



Particionamento

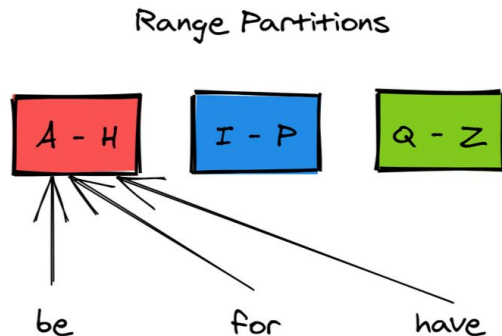
Universidade Tecnológica Federal do Paraná
Bacharelado em Ciência da Computação
Sistemas Distribuídos



Estratégias de Sharding

- Range Partitioning

Dados divididos em intervalos em ordem lexicográfica.



Vantagens:

- Intervalo da partição

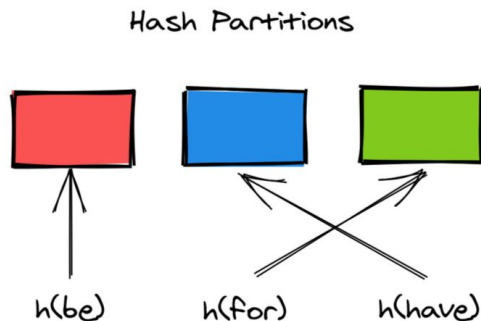
Desvantagens:

- Os dados precisam estar uniformes;
- Desbalanceamentos (hotspots).

Estratégias de Sharding

- Hash Partitioning

Dados divididos por um algoritmo hash



Vantagens:

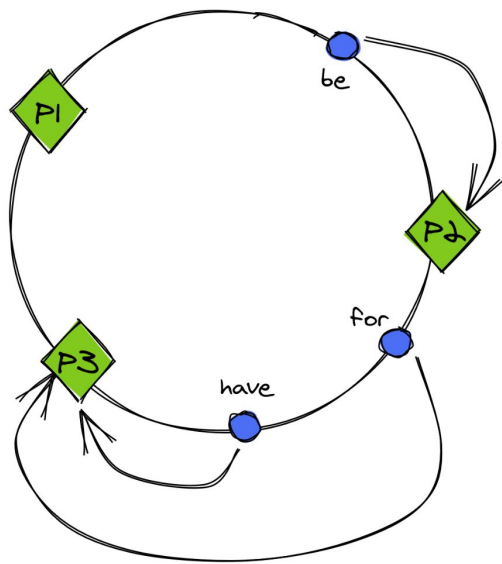
- Distribui as chaves de forma uniforme.

Desvantagens:

- Ainda há chances de hotspots;
- O balanceamento pode ser custoso ao adicionar um novo nó.

Estratégias de Sharding

- Hash Partitioning



Dados divididos por um algoritmo hash em anel

Vantagens:

- Distribui as chaves de forma uniforme
- Previne hotposts
- Menores custos de balanceamento

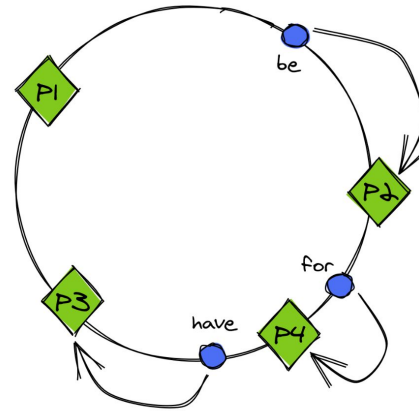
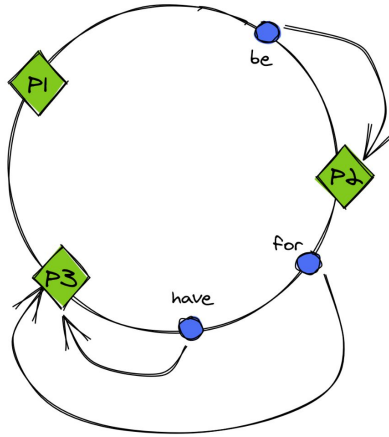
Desvantagens:

- A ordenação das partições é perdida, porém pode ser utilizada uma chave secundária

Estratégias de Sharding

- Hash Partitioning

Dados divididos por um algoritmo hash em anel



Rebalanceamento

- Objetivo: garantir que a carga de trabalho seja distribuída de forma equilibrada entre os nós.
- Durante o rebalanceamento, as partições de dados são movidas entre os nós para garantir um desempenho otimizado e evitar gargalos.
- À medida que o número de nós em um sistema distribuído aumenta ou diminui, é necessário ajustar a distribuição das partições de dados para manter o equilíbrio.
- O rebalanceamento traz benefícios como equilíbrio de carga e adaptação à mudança, garantindo escalabilidade e eficiência.

Particionamento Estático

- Criação de um número fixo de partições no momento da inicialização do sistema
- Quando um novo nó é adicionado, algumas partições são movidas dos nós existentes para o novo nó para manter o equilíbrio.
- O particionamento estático é útil quando a distribuição dos dados é conhecida antecipadamente e não muda com frequência.
- Muitas partições podem adicionar overhead e diminuir o desempenho, enquanto poucas partições podem limitar a escalabilidade do sistema.

Particionamento Dinâmico

- Criação de partições sob demanda, à medida que os dados crescem ou mudam.
- Inicialmente, uma única partição é criada para armazenar os dados.
- Quando a partição cresce além de um determinado tamanho ou se torna muito "quente" (muito acessada), ela é dividida em duas sub-partições, cada uma contendo aproximadamente metade dos dados.
- Uma das sub-partições é transferida para um novo nó, distribuindo a carga de trabalho.
- Da mesma forma, se duas partições adjacentes se tornarem pequenas o suficiente, elas podem ser mescladas em uma única partição.
- O particionamento dinâmico é útil quando a distribuição dos dados não é conhecida antecipadamente ou muda com frequência.
- No entanto, o particionamento dinâmico pode adicionar complexidade ao sistema, pois cada partição é independente e transações podem ser necessárias para atualizar várias partições atômicamente.

Referências

- VITILLO, Roberto. Understanding Distributed Systems. Fevereiro, 2021.