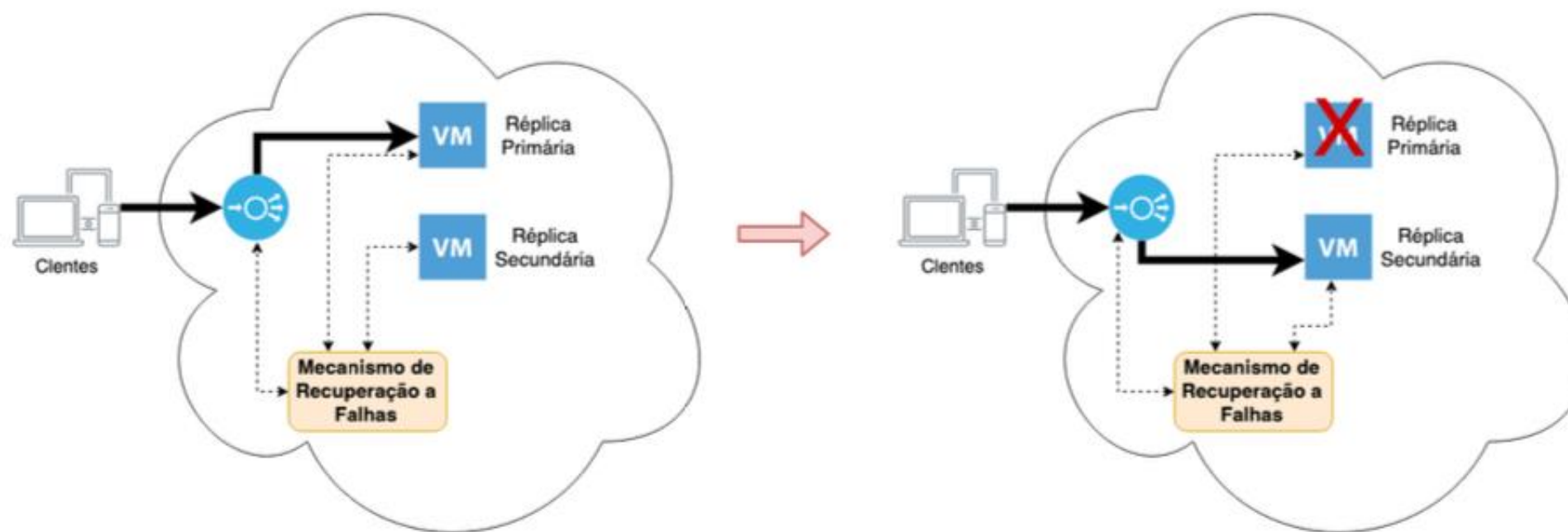
Escalonamento Horizontal



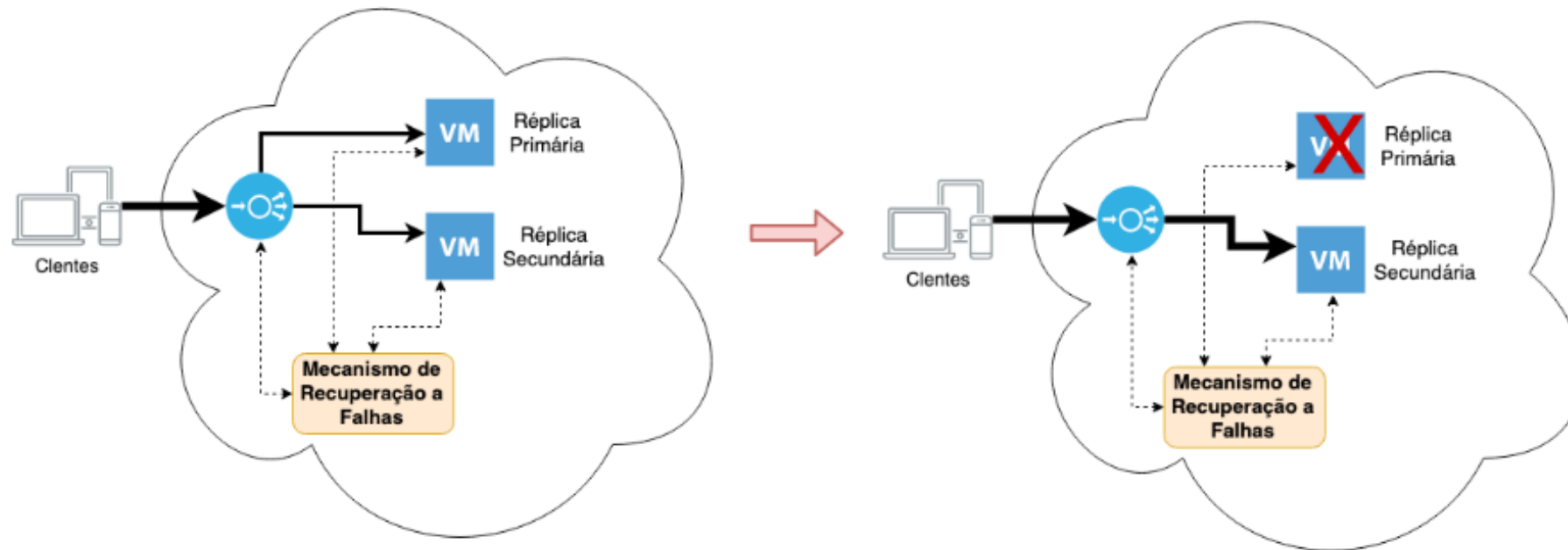
- **Balanceamento de carga:** é um mecanismo para distribuir a demanda de trabalho entre as réplicas de um serviço.



- **Recuperação de falhas:** visa identificar a ocorrência de falhas para que as requisições seja redirecionadas somente para as réplicas do serviço que estejam ativas e funcionando corretamente.



(a) Modelo Ativo-passivo



(b) Modelo Ativo-ativo

# 03.

Matérias para as próximas provas

- **Prova do 2º bimestre:** estudar os seguintes temas:

Aula	Tema
07	Provedores de computação em nuvem
08	Migração de aplicações para a nuvem
10	Serviços de processamento de dados
11	Modelos em arquitetura em nuvem
12	Qualidade de serviço em nuvem

- **Provas de 2ª Chamada e Exame:** estudar os seguintes temas:

Aula	Tema
03	Fundamentos de computação em nuvem
04	Modelos de serviços em computação em nuvem
05	Modelos de implantação em computação em nuvem
06	Tecnologias de suporte à nuvem
07	Provedores de computação em nuvem
10	Serviços de processamento de dados
11	Modelos em arquitetura em nuvem

- Observações:
  - Provas sem consultas.
  - Provas com pelo menos 10 questões de múltipla escolha.
  - Pontuações:
    - 2º Bimestre: 4000 pontos
    - 2ª Chamada: conforme o quadro no próximo slide
    - Exame: 5000 pontos



- Observações:
- Pontuação para a prova de 2ª chamada

Fez a prova do 1º bimestre?	Fez a prova do 2º Bimestre	Valor da 2ª chamada
Não	Sim	1000 pontos
Sim	Não	4000 pontos
Não	Não	4000 pontos