

noSQL

- pego um arquivo xml é jogado no banco é livre de esquema e é fácil de fazer uma análise de bases;
- É uma ideia de que os dados que são grandes não precisam ser salvos de forma linear e sim de forma constante. Ou seja, é uma base de dados do tipo necessário para Bigdata;

Database distribuído

- O conceito mais geral é de distribuir as tarefas;
- Cada computador é visto como um nó;
- Assim tem uma ideia de que cada computador pode ter um poder de processamento diferente;

Big data

- Não tem um limite;
- É um conjunto de dados que está acima da capacidade de software data base típico

3V

- Volume: cresce exponencialmente;
- Variedade: diferentes resoluções precisam ser levadas em consideração;
- Velocidade: Dado é gerado muito rápido e precisa ser processado;

quem está trabalhando com big data

PROF

- Instrumentos científicos;
- Redes sociais;
- satélites;

Big data x base de dados relacional

- Não é escalável;
- Não são projetados originalmente para serem distribuídos (por isso surgiu o NewSQL);

Escalabilidade

- Se o dado está crescendo ele consigo crescer o armazenamento dele;

- A ideia é que eu quero tornar meu sistema confiávelmente;
- tipos de escalabilidade:
 - horizontal, aumenta a quantidade de nós;
 - Vertical, aumenta o tamanho de cada nó;

modelos de base de dados x sistemas de gerenciamento de bases de dados

NewSQL

- Trás a tentativa de escalar os bancos de dados tanto verticalmente com horizontalmente;
- É Tentar escalar os bancos de dados e realcionar distribuição de dados com os dados de um SQL comum;

Brazil Data cube

- Projeto que trabalha com grande quantidade de volumes de dados;
- Modelas as imagens como cubos de imagens;
- Coloca varias imagens no tempo e coloca as bandas;
- Bota lat long, e o tempo e para cada um ainda temos as bandas espectrais e ainda temos com valores de dados sobre sistemas;
- exemplo de manipulação de
 - Brazil inteiro e em um espaço;
 - AS imagens são armazenados em cogs em um sistema de arquivo distribuido para armazenar as imagens,
 - Faz uma compuzição temporal em que a gente;
 - Faz uma caracterização de busca pelo melhor pixel durante um certo periodo a gente compoe uma imagem que representa aquele periodo;
 - Armazenado em COG(Cloud optmized Geotiff);
 - são salvos algumas informações as bandas espectrais da imagem original , índices de vegetação, máscara de nuvens,
 - Principal objetivo: estudar análise de cubos de dados

- Onde se costuma a ter a ideia de buscar diversos sistemas capazes de capturar as variações ao longo do tempo;
- Existem varios estudos na área de cubos de dados;
- Um exemplo é o mapa de uso e cobertura por que tem quebras de cada uma série através de
- Dado ard, para conseguir ter a reflectancia e a geração de mascara de núvem por exemplo o produto dois e existem máscaras de superficies;
- Hoje tem se cubos mensais a cada 16 dias ou seja as vezes pode se ter mais de uma imagens de satélite