

** Distribuição de informação – Partição **

Resumo

- Exploração diferentes formas de particionar um grande conjunto de dados em subconjuntos menores
- O objetivo principal do particionamento é espalhar os dados e o query load uniformemente em várias máquinas, evitando os hot spots
- Requer escolher um esquema de particionamento e reequilibrar as partições de vez em quando à medida que os nós são adicionados ou removidos do cluster
- Duas abordagens principais para particionamento:
 - Key range partitioning
 - Hash partitioning
- No Hash partitioning, é comum:
 - criar um número fixo de partições antecipadamente
 - atribuir várias partições a cada nó
 - mover partições inteiras de um nó para outro quando os nós são adicionados ou removidos
- Abordagens híbridas também são possíveis, por exemplo:
 - usando uma parte da chave para identificar a partição e outra parte para a ordem de classificação
- Particionamento e índices secundários. Um índice secundário também precisa ser particionado:
 - Document-partitioned index
 - Term-partitioned index (Índice com partição parcial)
- Técnicas para reequilibrar e rotear consultas para a partição apropriada