INSTITUTO FEDERAL
Norte de Minas Gerais
Campus Januária

SISTEMAS DISTRIBUÍDOS

ESCALABILIDADE DE DADOS

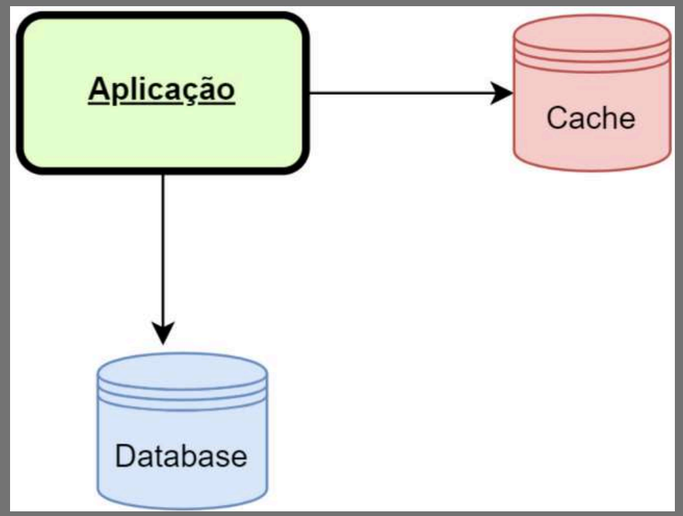
Discente: Mariane Silva Oliveira

INSTITUTO FEDERAL
Norte de Minas Gerais
Campus Januária

CONCEITOS
CACHING

Técnica de armazenamento temporário que provê o acesso rápido e direto a dados sem acessar diretamente a base de dados da aplicação.

O caching gera vantagem na escalabilidade e performance.



1

Hit: O dado buscado se encontra no cache
Miss: O dado buscado não se encontra no cache, fazendo com que tenha que buscá-lo na base de dados.

2

Write through, Write-back (Write-behind), Read through / Lazy Loading e Cache-aside Pattern são algumas estratégias de caching.

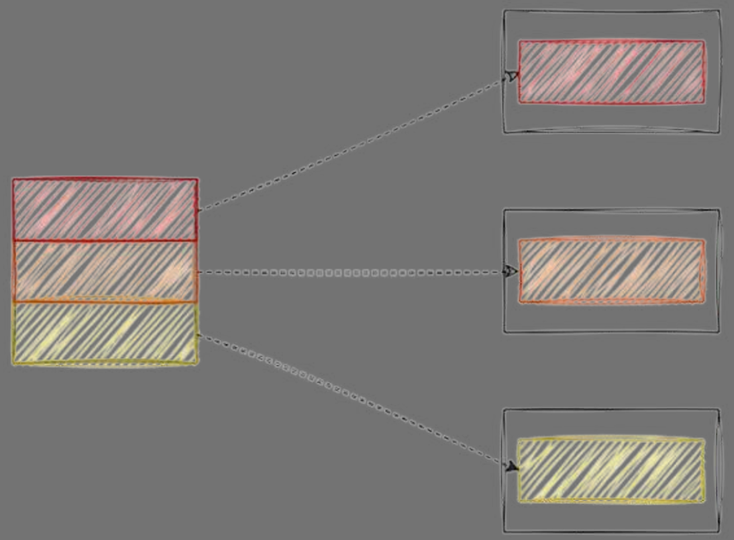
3

Implementações:
Na memória: consiste no armazenamento de dados de caching na própria memória do serviço.
Distribuído: distribui-se o cache entre múltiplas instâncias do servidor de cache.

INSTITUTO FEDERAL
Norte de Minas Gerais
Campus Januária

CONCEITOS **SHARDING**

Sharding ou particionamento é a criação de diferentes subconjuntos do banco de dados original, chamados de shards ou partições. Cada shard pode ser armazenado em diferentes servidores ou em nós de um sistema distribuído.



1 Permite que o sistema gerencie de forma eficiente e segura grandes volumes de dados e facilita a expansão sem a necessidade de reestruturar a base de dados original por completa.

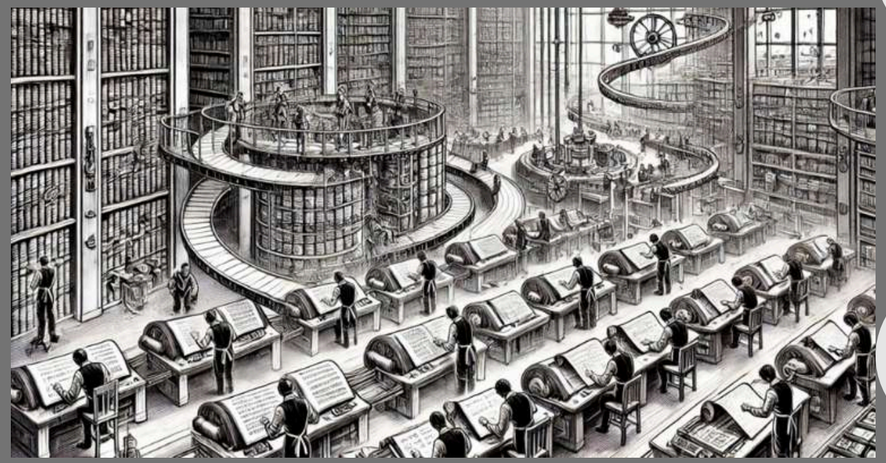
2 Vantagem na escalabilidade, desempenho e disponibilidade em caso de falhas de servidor.

3 **Sharding key:** chave de partição utilizada para determinar como e em qual partição os dados serão armazenados. Existem várias estratégias para defini-la.
Partições Hot: Uma ou mais partições recebem uma carga de trabalho muito alta em comparação com as outras, prejudicando o desempenho.

INSTITUTO FEDERAL
Norte de Minas Gerais
Campus Januária

CONCEITOS
REPLICAÇAO

Criação de uma ou mais cópias dos dados em locais diferentes. A replicação permite escalar sistemas distribuídos de forma inteligente, garantindo desempenho de aplicativo contínuo.



1

Vantagem na consistência, disponibilidade e tolerância a falhas.

2

Alguns dos métodos usados:
Replicação Primary-Replica: um nó primário recebe todas as operações de escrita e, em seguida, replica essas operações para um ou mais nós secundários.
Replicação Primary-Primary ou Multi-Master: múltiplos nós podem atuar simultaneamente como primários.

3

Algumas das estratégias usadas:
Replicação Total e Parcial
Replicação Síncrona
Replicação Assíncrona
Replicação Semi-Síncrona
Replicação por Logs

INSTITUTO FEDERAL
Norte de Minas Gerais
Campus Januária

BANCO DE DADOS RELACIONAL

Modelagem estruturada. Os dados são organizados em tabelas com linhas e colunas. Essas tabelas podem ser relacionadas entre si.

PROBLEMAS DE ESCALABILIDADE

- Escabilidade de bancos relacionais geralmente é vertical, à medida que os dados aumentam, pode ser necessário investir em hardware caro para manter o desempenho.

- Desafioso em casos onde a escalabilidade horizontal é preferível.

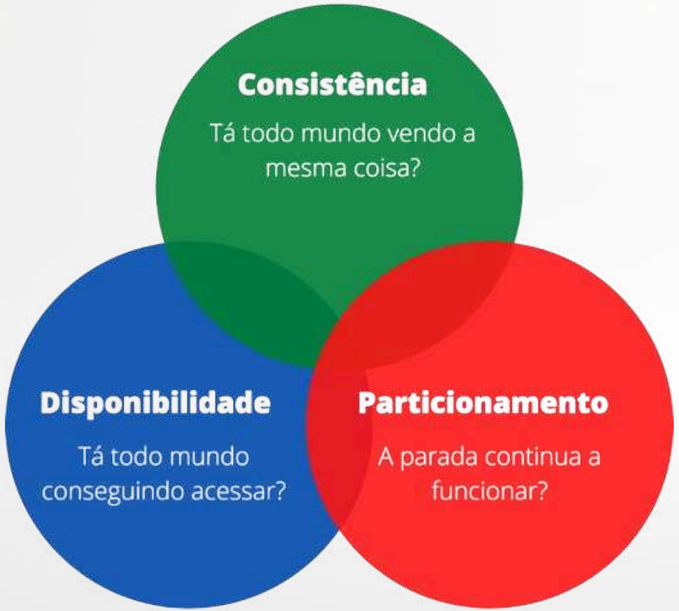
INSTITUTO FEDERAL
Norte de Minas Gerais
Campus Januária

TEOREMA CAP

CONSISTENCY (CONSISTÊNCIA), AVAILABILITY (DISPONIBILIDADE) E PARTITION TOLERANCE (PARTIÇÃO TOLERANTE A FALHAS).

Afirma que em um sistema com distribuição de dados, é impossível obter simultaneamente mais de duas das três garantias: consistência, disponibilidade e partição tolerante a falhas.

Não existe a possibilidade de criar um sistema distribuído sem partição tolerante a falhas, é necessário escolher entre consistência e disponibilidade.



1. A decisão deve depender da necessidade do seu negócio e o peso desses aspectos em relação a ele.
2. Escolher entre um deles não elimina completamente o outro, apenas são inversamente crescentes.
3. É possível obter um resultado interessante combinando estratégias.

INSTITUTO FEDERAL
Norte de Minas Gerais
Campus Januária

CONCEITOS

CONSISTÊNCIA

Bloqueio de todos os nós para gravações adicionais até que os nós que foram desativados voltem a ficar online. Garantindo que todos os nós terão os mesmos valores de entidade.

DISPONIBILIDADE

Quando alguns nós estiverem inativos, os outros nós estarão disponíveis para os usuários realizarem operações de gravação atualizando os dados. Nós que estão inativos não são atualizados com os novos dados.

PARTIÇÃO TOLERÂNTE A FALHAS

Significa que o cluster deve continuar funcionando mesmo se ocorrer uma ou mais falhas de comunicação entre nós no sistema.

INSTITUTO FEDERAL
Norte de Minas Gerais
Campus Januária

NÍVEIS DE CONSISTÊNCIA

FORTE

Todos nós, servidores em todo o mundo, devemos conter o mesmo valor de uma entidade a qualquer momento.

FRACA

Nem todos os nós, servidores em todo o mundo, contem o mesmo valor de uma entidade.

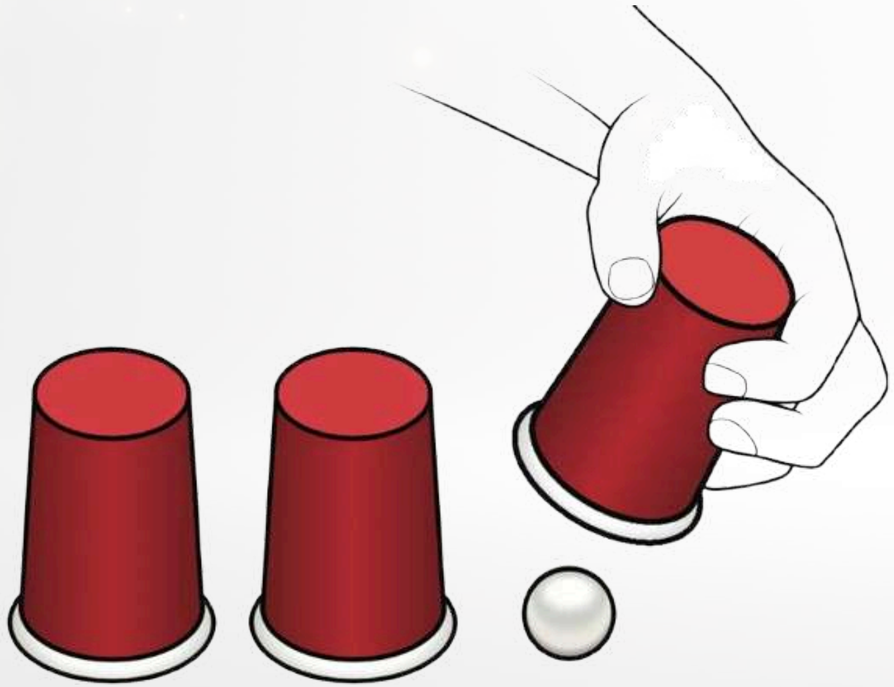
EVENTUAL

Modelo de consistência de dados que permite que os armazenamentos de dados estejam altamente disponíveis.

INSTITUTO FEDERAL
Norte de Minas Gerais
Campus Januária

CONSISTÊNCIA EVENTUAL

O termo foi cunhado no artigo de Bayou no ano de 1995 por Douglas Terry e significa que, para cada objeto em uma coleção de objetos, todas as réplicas de cada objeto eventualmente terão o mesmo valor.



1 Armazenamentos de dados altamente disponíveis.

2 Vantagens:
O sistema pode adicionar novos nós rapidamente
Permite que os usuários finais executem operações de gravação continuamente, sem a necessidade de bloqueio.

3 O sistema mantendo a contagem de usuários simultâneos assistidos a uma transmissão de vídeo ao vivo, lidando com grandes quantidades de dados analíticos.

INSTITUTO FEDERAL
Norte de Minas Gerais
Campus Januária

OUTRA APLICAÇÃO

Diferentes regiões geográficas tem várias zonas de data center de uma rede social: Norte, Leste, Oeste e Sul. Cada uma dessas zonas de data center tem vários clusters com vários nós de servidor em execução.

Bancos de dados são executados em diferentes zonas do data centers, permitindo que o site persista os dados com eficiência. O serviço de armazenamento de dados é altamente disponível. Mesmo que alguns nós fiquem inativos, o serviço de persistência como um todo ainda está ativo

Suponha que uma celebridade crie uma postagem que se torne viral, e todos ao redor do mundo comecem a curtir. Em determinado momento, um usuário na Austrália curte a postagem, o que aumenta a contagem de curtidas ou likes de 100 para 101.

Ao mesmo tempo, um usuário no Brasil, em uma zona geográfica diferente vê a contagem de likes como 100, não 101. Porque o novo valor precisaria de algum tempo para passar da Austrália para a América para atualizar os nós do servidor em execução, atingindo um estado consistente globalmente.

INSTITUTO FEDERAL
Norte de Minas Gerais
Campus Januária

REFERÊNCIAS

CASSIO. Caching: o que é e como funciona. Disponível em: <https://www.softplan.com.br/tech-writers/caching-o-que-e-e-como-funciona/>.

MATHEUS. System Design - Sharding e Particionamento de Dados. Disponível em: <https://fidelissauro.dev/sharding/>. Acesso em: 28 set. 2025.

MATHEUS. System Design - Replicação de Dados. Disponível em: <https://fidelissauro.dev/replicacao/>. Acesso em: 28 set. 2025.

SOARES, E. Banco de Dados Relacional: melhor ou pior? Disponível em: <https://www.dio.me/articles/banco-de-dados-relacional-melhor-ou-pior>.

ROGERIO ANGELISKI. Teorema CAP - O que é isso? Disponível em: <https://angeliski.com.br/teorema-cap-o-que-e-isso>. Acesso em: 28 set. 2025.

STORM, T. Capítulo 4: Consistência, Consistência Eventual e Teorema CAP. Disponível em: <https://medium.com/@tanstorm/cap%C3%ADtulo-4-consist%C3%AAncia-consist%C3%AAncia-eventual-e-teorema-cap-cf0c0a23ad4c>.

INSTITUTO FEDERAL
Norte de Minas Gerais
Campus Januária

OBRIGADA!