

Hadoop ou Bancos de Dados Relacional?

Sempre que emerge uma nova tecnologia emerge também uma tendência a menosprezar as tecnologias existentes anteriormente. Tenho visto isso com a relação Hadoop e os Bancos de Dados Relacionais (BDR), e por isso resolvi compartilhar alguns pensamentos.

Juro que não estou escrevendo como uma DBA apaixonada, que teve os sentimentos feridos ao ler e ouvir críticas aos bancos de dados relacionais...

Em primeiro lugar o Hadoop não é um banco de dados, embora dados possam ser armazenados nele, ele é um framework para processar grandes quantidades de dados. É composto pelo HDFS, um sistema de arquivos distribuído, e pelo Map Reduce, que é uma camada de computação distribuída.

Isso por si só já mostra que a comparação é questionável, mas vamos continuar...

É sempre bom analisar as características do Hadoop e dos BDR e ver qual deles possui as características relevantes para o seu projeto, o objetivo é extrair o melhor dos dois mundos.

O que o Hadoop tem de bom?

- Código aberto;
- Processamento em batch;
- Utiliza hardware commodity;
- Trata com maestria grandes quantidades de dados, principalmente não estruturados e semi estruturados;
- Um grande ecossistema (várias ferramentas que podem ser usadas com ele);
- Tem o funcionamento excelente quando os dados são gravados e não são alterados.

E os BDR o que tem de bom?

- A linguagem SQL (♥);
- O suporte a transações e por consequência suporte às propriedades ACID (atomicidade, consistência, isolamento e durabilidade);
- O suporte ao processamento interativo;

- Armazena principalmente dados estruturados, ou seja, que possuem um esquema definido;
- Utiliza servidores potentes;

Para os profissionais de Arquitetura de Dados, os desafios estão ficando cada vez mais interessantes. Saímos do tempo onde modelávamos o banco de dados e acabava aí a nossa atuação. Agora é preciso analisar as características dos dados, por exemplo:

- seu volume;
- como eles serão usados;
- como serão acessados;
- requisitos de segurança;
- requisitos de integridade referencial;
- E etc...

E só depois afirmar onde os dados serão armazenados.

Desta forma, não há como afirmar que os BDR são melhores que o Hadoop ou vice-versa. Primeiro porque são diferentes, segundo porque são usados em situações diferentes, e terceiro porque devem "conviver juntinhos" por muito tempo. Sendo assim é bom abrir espaço no coração e gostar dos dois.

Até a próxima!!!

@DanMonteiroDBA

Notas:

- Pode usar este artigo a vontade!!! Mas não esquece de me citar ☺ ;
- Qualquer dúvida, consideração, crítica construtiva e afins... Entra em contato comigo:
 - o [Twitter.com/DaniMonteiroDBA](https://twitter.com/DaniMonteiroDBA) (@DaniMonteiroDBA)
 - o DaniMonteiroDBA@gmail.com
 - o [LinkedIn.com/DaniMonteiroDBA](https://www.linkedin.com/company/danimonteirodba)
 - o [Medium.com/DaniMonteiroDBA](https://medium.com/@danimonteirodba)
 - o facebook.com/DaniMonteiroDBA
 - o <https://github.com/DaniMonteiroDBA>
 - o [TempestadeDeDados.com.br](https://tempestadededados.com.br)