**UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO**

**1ª PROVA DE BD III ECE 2022-1 – PROF. SERGIO SERRA**

**LEIA AS INSTRUÇÕES E O CADERNO DE PERGUNTAS COM ATENÇÃO ANTES DE FAZER SUA PROVA**

OBSERVAÇÃO 1 - Esta prova é teórico-prática, pode ser feita em casa, sendo entregue através do SIGAA, conforme informado em sala de aula. Você pode consultar softwares a vontade. NÃO COPIE E COLE de outras fontes (Internet ou de colegas). As provas serão corrigidas por comparação. Provas assemelhadas serão zeradas. Essa prova vale 10 pontos.

OBSERVAÇÃO 2 - Você tem várias horas para fazer a prova, pode entregar **até dia 18/07/2022 as 17:59:00**. Ela deve ser anexada ao SIGAA. É responsabilidade EXCLUSIVA do aluno garantir que a prova foi submetida e recebida pelo professor. Certifique-se disso. A cada HORA DE ATRASO será atribuído 1,0 ponto de desconto da nota final da P1. Para isso usarei o relógio do Google. Não deixa para a última hora. Faça a sua prova com calma.  Além disso, você deverá entregar a prova ASSINADA! Para isso basta salvar sua assinatura junto ao arquivo da prova anexado, o arquivo final deve ter **formato PDF**. NÃO ENVIE VERSÕES DISTINTAS. Arquivos com vírus terão automaticamente a nota ZERO!

OBSERVAÇÃO 3 – A frequência do dia da prova será atribuída quando o professor receber o material enviado pelo aluno (caderno de perguntas e respostas de cada questão da prova). Caso não emita resposta, a frequência não será atribuída e a nota será ZERO. Caracterizará uma falta na avaliação.

OBSERVAÇÃO 4 – NÃO TENTE BURLAR O SISTEMA. Seja criativo, a prova não é pesada nem difícil, responda na ordem das perguntas e descreva em DETALHES AS TODAS as soluções encaminhadas para cada questão.

PROVA INDIVIDUAL, RESPOSTAS SOMENTE NA COR AZUL

NOME: Daniel Sant' Anna Andrade\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Nota: \_\_\_\_\_\_\_\_

PROVA INDIVIDUAL, ENTREGAR VIA SIGAA

**1ª QUESTÃO (2 pontos por item)**

Disserte tecnicamente sobre:

* O que são bancos de dados distribuídos? Por que são importantes no contexto na Internet e nuvens de computadores. Quais são as suas principais características técnicas?
* Quais são as principais estratégias de fracionamento de relações. Explique, em detalhes, como elas funcionam e por que essas estratégias são adotadas no contexto dos bancos de dados distribuídos

R = Banco de dados distribuído (BDD) é um banco de dados no qual nem todos os dispositivos de armazenamento estão conectados a um processador comum. Pode ser armazenado em vários computadores, localizados no mesmo local físico, ou podem ser dispersos por uma rede de computadores interconectados.

Bancos de dados distribuidos é um banco de dados onde suas informações estão dispostas em diversos servidores diferentes, pondendo estar todos os servidores no mesmo local físico ou então espalhados por uma rede. Os bancos de dados distribuidos muitas vezes tem um desempenho muito melhor, pois ele não utiliza apenas um processador para realizar suas consultas, e sim um conjunto de processadores de diferentes servidores.

As aplicações que utilizam um banco de dado distribuidos pode ser locais ou globais. A locais, os dados são fixos para cada site. As globais podem utilizar informações de diversos sites diferentes, que podem estar espalhados mundialmente.

Em bancos de dados distribuidos, temos o que se chama de sistemas gerenciador de bancos de dados distribuídos (SGBDDs), esses sistemas podem ser homogêneos e heterogêneos. Os sistemas homogêneos, todos os servidores utilizam o mesmo SGBD, esse sistema é muito mais fácil de projetar e gerenciar. Nos sistemas heterogêneos, os servidores utilizam SGBDs diferentes, isso pode afetar negativamente o processamento das transações. Os sistemas heterogênos geralmente não são muito viáveis, porém, em um mundo globalizado onde diversas coisas diferentes se conectam, é muito mais fácil manter esse sistema do que o homogêneo.

Apesar dos problemas, eles também se destacam em um ambiente globalizado, como:

* Menor custo para uma pequena rede do que para um único servidor potente.
* São utilizados fragmentos do banco de dados com informações específicas para locais específicos.
* Partes específicas podem ser modificadas sem comprometer todo o banco.
* Os dados são localizados próximo do local de maior demanda e os servidores são paralelizáveis, balanceando a utilização dos servidores.
* Uma falha no banco de dados, geralmente irá apenas afetar uma parte do banco de dados, e não ele por completo.

Para que o banco de dados distribuido possa manter seu desempenho, sua confiabilidade e sua disponibilidade, é realizada a técnica de fragmentação, onde suas informações são divididas em frgamentos de tamanhos menores, com cada fragmento sendo tratado como um banco de dados próprio. Isso permite que apenas as informações necessárias sejam enviadas para cada lugar.

Essa fragmentação pode ser de 3 tipos:

* Horizontal: Nessa fragmentação, a separação é feita entre as tuplas, e cada fragmento é um subconjunto das tuplas totais.
* Vertical: Nessa fragmentação, a separação é feita entre os atributos, onde cada fragmento é alguns dos atributos de cada relação utiliza. Um ponto importante é que todos os fragmentos verticais de uma relação possui o atributo primary key, para que ainda seja possível os indentificar.
* Híbrida: Essa fragmentação não possui nenhuma novidade em relação a anterior, os fragmentos tanto terão tuplas específicas como atributos específicos.

**2ª QUESTÃO (2 pontos)**

O que é ECG e ECL no contexto dos bancos de dados distribuídos? Explique por que isso é de importância fundamental. Explique quais são as principais abordagens e descreve em detalhes suas características técnicas e limitações. Dê um exemplo que ilustre a sua argumentação.

R = ECL (Esquema Conceitual Local): possui os dados que são normalmente fragmentados e replicados num banco de dados Distribuido. É necessário descrever a organização lógica de dados. São também chamadas de views do ECG.

ECG (Esquema Conceitual Global): possui a visão organizacional dos dados e descreve a estrutura lógica em de todos os locais, ignorando o fato do banco de dados ser distribuido. ECG é a união de vários ECL.

O esquema conceitual global não é passado diretamente para os diversos nós de um banco de dados distribuido. Isso faria com que muitos problemas pudessem ocorrer, como também seria muito ruim para a segurança geral do banco de dados. Para evitar que problemas ocorram, é utilizado o esquema conceitual local, onde apenas uma cópia (view) da parte necessária é separada em um esquema e enviado para os nós que a utilizarão. O ECG será uma combinações das diversas ECLs utilizadas no banco de dados distribuido.

**3ª QUESTÃO (4 pontos)**

Analise o esquema abaixo que representa uma base centralizada e proponha a fragmentação horizontal. Defina formalmente os mintermos, justifique suas escolhas e representa as novas relações fragmentadas

BOA SORTE