

Projecto Final de Bases de Dados Distribuídas Avançadas

Mestrado em Ciência de Dados

Sistema de Gestão para “Rede de Bibliotecas” Distribuídas com o Stack Hadoop

Objectivo: Desenvolver um sistema de gestão de bases de dados distribuído para uma rede de bibliotecas, recorrendo ao stack Hadoop, de forma a consolidar os conceitos e técnicas abordados na unidade curricular

1. Introdução e Configuração:
 - a. Instale e configure o Hadoop e o HDFS.
 - b. Compreenda a arquitectura e os fundamentos dos SGBD distribuídos, aplicados ao contexto de bibliotecas.
2. Design da Base de Dados:
 - a. Recorra ao HBase para estruturar uma base de dados distribuída relacionada com livros, autores, empréstimos e utilizadores.
 - b. Estructure as informações relevantes para o funcionamento das bibliotecas.
3. Controlo de Dados:
 - a. Acrescente o ZooKeeper para assegurar a gestão de concorrência, especialmente quando múltiplos utilizadores tentam reservar o mesmo livro.
 - b. Implemente soluções para cenários de múltiplos pedidos por determinados títulos.
4. Processamento de Transacções:
 - a. Use o Apache Phoenix para gerir empréstimos e devoluções, garantindo transacções ACID.
 - b. Implemente consultas SQL para gerir o stock e o estado dos livros.
5. Replicação de Dados:
 - a. Configure a replicação no HDFS, assegurando que os dados das bibliotecas estão consistentes em todas as localidades.
 - b. Avalie a eficácia da replicação na prevenção de perdas de dados.
6. Integração de Bases de Dados:
 - a. Com o Apache Sqoop, sincronize dados entre o sistema Hadoop e um SGBD relacional já existente na rede de bibliotecas

Entrega Final:

- (i) **(até 14 de Dezembro)** Elabore um relatório detalhando cada fase, justificando escolhas, identificando desafios e as abordagens adoptadas.
- (ii) **(14 de Dezembro)** Faça uma demonstração prática do sistema, mostrando funcionalidades e integrações. Reflecta sobre o valor acrescentado deste sistema à rede de bibliotecas e à experiência dos utilizadores.