



PLANO DE ENSINO 2016.2

1. DADOS DE IDENTIFICAÇÃO

Disciplina/Módulo: Técnicas de Implementação de Banco de Dados
Código, turma e horário: N592-16 N3AB5A (Lab N592-17 N5B)
Pré-requisito: N527 - Fundamentos de Bancos de Dados
Número de créditos/horas: 3.1 / 72 hs
Horário: N3AB5A (N592-16), N5B (N592-17)
Local: D-55 (N592-16) / M-35 (lab, N592-17)
Professor(a): José de **Aguiar** Moraes Filho, Dr.-Ing.
E-mail: jaguiar@unifor.br

2. SÍNTESE DO CURRÍCULO LATTES (em poucas linhas)

Doutor pela Universidade de Kaiserslautern, Alemanha, em 2010. Mestre em Informática Aplicada pela Universidade de Fortaleza, em 2003. Tem interesses de pesquisa na área de banco de dados tais como: mobilidade, XML, processamento de consultas e transações, novas mídias de armazenamento (memória flash), dentre outras.

3. OBJETIVO GERAL

Identificar e descrever os componentes básicos da arquitetura de um Sistema de Bancos de Dados (SBDs), possibilitando ao aluno a capacidade de analisar as funcionalidades de cada um destes componentes. Serão apresentados e avaliados fatores que influem na performance de SBDs. Adicionalmente, serão introduzidos tópicos avançados da tecnologia de Banco de Dados, como distribuição, XML e mobilidade.

4. OBJETIVOS / CONTEÚDOS

| OBJETIVOS | CONTEÚDOS | CH |
|---|---|-----------|
| Identificar os principais componentes de um sistema de banco de dados centralizado, relacionando-os com as funcionalidades de um sistema de banco de dados. | UNIDADE I - Introdução. 01.01 - Funcionalidades de sistemas de banco de dados; 01.02 - Arquitetura de sistemas de banco de dados. | 2 |

| | | |
|--|---|-----------|
| Apresentar o funcionamento do componente gerenciador de arquivos, descrevendo as principais técnicas de armazenamento de dados. | UNIDADE II - Armazenamento de dados. 02.01 - Meios físicos de armazenamento; 02.02 - Armazenamento em discos magnéticos; 02.03 - Técnicas de RAID; 02.04 - Arquivos. | 4 |
| Apresentar o funcionamento do componente gerenciador de buffer, descrevendo e analisando as principais técnicas paginação em sistemas de banco de dados. | UNIDADE III - Gerenciamento de buffers. 03.01 - Páginas; 03.02 - Mecanismo de paginação de SGBDs; 03.03 - Políticas de alocação de páginas. | 2 |
| Apresentar e analisar técnicas de indexação de dados, implementando as principais estruturas de índices para sistemas de banco de dados. | UNIDADE IV - Indexação. 04.01 - Classificação de índices; 04.02 - Índices primário; 04.03 - Índices multi-níveis; 04.04 - Índices secundários; 04.05 - Índices hash: hash estático; hash dinâmico; 04.06 - Estruturas de índices ordenados: árvores B; árvores B+; grid files. | 8 |
| Descrever o mecanismo de processamento de consultas. | Unidade V - Processamento de consultas. 05.01 - Conceituação de consultas a bancos de dados; 05.02 - Representação de consultas; 05.03 - Implementação de operadores relacionais; 05.04 - Fases do processamento de consultas; 05.05 - Otimização de consultas; 05.06 - Otimização de consultas em bancos de dados orientado a objeto. | 14 |
| Descrever o processamento de transações, conceituando modelo de transação e recuperabilidade. | UNIDADE VI - Processamento de transações. 06.01 - Problema de concorrência em banco de dados; 06.02 - Paradigma de transação; 06.03 - Execução correta de transações concorrentes; 06.04 - Recuperabilidade de schedules. | 14 |
| Descrever e analisar os principais protocolos de controle de concorrência para o modelo clássico de transações. | UNIDADE VII - Protocolos para o controle de concorrência em bancos de dados. 07.01 - Classificação de protocolos para controle de concorrência; 07.02 - Protocolos baseados em mecanismos de bloqueios; 07.03 - Múltipla granularidade de bloqueios; 07.04 - Protocolos que não utilizam mecanismos de bloqueios; 07.05 - Identificação e resolução de deadlocks e livelocks; 07.06 - O mecanismo de versões múltiplas. | 10 |

| | | |
|---|--|----------|
| Definir o processo de recuperação em sistemas de banco de dados, identificando as principais técnicas e ações para o processo de recuperação. | UNIDADE VIII - Recuperação e logging. 08.01 - Classificação de falhas e ações de recuperação de bancos de dados; 08.02 - Técnicas de logging; 08.03 - Conceito de checkpoints; 08.04 - Tipos de checkpoints; 08.05 - Construção do arquivo de log; 08.06 - Procedimento de recuperação após falha. | 8 |
| Identificar e analisar os principais fatores que produzem gargalos em sistemas de banco de dados, descrevendo mecanismos de benchmark, propondo estratégias para redução destes gargalos. | UNIDADE IX - Performance de sistemas de bancos de dados. 09.01 - Fatores que degradam a performance de sistemas de bancos de dados; 09.02 - Benchmarks; 09.03 - Tuning. | 4 |
| Identificar a estratégia de distribuição em sistemas de banco de dados. | UNIDADE X - Distribuição em sistemas de banco de dados. 10.01 - Tipos de distribuição; 10.02 - Bancos de dados cliente servidor; 10.03 - Bancos de dados paralelos; 10.04 - Bancos de dados distribuídos; 10.05 - Bancos de dados heterogêneos; 10.06 - Mobilidade em banco de dados. | 6 |

5. CRONOGRAMA:

2016.2

N592-16 – Técnicas de Implementação de Banco de Dados

Professor: *José de Aguiar*

| Horas/aula | Data | | Conteúdo – Plano de Aulas | |
|------------|------|------|------------------------------------|---|
| | | | AGOSTO 2015 | |
| 18 h/aula | Ter | 02 | 1. | Introdução – Apresentação do programa da disciplina |
| | | | 1.1. | Funcionalidades de Sistemas de Bancos de Dados |
| | | | 1.2. | Arquitetura de Sistemas de Bancos de Dados |
| | Qui | 04 | 2. | Armazenamento de Dados |
| | | | 2.1. | Meios Físicos de Armazenamento |
| | | 2.2. | Armazenamento em Discos Magnéticos | |
| | Ter | 09 | 2.3. | Técnicas de RAID |
| | | | 2.4. | Gerenciamento de Arquivos |
| | Qui | 11 | 3. | Gerenciamento de Buffers |
| | | | 3.1 | Páginas |
| | | | 3.2 | Mecanismo de Paginação de SGBDs |
| | | | 3.3 | Políticas de Alocação de Páginas |
| | Ter | 16 | 4 - | Indexação |
| | | | 4.1 | Classificação de índices; |
| | | | 4.2 | Índices primário; |
| | | | 4.3 | Índices multi-níveis; |
| | | | 4.4 | Índices secundários; |

| | | | |
|----------------------|------------|----|--|
| | <i>Qui</i> | 18 | 4.5 <i>Índices Hash</i> 4.5.1 <i>Hash Estático</i> 4.5.2 <i>Hash Dinâmico</i> 4.6 <i>Estruturas de Índices Ordenados</i> 4.6.1 <i>Árvores B</i> |
| | <i>Ter</i> | 23 | 4.6.2 <i>Árvores B+</i> 4.6.3 <i>Grid Files</i> 4.6.4 <i>Quad-trees</i> 4.6.5 <i>kd-trees</i> |
| | <i>Qui</i> | 25 | 5. <i>Processamento de Consultas</i> 5.1. <i>Conceituação de Consultas a Bancos de Dados</i> 5.2. <i>Representação de Consultas</i> |
| | <i>Ter</i> | 30 | 1ª Avaliação da 1ª NP |
| SETEMBRO 2015 | | | |
| 18 h/aula | <i>Qui</i> | 01 | Correção avaliação 5.2 <i>Implementação de Operadores Relacionais</i> |
| | <i>Ter</i> | 06 | 5.2 <i>Implementação de Operadores Relacionais (cont.)</i> |
| | <i>Qui</i> | 08 | 5.3 <i>Fases do Processamento de Consultas</i> |
| | <i>Ter</i> | 13 | 5.4 <i>Otimização de Consultas</i> |
| | <i>Qui</i> | 15 | ---- <i>Otimização de consultas – Laboratório ----</i> <i>Análise do Otimizador de Consultas do SQL Server e/ou Oracle</i> ---- <i>Lista de exercícios</i> |
| | <i>Ter</i> | 20 | ---- <i>Otimização de consultas – Laboratório ----</i> <i>Análise do Otimizador de Consultas do SQL Server e/ou Oracle</i> |
| | <i>Qui</i> | 22 | 6. <i>Processamento de Transações</i> 6.1 <i>Problema de Concorrência em Banco de Dados</i> 6.2 <i>Paradigma de Transação</i> 6.3 <i>Execução Correta de Transações Concorrentes</i> |
| | <i>Ter</i> | 27 | 6.3 <i>Execução Correta de Transações Concorrentes (cont.)</i> |
| | <i>Qui</i> | 29 | Entrega do Trabalho em equipe da 1ª NP |
| OUTUBRO 2015 | | | |
| 16 h/aula | <i>Ter</i> | 04 | 6.3 <i>Execução Correta de Transações Concorrentes (cont.)</i> |
| | <i>Qui</i> | 06 | 6.4 <i>Recuperabilidade de Schedules</i> ---- <i>lista de exercícios</i> |
| | <i>Ter</i> | 11 | 7. <i>Protocolos para o Controle de Concorrência em Bancos de Dados</i> 7.1 <i>Classificação de Protocolos para Controle de Concorrência</i> 7.2 <i>Protocolos Baseados em Mecanismos de Bloqueios</i> |
| | <i>Qui</i> | 13 | 7.3 <i>Múltipla granularidade de bloqueios;</i> 7.4 <i>Protocolos que não utilizam mecanismos de bloqueios;</i> |

| | | | |
|-----------|-----|---------|---|
| | Ter | 18 | 7.5 Identificação e resolução de deadlocks e livelocks; 7.6 O mecanismo de versões múltiplas. |
| | Qui | 20 | 8 Recuperação e Logging 8.1 Classificação de Falhas e Ações de Recuperação de Bancos de Dados |
| | Ter | 25 | 8.2 Técnicas de Logging |
| | Qui | 27 | Posicionamento, tira-dúvidas e acompanhamento do trabalho de implementação. |
| | | | NOVEMBRO 2015 |
| 16 h/aula | Ter | 01 | 8.3. Conceito de Checkpoints |
| | Qui | 03 | 8.4. Tipos de Checkpoints |
| | Ter | 08 | 8.5. Construção do Arquivo de Log 8.6. Procedimento de Recuperação Após Falha |
| | Qui | 10 | Laboratório: Controle de recuperação no SQL Server e/ou Oracle 9 Performance de Sistemas de Bancos de Dados 9.1 Benchmarks 9.2 Tuning |
| | Qui | 17 | 9.2 Tuning (cont) |
| | Ter | 22 | 10. Distribuição em Sistemas de Banco de Dados 10.1 Classificação 10.2 Bancos de Dados Cliente Servidor 10.3 Bancos de Dados Paralelos |
| | Qui | 24 | 10.4 Bancos de dados distribuídos; 10.5 Bancos de dados heterogêneos; 10.6 Mobilidade em banco de dados. |
| | Ter | 29 | 1ª. Avaliação 2ª NP |
| | | | DEZEMBRO 2015 |
| 6 h/aula | Qui | 01 | Correção avaliação |
| | Ter | 06 | Apresentação dos trabalhos em equipe |
| | Qui | 08 | Revisão Geral dos conceitos |
| | Seg | 12 – 17 | Provas Finais |
| | | | Data final para registro da 1ª. NP: 05/10/2016 Data final para registro da 2ª. NP: 07/12/2016 Período de Provas Finais: 12 a 17/12/2016 |

6. METODOLOGIA

Os conteúdos serão apresentados e discutidos com os alunos usando algumas das metodologias abaixo:

- Aulas expositivas dialogadas;
- Discussão/debate em sala usando livros, textos e/ou publicações na internet sobre atualidades relacionadas aos conteúdos abordados;
- Participação em Palestras Técnicas internas e/ou externas relacionados à área de banco de dados;
- Laboratórios de implementação prática de conteúdos da disciplina.
- Seminários
- Trabalhos individuais e/ou em grupos

7. AVALIAÇÃO

A avaliação da aprendizagem será feita por meio de:

- Aplicação de provas teóricas para cada uma das notas parciais;
- Trabalhos individuais ou em duplas a serem realizados para compor a segunda nota parcial;

A primeira nota parcial será definida pela *maior das notas de uma aplicação de prova e um trabalho em equipes (duplas)*, estas marcadas para os dias **30.08** e **29.09, respectivamente**. A avaliação para a segunda nota parcial envolverá também um trabalho prático (implementação) e uma prova (em **29.11**). Observe a data de entrega do trabalho (**06.12**), pois não serão recebidos trabalhos após estas datas. A segunda nota parcial será definida pela média aritmética entre as notas da aplicação de prova e do trabalho prático.

As provas sempre serão consultadas (inclusive a final). As provas serão sempre cumulativas, ou seja, cada prova abrangerá o conteúdo desde o início do período letivo até a última aula antes da prova. **Ao aluno(a) não será permitido acessar a Internet e usar telefones durante as provas. Somente será permitida consulta ao material individual do próprio do aluno(a) (i.e., livros, notas de aula, cadernos e anotações próprias)**

Lembramos o critério de frequência às aulas, sendo exigido um mínimo de 75% (setenta e cinco por cento) de comparecimento às aulas da disciplina.

Trabalho Prático (Implementação)

Para a 1a. NP

Será um trabalho comum para todas as equipes. A implementação poderá ser feita por equipes de no máximo dois alunos (duplas). A implementação deverá ser feita em JAVA.

O trabalho consistirá na implementação de uma **storage engine**, implementação dos componentes Gerenciador de Arquivos e Gerenciador de Buffer de um SGBD (ver componentes nos slides das notas de aula).

Para o Gerenciador de Buffer, deverá ser implementada a política LRU.

É altamente recomendado que a implementação do Gerenciador de Arquivos use a classe Random Access File do Java.

Os tipos de dados a serem persistidos são Inteiro e String. Porém o armazenamento dos tipos deverá ser em byte array.

Por simplicidade, cada tabela persistida corresponderá a um arquivo físico.

Ver layout físico do arquivo e dos blocos proposto no arquivo correspondente (Estrut-storage-engine.jpg) publicado no Unifor On Line.

O gerenciador de arquivos deve prover os seguintes serviços:

- criar arquivo
- ler bloco
- gravar bloco (existente)
- criar bloco
- excluir bloco

O gerenciador de buffer deve receber requisições das camadas superiores, com TupleIDs, e daí derivar o número do bloco requisitado (ID do bloco). Aplicando a política, ele buscará em memória o bloco e, se não encontrar, ele deve chamar o serviço do gerenciador de arquivo para ler o bloco do arquivo. Após o bloco estar em memória, o gerenciador de buffer fornecerá o ID do bloco para que a camada superior possa processar a tupla no bloco.

Para a 2a. NP

O aluno poderá optar por uma das 5 (cinco) opções abaixo, mutuamente excludentes.

1. O trabalho prático representará a implementação de um índice secundário em árvore B+ com chave de busca composta. Deverão ser disponibilizados os serviços de inserção, construção (carga) a partir dos dados de tabela, deleção e consulta ao índice por um ou mais valor de chave. Assumir cada nó da árvore corresponde a um bloco de dados.

2. O trabalho prático representará a implementação de um protocolo conservador para controle de concorrência. A implementação deverá ser feita em JAVA.

2.1. O protocolo poderá ser o 2V2PL preciso, ou

2.2. O protocolo poderá ser o 2PL preciso

Ambos com múltipla granularidade e detecção/prevenção de deadlocks. Na implementação do protocolo, a entrada deverá ser um conjunto de transações e suas operações qualquer e a saída deve mostrar a sincronização correta (schedule) das operações das transações. A detecção de deadlocks deverá utilizar a estratégia do grafo de espera.

3. O trabalho prático representará a implementação de um algoritmo de junção do tipo hash, podendo ser escolhido

3.1. Double Pipelined Hash Join – DPHJ, ou

3.2. Hybrid Hash Join – HHJ

A implementação deverá ser feita em JAVA. As duas tabelas para o uso e execução do algoritmo estarão armazenadas na storage engine cujo código os alunos construiram por ocasião da NP1. A implementação deverá tratar a leitura das tabelas independentemente, depois irá particionar cada tabela em buckets (build das hash tables) pelo atributo de junção, colocando tuplas correspondentes em cada bucket. A seguir, fará o teste (probe) das tuplas de cada par de buckets correspondentes. (Algoritmo de 2 passos). A implementação deverá tratar eventos de estouro de memória (memory overflow) e possíveis resultados intermediários duplicados. Afinal, nem todos os buckets caberão em memória. Podem ser usados objetos hashTable do Java ou implementar as próprias hash tables. Para o tratamento independente da leitura das tabelas sugere-se uso de threads do Java.

Para todos os trabalhos, os alunos deverão usar a storage engine implementada na NP1. Este código (em Java) proverá serviços de armazenamento em blocos e acesso a tuplas no disco. A carga de dados de tabelas será também fornecida.

A implementação poderá ser feita por equipes de no máximo dois alunos. A entrega do trabalho deverá ser feita no dia **06.12**, onde serão argüidos todos os membros da equipe sobre o trabalho implementado em quesitos como entendimento do problema/solução, clareza e modularidade do código, dentre outros. Trabalhos idênticos ou copiados terão nota ZERO.

Critério para a Escolha do Trabalho Prático (Implementação)

Especificamente para o trabalho da NP2, só será permitida mais de uma equipe no mesmo tópico (opção) caso todas as outras opções já tenham uma equipe alocada, pelo menos. Em outras palavras, não poderá existir uma opção que não vá ser implementada por uma equipe. A distribuição da quantidade de equipes em uma opção deve ser balanceada. Em outras palavras, não será permitida uma opção com muitas equipes e uma outra opção com apenas uma ou poucas equipes. Os alunos são livres para fazer sua opção desde que respeitem este critério.

8. BIBLIOGRAFIA

Bibliografia básica

1. Garcia-Molina, H.; Ullman, Jeffrey D.; Widom, J. Database System Implementation. Prentice Hall. 2000.
2. Bernstein, P. A.; Hadzilacos, V. and Goodman, N. Concurrency Control and Recovery in Database Systems. Addison-Wesley, 1987.
3. O'NEIL, P.; O'NEIL E. Database: principles, programming and performance. 2 ed. San Francisco: Morgan-Kaufman, 1999.

Bibliografia complementar

4. SILBERSCHATZ, A.; KORTH, H.; SUDARSHAN, S. Sistema de banco de dados. Rio de Janeiro: Elsevier, 2006.
5. ELMASRI, R.; NAVATHE, S. B. Sistemas de banco de dados. 4 ed. São Paulo: Pearson, 2005.
6. Brayner, A. Transaction Management in Multidatabase Systems. Schaker-Verlag. 1999.
7. Stonebreaker, M. Readings in Database Systems. Morgan Kaufmann. 1999. Third Edition.
8. GRAY, J.; REUTER, A. Transaction Processing: Concepts and Techniques, Morgan Kaufmann; 1st edition. 1992.
9. Artigos publicados nos principais periódicos e conferências da área de banco de dados:
 - Transactions on Database Systems, Data & Knowledge Engineering, Lecture Notes in Computer Science (LNCS), Journal of VLDB, SIGMOD, PODS, VLDB, DEXA, CoopIS, CAiSE.

9. INFORMAÇÕES COMPLEMENTARES

Sites Importantes:

SBC: Sociedade Brasileira de Computação: www.sbc.org.br

ACM: Association for Computing Machinery: www.acm.org

Atendimento

Em sala de aula, preferencialmente. Na sala J-30 (MIA), 2ª e 4ª noite. Excepcionalmente, por email.

Recomendações

1. O material (slides) fornecido pelo professor representa uma fonte de apoio e não a principal fonte de leitura e pesquisa. Os livros 1, 2, 3 e 4 da bibliografia, apresentada a seguir, são excelentes fontes de pesquisa e com exemplares na biblioteca. Os livros 2, 3 e 4 já apresentam versão em português (na biblioteca tem estas versões).
2. No início do material fornecido (slides) existe um “Guia Bibliográfico por Tópicos” no qual são indicados os principais capítulos/seções (se existem) de cada livro para cada tópico estudado. Isto não significa que o que está apontado no Guia seja a única fonte de pesquisa do aluno. Os alunos devem pesquisar outras fontes (inclusive de outros capítulos/seções do mesmo livro). O Guia apenas auxilia o aluno a melhor se situar frente ao tópico.
3. Não deixar para estudar próximo às avaliações. Defina um horário fixo de estudo diário.
4. Resolver listas de exercícios. Utilizar a aula para tirar dúvidas.
5. Iniciar a implementação do trabalho prático o mais rápido possível.

10. OUTRAS ATIVIDADES DESENVOLVIDAS PELO (A) ALUNO (A)

Os alunos deverão realizar estudos e pesquisas individuais ou em grupos e descritos neste plano e participar de fóruns no Unifor On-Line.

Ressaltamos que as atividades descritas acima visam aprofundar e/ou complementar os conteúdos abordados em sala de aula. Visitas externas e/ou internas a setores da UNIFOR que possuem atividades acadêmicas na área de banco de dados também poderão ser agendadas.