

Lista de exercícios arquiteturas para armazenamento e processamento de BIG DATA

1. Descreva a diferença entre processamento batch e processamento em tempo real no contexto de Big Data. Dê exemplos de casos de uso em que cada abordagem é mais apropriada.
2. Pesquise e explique as características fundamentais do framework Apache Spark. Como ele lida com processamento de dados em memória e em disco, e quais são os benefícios dessa abordagem?
3. Compare a arquitetura do Apache Hadoop MapReduce com a arquitetura do Apache Spark. Quais são as principais melhorias e inovações introduzidas pelo Spark em relação ao MapReduce?
4. Explique a diferença entre processamento paralelo e processamento distribuído. Como essas abordagens são aplicadas em sistemas de processamento de Big Data para otimizar o desempenho?
5. Explique a arquitetura Delta Lake e como ela aborda os desafios de armazenamento e processamento de grandes volumes de dados. Quais são suas principais características que o tornam adequado para cenários de Big Data?
6. Descreva a Lambda Architecture e como ela combina processamento em lote e processamento em tempo real para lidar com dados em larga escala. Quais são os benefícios de ter camadas de processamento separadas?
7. Compare a arquitetura de Delta Lake com a Lambda Architecture em termos de flexibilidade, complexidade e capacidade de gerenciar fluxos de dados. Em quais situações cada uma dessas abordagens pode ser mais apropriada?
8. Quais são os desafios de gerenciar a escalabilidade e a latência em sistemas de armazenamento e processamento de Big Data? Como as arquiteturas como Delta Lake e Lambda Architecture abordam esses desafios?
9. Qual é a importância da escolha da arquitetura correta para armazenamento e processamento de grandes volumes de dados? Quais fatores devem ser considerados ao selecionar a abordagem mais adequada para um determinado cenário?